James Hall
Ariel Lindorff
Pamela Sammons *Editors*

International Perspectives in Educational Effectiveness Research

Springer

International Perspectives in Educational Effectiveness Research

James Hall • Ariel Lindorff • Pamela Sammons
Editors

International Perspectives in Educational Effectiveness Research

Springer

Editors
James Hall
Southampton Education School
University of Southampton
Southampton, UK

Pamela Sammons
Department of Education
University of Oxford
Oxford, UK

Ariel Lindorff
Department of Education
University of Oxford
Oxford, UK

ISBN 978-3-030-44809-7 ISBN 978-3-030-44810-3 (eBook)
https://doi.org/10.1007/978-3-030-44810-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to all the hard-working practitioners and policy makers seeking to support equity and improve educational quality across the globe.

Foreword

International Perspectives in Educational Effectiveness Research is a much needed book. As is clear from its accounts of research, there is a fervent desire throughout the world for better education. As the book also makes clear, however, better education needs to include the pursuit of equality as well as of academic prowess.

Traditionally, education systems have been good at working with advantaged, motivated students. They have not been good at working with the least advantaged and those who – for whatever reasons – have little interest in traditional academic achievement. Better education, today, needs to be about creating successful and responsible citizens from all social backgrounds. It should also be about capitalising on the latest neurological developments and on artificial intelligence and should use the full gamut of digital opportunities that are becoming available.

Similarly, as the editors note, researchers in educational effectiveness should be concerned with not only 'what works – but also how, for whom, when, and why'. These are difficult questions for researchers to answer. Evidence needs to be drawn from the individual characteristics of students and teachers, the ways in which the participants are grouped, classroom processes, pedagogical techniques and the methods by which school systems are funded and organised. All of which, of course, are affected by the nature and culture of the societies in which they exist.

The roots of educational effectiveness research lie in a handful of studies undertaken in the United Kingdom, the United States and the Netherlands nearly 50 years ago. Since then there have been numerous studies employing new research and statistical techniques, undertaken in changed environments. There has also been a plethora of international studies, conducted by supranational bodies, testing large samples of students.

Throughout the world much schooling is now dominated by such testing of students, with the tests often used as a means of judging the efficacy of educational systems. Yet, as a number of contributors to this book demonstrate, assessment is complex and testing regimes frequently have unintended negative effects on their participants as well as on entire education systems.

In both sections of the book the editors have sought to integrate expertise from many disciplines. Readers will benefit from philosophical and historical analyses as well as from accounts provided by assessment specialists and statisticians. Ideas have been generated by theoreticians and practical researchers, by those who have taught in schools and those who have not. Most importantly, the editors have sought to bring together findings from research using traditional approaches to school effectiveness and findings from the large-scale international surveys. The need for international cooperation – especially at a time when nationalistic ideas are in the ascendency – is constantly reiterated, together with the need for trustworthy data, open statistical procedures and periodic reviews of what is known.

Since nothing stays exactly the same, the editors are correct to stress the need for findings from all the different kinds of educational effectiveness research to be frequently updated and the theories being deployed (especially if they are dynamic) constantly revised.

A contemporary summation of the theory and practice that has been learned over the years from these different sources and the ensuing implications for the design and development of future systems of education will undoubtedly be of value to fellow researchers engrossed in this field. Politicians and education officers, responsible for national education systems, and hence for the education of future generations of children, would also benefit greatly from taking note of its conclusions.

Professor Peter Mortimore
Former Director, Institute of Education
University of London
London, UK

Abbreviations

AERA	American Educational Research Association
AERJ	American Education Research Journal
ANEP	Administración Nacional de Educación Pública
APPA	Australian Primary Principals Association
BERA	British Education Research Association
CDC	Consejo Directivo Central
CIEP	Council of Initial and Elementary Education
CMO	Charter Management Organisation
CMOE	Chinese Ministry of Education
CPD	Continuing professional development
CVA	Contextual Value Added
DASI	Dynamic Approach to School Improvement
DFID	Department for International Development
DIEE	División de Investigación, Evaluación y Estadística OR Departamento de Investigación y Estadística Educativa
DIPF	German Institute for International Educational Research (Deutsches Institut für Internationale Pädagogische Forschung)
DRC	Democratic Republic of the Congo
EARLI	European Association for Research in Learning and Instruction
ECEC	Early Childhood Education and Care
ECER	European Conference on Educational Research
ECLS	Early Childhood Longitudinal Study
EEC	European Economic Community
EEF	Education Endowment Foundation
EEHE	Entrance Exam to Higher Education
EEIR	Educational Effectiveness and Improvement Research
EEPA	Educational Evaluation and Policy Analysis
EER	Educational Effectiveness Research
EESHS	Entrance Exam to Senior High School
EGRA	Early Grade Reading Assessment

ELS	Education Longitudinal Study
EPPSE	Effective Pre-school, Primary, and Secondary Education Project
EPR	Ecoles Privées
EPUC	Ecoles Publiques Conventionnées
EPUNC	Ecoles Publiques Non Conventionnées
ES	Effect Size
ESCS	Economic, Social and Cultural Status (PISA)
ESEA	Elementary and Secondary Education Act
ESRC	Economic and Social Research Council
ESSA	Every Student Succeeds Act
EU	European Union
EYE (EIT)	Early Years Evaluation (Evaluación Infantil Temprana)
FS	Factor Score
FSM	Free School Meals
GCSE	General Certificate of Secondary Education
GDP	Gross Domestic Product
GNI	Gross National Income
GPA	Grade Point Average
HISEI	Highest occupational status of parents (PISA)
IAEP	International Association for the Evaluation of Educational Progress
ICC	Intra-class Correlation Coefficient
ICSEI	International Congress for School Effectiveness and Improvement
IDACI	Income Deprivation Affecting Children Index
IEA	International Association for the Evaluation of Educational Achievement
IEEQC	Improving Educational Evaluation and Quality In China
ILP	Indigenous Leadership Partner
ILSA	International Large Scale Assessment
IRT	Item Response Theory
ISERP	International School Effectiveness Research Project
ISIP	International School Improvement Project
ISSPP	International Successful School Principalship Project
ISTOF	International Schedule for Teacher Observation and Feedback
IT	Information Technology
ITDEQC	Improving Teacher Development and Educational Quality in China
LEA	Local Education Authority
LGBT+	Lesbian, Gay, Bisexual, Transgender/Transsexual plus (inclusive of other groups e.g. asexual, intersex, queer, questioning, etc.)
LIFT	Learning in Families Together
LOSO	Longitudinal Research in Secondary Education (Longitudinaal Onderzoek Secundair Onderwijs)
MAT	Multi-Academy Trust
MMEER	Mixed Methods Educational Effectiveness Research
MMIRA	Mixed Methods International Research Association

MOREI	Methods of Researching Educational Effectiveness and Improvement
NCEA	National Certificate of Educational Achievement
NCES	National Center for Education Statistics
NCLB	No Child Left Behind
NICHD	National Institute of Child Health and Human Development
NIES	National Institute for Education Sciences
NLE	National Leader of Education
NPD	National Pupil Database
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PALL	Principals as Literacy Leaders
PALLIC	Principals as Literacy Leaders in Indigenous Communities
PASEC	Programme d'Analyse des Systèmes Educatifs de Confemen
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PLASC	Pupil Level Annual Schools Census
PLC	Professional Learning Community
PLP	Personal Leadership Profile
PSM	Propensity Score Matching
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SD	Standard deviation
SDG	Strategic Development Goal
SE	Standard Error
SEC	Socioeconomic Composition
SEN(D)	Special Educational Needs (and Disability)
SER	School Effectiveness Research
SES	Socioeconomic status
SESI	School Effectiveness and School Improvement
SIG	Special Interest Group
SIR	School Improvement Research
SLE	School learning environment
SREE	Society for Research in Educational Effectiveness
TALIS	Teaching and Learning International Survey
TER	Teacher Effectiveness Research
TLB	The Learning Bar
TIMSS	Trends in International Mathematics and Science Study
TSA	Teaching School Alliance
UK	United Kingdom
UKAID	UK Department for International Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
US(A)	United States (of America)
WWC	What Works Clearinghouse

Contents

1 Introduction 1
James Hall, Ariel Lindorff, and Pamela Sammons

Part I Setting the Scene for International Perspectives in Educational Effectiveness Research: Historical Context, Theory, and Extensions

2 International Perspectives in Educational Effectiveness Research: A Historical Overview 9
Ariel Lindorff, Pamela Sammons, and James Hall

3 Developing and Testing Theories of Educational Effectiveness Addressing the Dynamic Nature of Education 33
Leonidas Kyriakides, Bert Creemers, and Anastasia Panayiotou

4 The Fifth Phase of Educational Effectiveness Research: The Philosophy and Measurement of Equity 71
Anthony Kelly

5 Extending Educational Effectiveness: The Middle Tier and Network Effectiveness 101
Daniel Muijs

6 Extending Educational Effectiveness: A Critical Review of Research Approaches in International Effectiveness Research, and Proposals to Improve Them 121
David Reynolds, Anthony Kelly, Alma Harris, Michelle Jones, Donnie Adams, Zhenzhen Miao, and Christian Bokhove

7 Policies and Practices of Assessment: A Showcase for the Use (and Misuse) of International Large Scale Assessments in Educational Effectiveness Research 147
Eckhard Klieme

Part II Examples of Educational Effective Research from Around the Globe

8 Educational Effectiveness Research in Africa: The Case of the Democratic Republic of the Congo (DRC) 185
Gratien Bambanota Mokonzi, Jan Van Damme, Bieke De Fraine, Paul Masimango Vitamara, Gaston Mabela Kimbuani, Augustin Tshite Mukiekie, Stanislas Lukula Maroyi, Oscar Asobee Gboisso, and Jean Paul Legono Bela

9 School and Teacher Value Added Performance and the Relationship with Teacher Professional Development in Mainland China 209
Sally M. Thomas

10 Three Decades of Educational Effectiveness Research in Belgium and the Netherlands: Key Studies, Main Research Topics and Findings 231
Marie-Christine Opdenakker

11 The Impact of Socioeconomic Segregation in U.S. High Schools on Achievement, Behavior, and Attainment and the Mediating Effects of Peers and School Practices 287
Gregory J. Palardy

12 Leadership for Learning in Diverse Settings: School Leaders Setting the Agenda in Australia and New Zealand 327
Tony Townsend, Mere Berryman, David Gurr, and Lawrie Drysdale

13 A National Evaluation of Kindergarten Outcomes: Findings from Uruguay 361
Alma Y. Lopez and J. Douglas Willms

14 Continuing Towards International Perspectives in Educational Effectiveness Research 383
James Hall, Pamela Sammons, and Ariel Lindorff

Index 407

Editors and Contributors

Editors

James Hall is an Associate Professor in the Southampton Education School, University of Southampton, UK. e-mail: j.e.hall@soton.ac.uk

Ariel Lindorff is a Departmental Lecturer in Research Methods in the Department of Education, University of Oxford, UK. e-mail: ariel.lindorff@education.ox.ac.uk

Pamela Sammons is an Emeritus Professor in the Department of Education, University of Oxford and an Emeritus Fellow of Jesus College, University of Oxford, Oxford, UK. e-mail: pamela.sammons@education.ox.ac.uk

Contributors

Donnie Adams is a Senior Lecturer in the Faculty of Education, Institute of Educational Leadership, University of Malaya, Malaysia. e-mail: donnieadams@um.edu.my

Jean Paul Legono Bela is a Senior Lecturer in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: jean-paul.bela@unikis.ac.cd

Mere Berryman is a Professor in the Faculty of Education, Waikato University, New Zealand. e-mail: mere.berryman@waikato.ac.nz

Christian Bokhove is an Associate Professor in the Southampton Education School, University of Southampton, UK. e-mail: c.bokhove@soton.ac.uk

Bert Creemers is a Professor Emeritus and Honorary Professor in Educational Sciences in the Faculty of Behavioural and Social Sciences, University of Groningen, Netherlands. e-mail: b.p.m.creemers@rug.nl

Bieke De Fraine is an Associate Professor in the Center for Educational Effectiveness and Evaluation/KU Leuven, Belgium. e-mail: bieke.defraine@kuleuven.be

Lawrie Drysdale is an Associate Professor in Educational Leadership in the Melbourne Graduate School of Education, University of Melbourne, Australia. e-mail: drysdale@unimelb.edu.au

Oscar Asobee Gboisso is a Senior Lecturer in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: oscar.gboisso@unikis.ac.cd

David Gurr is an Associate Professor in Educational Leadership in the Melbourne Graduate School of Education, University of Melbourne, Australia. e-mail: d.gurr@unimelb.edu.au

Alma Harris is Professor of Education at Swansea University School of Education, Swansea University, UK. e-mail: alma.harris@swansea.ac.uk

Michelle Jones is an Associate Professor and Head of Swansea University School of Education, Swansea University, UK. e-mail: michelle.s.jones@swansea.ac.uk

Anthony Kelly is a Professor of Education in the Southampton Education School, University of Southampton, UK. e-mail: a.kelly@soton.ac.uk

Gaston Mabela Kimbuani is a Professor in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: gaston.kimbuani@unikis.ac.cd

Eckhard Klieme DIPF|Leibniz Institute for Research and Information in Education, Frankfurt, Germany. e-mail: klieme@dipf.de

Leonidas Kyriakides is a Professor of Educational Research and Evaluation and Head of the Department of Education, University of Cyprus, Cyprus. email: kyriakid@ucy.ac.cy

Alma Y. Lopez is a Senior Account Success Manager and Business Development Lead at The Learning Bar Inc., Canada. e-mail: alma.l@thelearningbar.com

Stanislas Lukula Maroyi is a Senior Lecturer in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: stanislas.maroy@unikis.ac.cd

Zhenzhen Miao is an Assistant Professor in the College of Elementary Education, Jiangxi Normal University, China. e-mail: z.miao@jxnu.edu.cn

Gratien Bambanota Mokonzi is a Professor in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: gratien.mokonzi@unikis.ac.cd

Daniel Muijs is Head of Research in the Office for Standards in Education, Children's Services and Skills, UK. e-mail: daniel.muijs@ofsted.gov.uk

Augustin Tshite Mukiekie is an Associate Professor in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: augustin.mukiekie@unikis.ac.cd

Marie-Christine Opdenakker is an Associate Professor of Education at the University of Humanistic Studies, Utrecht, The Netherlands. e-mail: m.opdenakker@uvh.nl (alternative: mariechristineopdenakker@gmail.com)

Gregory J. Palardy is an Associate Professor in the Graduate School of Education, University of California, Riverside, USA. e-mail: gjpalardy@gmail.com

Anastasia Panayiotou is a post-doctoral researcher in the Department of Education, University of Cyprus, Cyprus. e-mail: panayiotou. anastasia@ucy.ac.cy

David Reynolds is an Emeritus Professor at the University of Exeter. e-mail: David.Reynolds@swansea.ac.uk

Sally M. Thomas is a Professor of Education in the School of Education, University of Bristol, UK. e-mail: S.Thomas@Bristol.ac.uk

Tony Townsend is an Adjunct Professor at the Griffith Institute for Educational Research, Griffith University, Queensland, Australia and the University of Tasmania, Hobart, Australia. e-mail: t.townsend@griffith.edu.au

Jan Van Damme is an Emeritus Professor in the Center for Educational Effectiveness and Evaluation/KU Leuven, Belgium. e-mail: jan.vandamme@kuleuven.be

Paul Masimango Vitamara is a Professor in the Service de Planification et d'Évaluation en Éducation, Université de Kisangani, the Democratic Republic of the Congo. e-mail: paulvitamara@yahoo.fr

J. Douglas Willms is President of The Learning Bar Inc., Canada. e-mail: doug.willms@thelearningbar.com

Chapter 1
Introduction

James Hall, Ariel Lindorff, and Pamela Sammons

This book is about teachers, schools, school networks, and educational systems, how they influence the academic and socio-emotional outcomes of students, and the extent to which these impacts are equitable across different groups of students. The book presents a global perspective on this topic from leading academics working in the field of Educational Effectiveness Research (EER) – the field of science that studies variation, quality and equity in education across schools, teachers, networks and systems, attending not only to "what works" but also how, for whom, when, and why.[1] It takes an explicitly international perspective reflecting the growth of EER over the last half century and recognising the increased impacts of globalisation in education policy and practice.

This book is, in part, a response to our need for a contemporary volume that presents a global comparison of EER both as it is, and as it could be. This volume is not intended to supplant the handbooks that have preceded it (Chapman, Muijs, Reynolds, Sammons, & Teddlie, 2016; Teddlie & Reynolds, 2000; Townsend, 2007), but instead to complement, extend, and provide a new perspective and an update to these important contributions to the field. This we have aimed to achieve by bringing together contributions that advance our understanding of the theories and philosophies of educational effectiveness, and studies by leading international experts in EER who have carried out innovative studies across all of the permanently

[1]When this body of knowledge is used to study how schools can improve (i.e. as in School Improvement Research) then the term EER can be extended to Educational Effectiveness and Improvement Research (EEIR; as in Chap. 6).

J. Hall (✉)
Southampton Education School, University of Southampton, Southampton, UK
e-mail: j.e.hall@soton.ac.uk

A. Lindorff · P. Sammons
Department of Education, University of Oxford, Oxford, UK
e-mail: ariel.lindorff@education.ox.ac.uk; pamela.sammons@education.ox.ac.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_1

populated continents of the world. The book is therefore divided into two parts that each serve a different purpose:

Part I sets the scene for the book by providing a historical context to international perspectives in EER, the development of theories and philosophies of EER, and illustrations of where the remit of EER is being extended at the time of writing. In Chap. 2 (Lindorff, Sammons & Hall), we place the book in the appropriate historical context by providing an overview of the various contributions to, and developments in, international perspectives in EER. Subsequent chapters in this part go on to explore theoretical aspects in more depth.

In Chap. 3, Leonidas Kyriakides, Bert Creemers and Anastasia Panayiotou discuss the development of theories in EER, and the development of the dynamic model of educational effectiveness (Creemers & Kyriakides, 2006, 2008). The authors describe the factors of the model at system, school, teacher and student levels, and discuss how meta-analyses and empirical studies have tested the validity of the dynamic model in a variety of international contexts. Kyriakides and colleagues note the need for more international comparative studies to further develop the model and to test its validity in more diverse settings. An emphasis is placed on the importance of measuring and promoting equity as well as quality, and suggestions for improving upon the dynamic model – as well as broader implications for promoting quality and equity in education via theory-driven, evidence-based school improvement – are proposed.

In Chap. 4, there is a further in-depth focus on equity and what it means in education in relation to EER. Anthony Kelly provides a critical discussion of the philosophies that underlie EER, problematising what he terms "a creeping utilitarianism" in the field. He proposes an alternative philosophical orientation based on John Rawls's theory of justice to underpin investigations of equity in EER. This is followed by the presentation of several different statistical metrics to better measure equity and a description of their relative advantages, disadvantages, as well as their technical properties. This discussion suggests that such metrics can complement and extend traditional Contextual Value Added (CVA) statistical models that have featured strongly in the study of school effects in many EER studies across the globe.

In Chap. 5, Daniel Muijs points to the increased scope of EER approaches in different contexts, moving beyond studies that largely focus on institutional effects linked to schools, class membership or teacher effects. His chapter reflects upon our knowledge regarding educational effectiveness of middle-tier structures beyond the level of the school. Focusing particularly on networking and collaboration in education that have received much attention in different countries in the last two decades, Muijs describes how EER has investigated variation and effectiveness within these structures. Drawing on empirical research conducted in England on a range of different collaborative structures, Muijs discusses the mixed evidence regarding the extent to which – and how – networking and collaboration can impact school improvement, specifically in terms of promoting better pupil outcomes.

The next two chapters document the varied relationships that exist between EER and comparative international research including International Large Scale Assessments (ILSAs). In Chap. 6, David Reynolds, Anthony Kelly, Alma Harris, Michelle

Jones, Donnie Adams, Zhenzhen Miao, and Christian Bokhove provide a critical review of historic research methodologies and findings in international "effectiveness research" – including in ILSAs. They propose ways in which EER (and the broader concept including educational improvement – EEIR) might provide useful insight in this area and make recommendations for future studies.

In Chap. 7, Eckhard Klieme discusses past, present and future uses of ILSA findings and methods for the development of international perspectives in EER, approaching this topic from a different but complementary perspective to that of Chap. 6. Klieme outlines some methodological limitations of ILSAs and highlights advances that have been made towards addressing these limitations. He reflects on the impact of ILSAs on national policies and discusses some of the ways in which results have been misinterpreted or misused. Building on this discussion, Klieme proposes ways in which ILSA and EER can be, and have been, mutually informative to enrich our understanding of effectiveness in different country contexts. He illustrates this relationship and provides a review of methodological considerations via a re-analysis of PISA 2015 data focusing on policies and practices of assessment.

Part II provides a contemporary snapshot of studies of educational effectiveness research from around the world over six chapters. The foci of these studies vary in response to the contemporary policies, practices, and social issues within in each country and vary in response to the extent of each country's existing EER knowledge base.

In Chap. 8, Gratien Mokonzi Bambanota, Jan Van Damme, Bieke De Fraine, Paul Vitamara Masimango, Gaston Kimbuani Mabela, Augustin Mukiekie Tshite, Stanislas Maroyi Lukula, Oscar Gboisso Asobee, and Jean Paul Bela Legono report results from a longitudinal study of the effectiveness of schools in the *Democratic Republic of the Congo (DRC)*, conducted by a collaborative research team from the Université de Kisangani in the DRC and the Katholieke Universiteit Leuven in Belgium. The authors provide contextual information about education in the DRC and note how this study addresses a lack of research in the region, particularly with regard to studies conducted by local researchers. Bambanota and colleagues focus on results relevant to primary school effectiveness in mathematics to illustrate the importance of local knowledge and understanding to support appropriate context specificity, here addressed through the use of observations of instructional practice and discussions with practitioners to inform the choice of variables for study.

In Chap. 9, Sally M. Thomas gives an overview of EER to date in mainland *China* and summarises findings from two linked projects ("Improving Teacher Development and Educational Quality in China", and "Improving Educational Evaluation and Quality In China") that were conducted by researchers from the University of Bristol in the UK in collaboration with the Chinese National Institute for Education Sciences in Beijing. Thomas describes and interprets results concerning Chinese secondary schools' raw and value-added teacher and school effects, trends over time, regional variation, differential effects for different groups of students and for different subject areas, and associations between teacher professional development and students' value added progress. Findings illustrate the relevance of local as well as national context, and Thomas adds a comparative perspective by highlighting some

differences and similarities between the results from these studies in China and those from previous research in the UK.

In Chap. 10, Marie-Christine Opdenakker provides a review of the historical development of EER in the *Netherlands* and *Belgium*, where EER has a relatively long and established history. Drawing on a large body of literature generated in these countries, she summarises and discusses findings related to teacher behaviours, learning environment characteristics, group composition, school characteristics, and considerations of the important topic of generic versus differentiated educational effectiveness. The chapter concludes by reflecting on what is known about education at the levels of the school and of the class based on the evidence from the Netherlands and Belgium and suggests several directions for future research.

In Chap. 11, Gregory J. Palardy addresses a gap in the previous EER literature by assessing the impact of socioeconomic segregation on high school students in the *USA*. Palardy investigates the extent to which peer influences and school processes mediate the effects of socioeconomic segregation and highlights the importance of considering a range of outcomes, not only academic performance. Directions for future research on this topic are described, and practical implications are proposed in order to guide policy with the potential to address the issue of socioeconomic segregation in U.S. schools. This chapter thus also makes an important link back to considerations of equity that are discussed in Part I through its focus on socioeconomic segregation at the school level and its consequences.

In Chap. 12, Tony Townsend, Mere Berryman, David Gurr, and Lawrie Drysdale consider the roles of school leaders in establishing learning climates and effectiveness, with the latter broadly conceived to encompass several linked aspects: engagement, learning, achievement, and equity. The authors draw on findings from two case studies of improvement/reform initiatives in *Australia* and *New Zealand*, as part of a larger multi-year international project across 14 countries, to discuss the personal characteristics, behaviours, and complex roles of successful school leaders within the context of reform/improvement initiatives in these countries. The Australian case study focusses on the Principals as Literacy Leaders (PALL) program, while the New Zealand case study focusses on experiences of non-indigenous secondary-school principals within a national policy initiative aiming to improve the educational experiences of Maori students. Townsend and colleagues underscore the importance of how successful school leaders engage with and have influence upon their contexts rather than simply being constrained by them and propose implications for practice with regard to school leadership as critical, political, and key to improvement efforts towards promoting quality and equity in education.

In Chap. 13, Alma Y. Lopez and J. Douglas Willms report results from a nationwide assessment of student outcomes at the end of Kindergarten in *Uruguay,* which the authors note is the first population-based national study of kindergarten outcomes in Latin America. The authors present and interpret findings related to the variation in language and cognitive skills between and within kindergarten classrooms, the extent to which child-level variation is associated with gender and age (birth month within year or the "maturity effect"), and effects of classroom-level SES. Lopez and Willms present three strengths of this research: the population-based

nature of the study, its use of a reliable assessment that was contextualised to meet the needs of Uruguayan teachers, and its measurement of children's skills at kindergarten entry as well as at the end of kindergarten for those initially identified as vulnerable. They propose implications for policy and practice with regard to: identifying strengths and weaknesses in the school system, identifying vulnerable children, informing classroom practice and teacher pre-service and in-service training, and involving parents and setting instructional goals, with the ultimate aim of improving outcomes and reducing inequalities.

This book closes with a chapter by the editors (Chap. 14) that draws together the theories and ideas discussed in Part I, the relationships that can be shared between international perspectives in EER and ILSAs, and the examples of contemporary EER from across the world that are presented in Part II. Themes that are common across these chapters are explored and a heuristic device is proposed to aid the development, testing, and extension of theories of EER. These themes and this device are then used to support our suggestions and speculations on future directions both for international perspectives in EER, and for EER as a major field of inquiry that is now reaching maturity after its rapid growth over the last half-century.

References

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., & Teddlie, C. (Eds.). (2016). *The Routledge international handbook of educational effectiveness and improvement: Research, policy, and practice*. London, UK: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366. https://doi.org/10.1080/09243450600697242

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *International handbook of school effectiveness research*. London, UK: Falmer.

Townsend, T. (Ed.). (2007). *International handbook of school effectiveness and improvement*. Dordrecht, The Netherlands: Springer.

Part I
Setting the Scene for International Perspectives in Educational Effectiveness Research: Historical Context, Theory, and Extensions

Introduction

The chapters in this section provide a backdrop for the examples of Educational Effectiveness Research (EER) in Part 2 of this volume.

The historical overview of EER provided in Chap. 2, including its convergence and synthesis with educational improvement. This feeds into discussions of theoretical frameworks and theoretical perspectives in Chaps. 3 and 4, both of which place an emphasis on equity as well as quality as central concerns for educational research and practice.

Chapters 5, 6 and 7 present extensions of EER that build on the historical and theoretical framing of prior chapters. In Chap. 5, the focus is placed on the "middle tier" (e.g. districts, local authorities) and networks of schools. Chapters 6 and 7 shift to a global scale and present different perspectives on the problems, prospects, and synergies of EER and international comparative research with a focus on International Large Scale Assessment (ILSA).

Taken together, these chapters represent the diversity and range of perspectives in EER and provide the reader with the historical, theoretical, and contextual background to engage with contemporary EER from around the globe.

Chapter 2
International Perspectives in Educational Effectiveness Research: A Historical Overview

Ariel Lindorff, Pamela Sammons, and James Hall

2.1 Introduction

The origins of the modern field of Educational Effectiveness Research (EER) lie in debates surrounding the question, "Do schools make a difference?". One result was the development of the field of School Effectiveness Research (SER) – a field whose research is today encompassed within the more-broadly focussed EER. Extensive discussion of the evolution from SER to EER have been provided by Teddlie and Reynolds (2000), Creemers, Kyriakides, and Sammons (2010), Chapman et al. (2016), and Sammons, Davis, and Gray (2016). Over nearly 50 years, EER has evolved to become an intellectually coherent and widely recognised international field that has moved beyond a focus just on schools and their effects to now ask and answer increasingly complex and nuanced research questions about effects of teachers, classrooms, institutions, networks, and systems. It can be thought of in the present day as the broad endeavour to address questions including:

- How, why, when, where and for whom do education systems, teachers and schools make a difference with regard to a variety of student outcomes?
- What are the most appropriate designs and methodologies to undertake such research?
- How can findings from EER provide evidence and information that can be used to improve learning and life chances for all students, particularly disadvantaged and 'at risk' groups in different cultural contexts?

A. Lindorff (✉) · P. Sammons
Department of Education, University of Oxford, Oxford, UK
e-mail: ariel.lindorff@education.ox.ac.uk; pamela.sammons@education.ox.ac.uk

J. Hall
Southampton Education School, University of Southampton, Southampton, UK
e-mail: j.e.hall@soton.ac.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_2

Rather than being concerned merely with "what works" (a notion more commonly associated with Randomised Controlled Trials, RCTs), EER has historically adopted an ecological perspective that stresses the importance of investigating "natural" variation at different levels of education systems, neighbourhoods and schools (or other institutions), to shed light on issues of equity, quality, and susceptibility to intervention via improvement efforts. Although the roots of EER can be traced back to work conducted in a small number of countries (e.g. the USA, the UK, the Netherlands, and Australia), the field has become increasingly internationally diverse over time. This change has also been required given that questions of educational quality and equity are of universal relevance (even if the answers to these questions and the specific circumstances surrounding them may vary by culture and by country), and because education is increasingly seen as an important policy lever to achieve social and economic aims.

Although this chapter (and the volume within which it is situated) focus on EER, both acknowledge the contribution of educational improvement research towards the development of EER in three areas simultaneously: its internationalisation, its methodologies, and its theoretical perspectives (e.g. Harris & Chrispeels, 2006). Although educational improvement and educational effectiveness developed from different original perspectives and approaches, both share an emphasis on what education systems, institutions and teachers can do to foster positive outcomes for their students, and the two fields have grown together over time and been reciprocally informative. For example, Chap. 6 (Reynolds et al., 2020) uses the acronym EEIR to capture this shared focus on both effectiveness and improvement and both these terms feature within the name and objectives of the ICSEI organisation. The benefits of this are two-fold and mutual. On one hand, the more we learn about the characteristics and processes associated with positive student outcomes, the more insight we have about what can be improved and how (Mortimore, 1998). On the other hand, the more we understand the contexts and cultures within which education operates locally and globally, the better we can adapt our approaches to gain richer and fuller insight regarding the quality and equity of education around the world.

The internationalisation of EER also needs to be contextualised against the backdrop of the increased globalisation of education policy and practice in general. The following sections manifest this reality with a historical account of the international perspective in EER alongside broader contemporary trends towards globalisation in the policy sphere as well as the increasing prevalence and influence of international large-scale assessments (ILSA) and surveys. With the increasing attention of policymakers on the outcomes of international assessments, and their country's league table positions in such assessments and surveys, it is hardly surprising that education policy itself has become increasingly globalised. That is, policymakers have been increasingly concerned with the ability of schools to prepare children and young people to participate in the "global economy", and policy transfer between countries has become increasingly common (for examples see Chaps. 6 and 10 in this volume by Reynolds et al., 2020 and Opdenakker, 2020, respectively). Many supranational organisations such as the OECD and World Bank

continue to accord educational outcomes as key benchmark indicators and set ambitious global educational goals concerning access, participation, and increasingly a focus on quality, equity and improvement of educational outcomes that increasingly shape policy debate and influence national agendas.

A contemporary international example of the increasingly globalisation of educational policy and practice includes the current attempts to transplant, adapt and apply various East Asian approaches to teaching and learning into other settings. In the USA, this has included a push to adopt Singaporean mathematics teaching approaches and textbooks, with a fairly limited evidence base to support the effectiveness of this (e.g. Jaciw et al., 2016). In England, a similar interest in East Asian approaches has emerged with a large amount of resources spent in 2016 to encourage schools to use textbooks and teaching approaches modelled on those used in Shanghai and Singapore. The evidence base for such policy transfer, however, was not strongly rooted in robust empirical evidence; such evidence has been relatively limited and recent (e.g. Hall, Lindorff, & Sammons, 2016; Jerrim & Vignoles, 2015; Lindorff, Hall, & Sammons, 2019). That is, the perceived effectiveness of an educational approach in one cultural setting (East Asian) has often been taken as convincing evidence of its potential to shape policy initiatives intended to change educational practices in another context (Western systems). Raising mathematics has been identified as of particular policy interest in some countries with lower performance in international assessments. However, this assumption of simple cross-country transference is highly problematic, given the complexity of the diverse social, cultural and political contexts within which education in a particular setting is situated.

An additional development in the education policy sphere has been an increasing emphasis on evidence seen as "scientific",[1] particularly priority given to experimental studies which are often laurelled as being a 'gold standard' for producing policy-relevant findings. This is reflected in a shift in funding priorities within countries where this emphasis is especially dominant, and the proliferation of Randomised Controlled Trials (RCTs) and/or Quasi-Experiment Designs (QEDs) to shape policies and practice. For example, both the Education Endowment Foundation (EEF) in England and the What Works Clearinghouse (WWC) in the USA have been influential drivers of evidence-based practice. Both place explicit priority on experimental designs as the most 'trustworthy' form of evidence (Education Endowment Foundation, 2018; What Works Clearinghouse, 2017). Perhaps as a consequence, there has been some complication and confounding of terminology that has long been used in EER, and some regional trends of decline in its prevalence. For example, the Society for Research on Educational Effectiveness (SREE), founded in 2005, states that it "brings together individuals interested in the use of causal

[1]Note: From the authors' perspective, scientific evidence can take multiple forms. In presenting "evidence seen as 'scientific'" in a policy context, we do not propose that experimental evidence is more or less valuable, scientific or valid than observational evidence, merely that the former has – in recent years – appeared to gain prominence and perceived credibility over the latter in non-academic spheres.

inference to improve educational practice" (SREE, 2019). Without devaluing experimental (or quasi-experimental) research or the evidence arising from it, it is important to consider the implications for the field of EER of having, in effect, two uses of the word "effectiveness" in quite distinct ways – one grounded in experimental manipulation and causal inference, the other grounded in observation and analysis of natural variation in nested educational systems at different levels (e.g. national, regional, neighbourhood, school, classes and student). Both may contribute valuable evidence within and beyond the scope of EER, and indeed the more inclusive EEIR that encompasses improvement initiatives as well as studies of educational effectiveness, but the ambiguity of the term "effectiveness" may risk confusion of knowledge bases and professional and academic communities of practice.

Given the growing globalisation of education policy and practice, EER has undergone a dramatic and rapid transformation over the last few decades as comparison of major overviews illustrates (Chapman et al., 2016; Teddlie & Reynolds, 2000), and a greater emphasis on methodological issues and advancement of the field (Creemers et al., 2010). A considerable part of this transformation has involved looking outward from the few places from which the earliest EER (then within SER) emerged to other country contexts and the use of more sophisticated research designs and comparative perspectives. In the sections that follow, we endeavour to trace the history of international perspectives in EER, taking into account the wide variety of studies and developments that together comprise the broad range of manifestations and interpretations of an "international perspective". Rather than limiting our account to comparative international studies, we consider also the international conversations and collections sparked by single-country studies, meta-analyses and syntheses across diverse contexts, as well as steps toward internationalisation of EER that have not been limited to formal academic literature but have arisen from the development of an international community of researchers, policymakers and practitioners.

2.2 Early EER: The Beginning of an International Dialogue

The earliest research on the effectiveness of schools emerged as a response to studies in the USA by Coleman et al. (1966) and Jencks et al. (1972), both of whom found that the "effects" of schools on student outcomes were relatively small in comparison to the effects of student background characteristics; a conclusion also reached by the then contemporary Plowden Report in the UK (Plowden & Central Advisory Council for England, 1967). The implications drawn, then, were that schools did not make much of a difference in terms of student achievement, and schools could therefore do little to mitigate the effects of disadvantage. In reaction to this came studies in the USA (Brookover et al., 1978; Edmonds, 1979; Weber, 1971) and UK (Rutter, Maughan, Mortimore, & Ouston, 1979). These not only indicated that schools had an effect but provided evidence regarding what factors seemed to account for variation between schools in student outcomes. Reynolds (2000)

suggested that school/educational effectiveness studies in this early phase showed strongly "ethnocentric tendencies" (p. 232): studies were generally done in a small selection of settings and reviews of the relevant literature tended to emphasize heavily – if not exclusively – the research conducted in the author's own country.

Compared to studies examining the effectiveness of schools, Teacher Effectiveness Research (TER) arguably began much earlier. The history of this branch of EER can be traced back to studies of teachers perceived to be effective in the late nineteenth century (e.g. Kratz, 1896) and the beginnings of systematic observation (e.g. Stevens, 1912), but it was not until the 1960s that investigations of the relationships between teacher behaviours and student learning outcomes began to emerge (e.g. Gage, 1963). Teacher effectiveness has its roots in North America and shares some of the same general approaches as school effectiveness research in its search for effectiveness factors, albeit while focusing on the classroom rather than the school as a whole (Chapman et al., 2016; Creemers, 1994). As with SER though, TER studies too were generally done in a single context and tended to prioritise, review and acknowledge literature from within that context (e.g. Doyle, 1977).

The focus on within country and culture context in the early days of both school and teacher effectiveness research was not, however, to the exclusion of any international perspective. Instead, these early studies did initiate some degree of dialogue spanning beyond national boundaries. The Coleman et al. (1966) and Jencks et al. (1972) studies elicited responses that were not limited to their country of origin, and the seminal studies by authors including Edmonds (1979), Rutter et al. (1979), and Mortimore, Sammons, Stoll, Lewis, and Ecob (1988) were, and continue to be, frequently cited in literature from outside of their original settings. In that sense, although early reviews of educational effectiveness literature did not always clearly identify or discuss context or national differences, international perspectives were already taking shape – albeit not always explicitly and to a somewhat limited extent – in the form of that international intellectual exchange in the academic sphere. As Mortimore (2001) put it, "...The early existence of independent research projects in two countries asking similar questions and drawing, to a certain extent, on similar methodologies demonstrated the potential for further global investigations" (p. 236).

2.3 The Contribution of the International Congress for School Effectiveness and School Improvement: An International Community of Research and Practice

An international community was already starting to take shape only about a decade after the earliest school effectiveness studies. 1988 marked a pivotal moment in the history of EER: The first meeting of the International Congress for School Effectiveness took place in London, with 14 participating countries (see Townsend, 2007, for a more comprehensive narrative account of the origins of this organisation). The

Congress was intentionally convened not only as a community of researchers, but as an opportunity to bring researchers, policymakers and practitioners together to share ideas and to shape and improve research and practice. Not long after that first meeting, the name of the organisation was updated to the International Congress for School Effectiveness and Improvement (ICSEI) to reflect the ambition to promote convergence of the school effectiveness and school improvement fields. Soon afterward, the journal *School Effectiveness and School Improvement* was established (and associated with ICSEI), with its first issue published in 1990. It has since provided one of the key outlets for research in the field of EER. As of 2019, and as a marker of the extent to which the international EER community has diversified and expanded, ICSEI members come from over 50 countries and a deliberate effort is made by ICSEI to ensure that annual meetings of the organisation take place on different continents each year to promote an inclusive culture and to facilitate the involvement of practitioners and policymakers who might be less able to attend international or distant meetings.

ICSEI is by no means the only forum for the EER and school improvement community, though it is an influential one. Over time, other groups have formed, including the Society for Research in Educational Effectiveness (SREE) with its affiliated journal, the *Journal for Research in Educational Effectiveness* (first published in 2008), the Educational Effectiveness special interest group (SIG 18) in the European Association for Learning and Instruction (EARLI), the Educational Effectiveness and Quality Assurance network of the European Conference on Educational Research (ECER), the School Effectiveness and School Improvement SIG in the American Educational Research Association (AERA) and the Educational Effectiveness and Improvement SIG in the British Education Research Association (BERA). ICSEI, however, is notable in that it has taken as part of its core aims facilitating research and improvement activities through international collaboration and knowledge-sharing, forging links between educational effectiveness and educational improvement, and bridging between research, practice and policy in different international contexts across the globe (e.g. see Schildkamp, 2019, for text from the 2019 President of ICSEI that reflects these aims).

ICSEI has also sought to support international advancement of EER methodology through the creation of the MoREI (Methods of Researching Educational Effectiveness and Improvement) network. This has supported a number of cross-country international collaborations and an interest in developing measures that can be used in different contexts to facilitate comparative research. The International Schedule for Teacher Observation and Feedback (ISTOF; Teddlie, Creemers, Kyriakides, Muijs, & Yu, 2006; Muijs et al., 2018) provides an example of a research collaboration that developed a new instrument based on EER and TER research evidence and expert opinion from 20 participating countries. It is intended to support both research and improvement activities reflecting ICSEI's core aims. We consider the ISTOF project in more detail in Sect. 2.6.

Country reports presented at some of the ICSEI meetings have highlighted the variety of EER studies taking place around the world, and have been published as collections with bridging commentary (e.g. Creemers, Peters, & Reynolds, 1989;

Reynolds, 1996; Townsend, Clarke, & Ainscow, 1999). These collections and the editors' reflections on them contributed, at the time of their publication, a new form of international perspective, looking across contexts and cultures for thematic contributions to as well as similarities and differences in research, policy and practice and the interchange between them. The countries and regions of the world represented in these collections of country reports expanded; by the time Townsend, Clarke and Ainscow's *Third Millennium Schools* (1999) was published, 20 countries were represented, covering all major regions of the globe.

Efforts to create a more inclusive and international EER community have been concerted and ambitious, but their success has arguably been tempered by practical challenges. ICSEI, for example, has expanded its membership but remains to some extent dominated by researchers and policymakers from North America, the UK and Continental Europe. As noted above, the organisation intentionally avoids holding its annual meetings exclusively in these regions, ensuring that it rotates between different regions and includes both high- and low-income host countries, but even with subsidised participation for participants from low-income countries and explicit policy governing the variation in host countries, there are doubtlessly still barriers to participation for those coming from under-funded institutions and low-income nations.

2.4 The Contribution of Single-Setting Studies: A Growing Body of Evidence

The ICSEI country reports discussed in Sect. 2.3 constitute just a sample of the single-country EER studies that have been, and are being, conducted, but they also provide an indication of the spread of EER internationally. By the end of the 1900s although EER studies remained concentrated within the countries in which the field was first established, the settings in which single-country EER studies were taking place had diversified considerably. The subsequent contribution of these studies to an international perspective in EER has been manifold. They have: (1) Allowed for the development and testing of theory in a variety of ways; (2) Provided insight into the context specificity of circumstances and challenges of education systems, schools, teachers and students in particular countries; and (3) Advanced the methodology of the field. Each of these is now considered in turn.

First, critics of early EER drew attention to a perceived lack of theory that underpinned studies of educational effectiveness (Scheerens, 2016). Over time, EER theory has developed to better frame the complex underlying realities of schools, classrooms and education systems (Creemers et al., 2010). The dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) has been particularly influential, accounting for the factors at the classroom, school, and system level – and interactions between these – theorised to be associated with student outcomes (see Chap. 3; Kyriakides, Creemers & Panayiotou, 2020). Scheerens's

(2015) conceptualisation of educational *ineffectiveness* has also constituted an important contribution to theory by drawing attention to factors associated with negative student outcomes and/or small effect sizes. Studies in the context of countries where EER took hold early on have often been responsible for contributing to these developments in EER theory, and have extended it to take into account factors and organisational structures beyond the scope of schools and classrooms. As Muijs (2020) describes in Chap. 5 of this volume, for example, researchers in the UK and USA have gone beyond the proto-typical early EER considerations of school and classroom levels to investigate effectiveness, equity and improvement in education systems/countries (e.g. Van Damme, Liu, Vanhee, & Pustjens, 2010), districts/local authorities (e.g. Caldas & Bankston, 1999; Tymms et al., 2008), and networks of schools (e.g. Ainscow, Dyson, Goldrick, & West, 2012; Chapman & Muijs, 2014; Lindorff, 2016; Muijs, Ainscow, Chapman, & West, 2011).

Second, the proliferation of single-country studies that take place in low-income countries and in countries with a more recent history of EER has contributed insights into the variation of school and teacher factors that account for differences in student outcomes, the variation (and sometimes consistency) of challenges faced by school, teachers, educational networks and systems, and the extent to which these factors and challenges "travel" across countries. For example, Creemers (1999) found that the classroom-level factors linked with student learning outcomes in Indonesian schools were as expected based on the EER knowledge generated in higher-income Western nations. However, the factors at the system level that affected practices within these classrooms were not necessarily accounted for by this prior research carried out elsewhere. Furthermore, some studies outside of established EER national settings evidence a more dramatic reconceptualization of definitions and relationships; the work of Harrison and Kuint (1998) in Israel, for example, demonstrated how effectiveness was reframed in terms of effective school responses to emergency in Israel. In other words, as researchers in a widening variety of locations began to engage with EER and to conduct EER, this did far more than provide specific empirical insights from individual countries; it also afforded the opportunity to test, develop and extend theory, and it did so while raising new and important questions about context and the definition of "effectiveness".

Third, a substantial body of methodological literature in EER had been produced by the end of the twentieth century, with this methodology advanced so as to enable researchers to answer increasingly complex questions concerning the who, what, why, when, where, and how of educational effectiveness (Creemers et al., 2010). However, the extent to which the EER that takes place outside of countries in which EER was earliest established has capitalised on these methodological advances has varied. Multiple reasons for this have been suggested, but researcher capacity and training may be two key explanations. For example, Murillo (2007), in his account of how the EER literature has developed in Latin America since the first studies in the 1970s, notes the proliferation of "more and better studies, greater awareness of the specialized literature, and a new generation of well-trained researchers" (p. 86), and calls attention to the increasing use of sophisticated research methods (e.g. multilevel models, appropriate controls for student background) in that region.

Recent examples include innovative EER designs investigating school and teacher effects in Chile (see for example Ortega, Malmberg, & Sammons, 2018a, 2018b). By contrast, it has been suggested that EER studies in Africa were for a long time limited in both number and methodological sophistication by the relative dearth of trained indigenous researchers who could undertake such studies (Fleisch, 2007), alongside practical difficulties in both appropriate measurement and data collection. The researcher capacity issue has begun to be addressed in some countries (see Chap. 8 of this volume by Bambanota et al., 2020, for an example from the Democratic Republic of the Congo), but the importance of involving researchers with local knowledge and understanding of education systems in low-income countries is an area that still requires attention and effort in the interest of a more inclusive, informed and context-sensitive international EER knowledge base.

It is worth noting that although the vast majority of EER has consisted of single-country studies, there have also been numerous collaborative efforts that have brought together and discussed collections of studies from around the world in books and in special issues of peer-reviewed journals. Some of these collections, like those containing collected ICSEI country reports noted above, have specifically focused on scoping the field with an emphasis on international perspectives. Others have contributed in a more implicit way to international perspectives in the field by gathering researchers together from around the globe to share insights on a particular topic (for an example in teacher effectiveness, see Charalambous & Praetorius, 2018; in early childhood research, see Sammons, Anders, & Hall, 2013; and in methodology, see Sammons & Luyten, 2009).

Of course, there are limitations to the insights that can be gleaned from studies conducted in single countries. For example, the variations in local capacity, methodology and measurement noted above, plus differences in the languages of publication, have led to an unbalanced knowledge base in favour of English-language publications and studies conducted in higher-income countries. Additionally, although single-country studies can generate useful understandings of contexts, a comparative perspective is necessary to more fully understand the interactions between contextual and cultural characteristics and features of education systems, schools and classrooms, this being an important line of enquiry for future research in EER (Reynolds et al., 2016b).

2.5 The Contribution of Literature Reviews and Meta-Analyses as Syntheses of Evidence

As the number of EER studies has increased, and with the increasing diversity of countries in which EER is taking place, there is a growing contribution towards international perspectives in EER beyond that of single country studies. Reviews and meta-analyses of these studies are increasingly being carried out and these seek to synthesise findings across contexts and (sometimes) countries in order to advance

the EER knowledge base. On the most basic level, such syntheses of evidence facilitate investigations of which system, network, school, teacher and classroom factors are most strongly related to student outcomes, and the extent to which variation at each level of EER (see Chap. 14; Hall, Lindorff, & Sammons, 2020) is consistent across different settings. Beyond that, these meta-analyses and reviews have also allowed for the extension and testing of theory in EER and have raised important questions about context-specificity.

Existing literature reviews provide information about the range of countries in which EER has been conducted, and offer insight regarding some of the differences and similarities between these countries in terms of the research methodologies used, empirical findings, and contextual and cultural factors. Although early reviews reflected the above-noted "ethnocentricity" (see Sect. 2.2) in generally referring to studies within the authors' own national context(s), examples from the first *International Handbook of School Effectiveness Research* (Teddlie & Reynolds, 2000) onwards have more commonly explicitly attended to the international aspect and considered patterns, for example, across low- and high-income countries (e.g. Scheerens, 2001; Thomas, Kyriakides, & Townsend, 2016). Recent state-of-the-art narrative reviews of the literature in educational effectiveness (Reynolds et al., 2014), teacher effectiveness (Muijs et al., 2014), and school and system improvement (Hopkins, Stringfield, Harris, Stoll, & Mackay, 2014) have all to some extent noted the need for, and ways of, dealing with context-specificity, including what Thomas et al. (2016) term "micro aspects of context" (i.e. context at the level of the school and of the student; p. 220). Reviews focusing on specific regions or combinations of countries have also usefully provided syntheses of localised EER findings and foci, both in settings in which EER has had a relatively long tradition (see Chap. 10 of this volume by Opdenakker, 2020, for an example from Belgium and the Netherlands) and in regions where EER has been more recently adopted (e.g. see Yu, 2007, for an example on Sub-Saharan Africa).

While literature reviews can provide syntheses of the EER evidence in a narrative framework (e.g. the state-of-the-art reviews noted above), meta-analyses directly compare effect sizes across EER studies to generate generalisations including – but not limited to – aspects such as the relative size of school and teacher effects, the variations of these, and the associations between system, school and teacher variables with student outcomes (e.g. Scheerens & Bosker, 1997). It is worth noting that the development of a multilevel approach to meta-analysis (Raudenbush & Bryk, 1985) has contributed to the provision of more robust syntheses of the EER evidence base, just as multilevel modelling constituted an important advance in primary EER methodology. However and as noted above for literature reviews, not all meta-analyses have focused on or necessarily even acknowledged the locations within which each included study was conducted (e.g. Hattie, 2008; Seidel & Shavelson, 2007), but there have been several that consider country of origin. For example, a fairly recent meta-analysis by Scheerens, Witziers, and Steen (2013) did this explicitly via a country moderator variable included in their multilevel analysis, although partially (the countries included the USA, the Netherlands, and "other countries"; p. 627), and found some significant differences between countries in the effects of

variables including parental involvement (with higher effect sizes in the USA than elsewhere) and curriculum quality (with lower effect sizes in the Netherlands than elsewhere). Furthermore, Kyriakides, Christoforou, and Charalambous (2013), who used meta-analysis as a means of validating factors at the teacher level within the dynamic model of educational effectiveness, similarly included a country/area variable (including the USA, Europe, Asia, and other countries) though they found that this did not significantly predict effect sizes. This was interpreted to suggest that the teaching factors of the dynamic model seem to be generic, at least across the countries/areas included in that particular analysis. Common across both these examples, though, is the mixture of countries with groups of countries. This speaks to the disproportional weighting of existing EER research across countries and the continuing need to develop the international perspective in EER. In particular, there is scope for further research to investigate associations between effect sizes and finer-grained country variables – across a wider range of countries, perhaps – within a multilevel meta-analytic approach.

Of course, there are some inherent challenges and limitations from synthesising empirical findings, be this via narrative review or statistical meta-analysis. The methods used in individual studies vary, often making it difficult to untangle with any certainty that differences in findings might have resulted from underlying differences across countries rather than from differences in methodology. Further, depending on the approach taken to synthesising evidence across multiple studies, there is a risk of overlooking the importance of context. Avoiding this requires attention to the challenges encountered, methodological approaches, and limitations of the studies considered, as well as "effects" at each level (e.g. school, classroom, pupil) and of individual variables, to make sense of any differences across contexts.

2.6 The Contribution of International Studies Within and Beyond EER: Evidence on Cross-Country Comparisons

Despite the above noted dominance of single-country studies within the EER literature, comparative international studies focusing on student outcomes – particularly educational achievement – in order to evaluate the equity and quality of education systems have existed for approximately same period (about half a century) as have studies of school effectiveness (see Sect. 2.2). International large-scale assessments (ILSAs) began in the 1960s, with a study of mathematics achievement in 12 countries led by the International Association for the Evaluation of Educational Achievement (IEA; Foshay, Thorndike, Hotyaat, Pidgeon, & Walker, 1962).

In the five decades that followed the first IEA 12-country study, ILSAs have gained increasing prominence though why this is the case is likely due to multiple factors. Plausible reasons include the international policy trend towards globalisation, ILSA's contributions to that trend, or a combination of the two. Either way, the

number of countries participating in ILSAs has increased over time, from 12 countries in the 1962 IEA mathematics study to 80 jurisdictions in the 2018 administration of the Programme for International Assessment (PISA; OECD, 2018). Results of country comparisons have consequently drawn considerable attention from policymakers. There is considerable evidence to suggest that ILSAs and associated surveys have informed policy changes in a number of countries, particularly with regard to curriculum reform and performance standards (Lietz & Tobin, 2016), although in some cases reforms using ILSA results as a rationale have been based on misinterpretations or policy-maker cherry-picking (see Klieme, 2020, in Chap. 7 of this volume). The Trends in International Mathematics and Science Study (TIMSS; first conducted in 1995) and Progress in International Reading Literacy Study (PIRLS; first administered in 2001) conducted by the IEA, and the PISA (conducted every 3 years since 2000) and Teaching and Learning International Survey (TALIS; first administered in 2008) conducted by the Organisation for Economic Co-operation and Development (OECD) have been particularly influential, with country league table rankings widely publicised in the mass media.

Critics of these studies have pointed out a number of methodological and practical limitations. Reynolds et al. (2020) summarise these in detail in Chap. 6 of this volume – though noting that not all of these limitations apply to each ILSA – pointing to challenges in designing assessments with cross-cultural validity, sampling issues, often cross-sectional design, limited measurement of student background variables, limited measurement of teaching and learning processes, heavy emphasis on resource-based school factors, and limited measurement of non-academic student outcomes. Some of these limitations have been addressed by more recent developments; for example, the most recent PISA studies have adopted conceptual frameworks drawing on EER with input from established EER researchers (Klieme & Kuger, 2014; OECD, 2009), and a pilot of direct observation of video-recorded mathematics lessons to extend TALIS is currently underway at the time of writing this chapter (OECD, 2017) involving eight countries/jurisdictions including China (Shanghai), Spain (Madrid), Chile, Colombia, Japan, Mexico, England, and Germany.

While ILSAs have been in some ways quite distinct from EER, and the educational effectiveness literature tended for some time not to refer to ILSA literature (Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002), in some respects there has been reciprocal influence and synergy between the two. As noted above, EER researchers (e.g. Eckhard Klieme and Jaap Scheerens) have acted as experts and advisors for the OECD in order to develop and extend conceptual and analytic frameworks, and have featured in teams conducting OECD country reviews that have drawn on EER and education improvement perspectives (see OECD, 2013, p. 663 for the list of country review team members; see Mortimore, Field, & Pont, 2004; Nusche, Laveault, MacBeath, & Santiago, 2011; Santiago, Gilmore, Nusche, & Sammons, 2012; Shewbridge, Ehren, Santiago, & Tamassia, 2012, for examples of thematic reviews and reports). Conversely, EER researchers have also used ILSA data for secondary analyses to extend the EER knowledge base and to develop and test EER theory, dating back to re-analysis of IEA data to investigate the

generalisability of effectiveness factors across countries (see Scheerens, Vermeulen, & Pelgrum, 1989, for an early example of this). More recent secondary analyses of ILSA data have provided new types of insights that are particularly pertinent to EER. For example, a secondary analysis of TIMMS data (e.g. Kyriakides, 2006b) indicated a smaller proportion of variance between students, suggesting larger higher-level effects than generally identified in national studies, and secondary analyses of PISA data have provided insight into the functioning of specific factors across countries (e.g. Ning, Van Damme, Van Den Noortgate, Yang, & Gielen, 2015, on classroom disciplinary climate) and introduced country-level effectiveness measures for assessing contextualised effectiveness rather than simply identifying high-performing systems (e.g. Lenkeit & Caro, 2014), to name a few.

Beyond the re-analysis of data from ILSAs, researchers situated within EER have also drawn attention to the importance of primary comparative international studies within the field in order to extend existing theory and understanding of the impact of culture and context on schools, teachers and students. Some of the key reasons suggested (e.g. by Kyriakides, 2006a; Reynolds, 2000) for why international EER studies have an important contribution to make have included:

- Provision of an evidence base for (or against) the trend towards policy transfer across countries
- Correct and thorough identification of the power of school and classroom variation and effects of school and classroom variables beyond what is possible within single-country studies
- Generation of more sensitive theoretical explanations both to inform research design and to inform interpretation and understanding of results

Reynolds (2000) has further specified the need to study the relationships between processes, outcomes, and contexts in order to understand how different instructional variables relate to student outcomes in different contexts, and also how different school and education system features may foster similarly effective classroom practices in different contexts.

In response to this perceived need for comparative international studies within EER, one of the earliest examples of such work was the International School Effectiveness Research Project (ISERP; Reynolds, 2006; Reynolds et al., 2002). The aims of this study were to investigate "which factors are associated with student academic and social outcomes across countries and which factors are restricted to certain cultural context", as well as "which factors are associated with student academic and social outcomes across countries, for students with different characteristics" (i.e. differential effectiveness within and across countries; Reynolds, 2006, p. 539). The mixed-method research design was innovative and comprehensive, including observations of classroom and school processes, affective and social outcome measures, and a longitudinal cohort design. However, the study also highlighted some of the fundamental challenges in conducting such comparative international studies. Cultural differences in the discourse surrounding educational effectiveness, together with the above-mentioned dominance of certain countries within the EER knowledge base, meant that ensuring the cultural relevance of

educational factors was a challenge. Similarly, social outcomes were found to be more culturally-specific than academic ones, and therefore more difficult to measure and compare across regions and cultures; cross-cultural high-inference judgments of teaching practice were similarly found to be problematic. Nonetheless, the challenges and limitations of ISERP are arguably as (or more) useful towards an international perspective in EER as are the actual findings of the study; the detailed and transparent accounts of the research process (e.g. in Reynolds et al., 2002) have the potential to guide future comparative international EER efforts in addressing context specificity and cultural differences.

Within teacher effectiveness, the above-mentioned International System for Teacher Observation and Feedback (ISTOF; Muijs et al., 2018; Teddlie et al., 2006) was an ambitious effort to develop and validate an instrument for classroom observation across 20 participating countries (with at least some representation of regions including Europe, East Asia, South Asia, Southeast Asia, Africa, South America, and North America). The ISTOF instrument, developed using a modified Delphi technique drawing interactively on expert opinion and review, has since been validated and used in additional settings beyond the countries that contributed to its development (e.g. Soderlund, Sorlie, & Syse, 2015) as well as used for additional comparative studies (see Miao, Reynolds, Harris, & Jones, 2015, for an example comparing teaching in England and China).

Comparative international research within the area of school improvement has been characterised as having a more international character earlier in its history than was the case for educational effectiveness (Reynolds, 2000), and has also fed into the designs of studies at the intersection of EER and school improvement. The International School Improvement Project (ISIP; Bollen & Hopkins, 1987) provides an example of an early school-improvement focussed project on an international scale (with 14 participating countries), sponsored by the OECD though notably in contrast to the ILSA paradigm. Although Hopkins (1990) proposed that ISIP's focus on improvement strategies, goal-setting and processes marked a clear contrast with the effective schools knowledge base at the time, the detailed contextual information gleaned from the study's in-depth approach had the potential to inform more context-sensitive methods and measures in EER. Another example, the International Successful School Principalship Project (ISSPP), which built upon an approach used by Day and colleagues (2000) in the UK, was initiated in 2001 with researchers from seven countries and grew to include 14 countries by 2010. This study explored the qualities, behaviours and contexts of successful school principals, with a broad conceptualisation of success based on multiple indicators including but not limited to evidence of student achievement. The ISSPP constituted a large-scale, case-study based investigation of the role of school leaders in school improvement, and generated insight within and across contexts (e.g. Crow, Day, & Møller, 2017; Gurr, 2014; Johnson, Møller, Jacobson, & Wong, 2008; also see Chap. 12 in this volume by Townsend, Berryman, Gurr, & Drysdale, 2020); findings also provided information about the relationship between principal leadership and student learning (e.g. Day, Gu, & Sammons, 2016), which has significant implications for EER. The Effective School Improvement project (Creemers, 2002; Creemers, Reynolds &

Swint, 1996) comprised another important example of a comparative international study that advanced an international perspective, with an explicit link between the case study approach used and a set of guiding principles drawn from the existing EER knowledge base.

Replications of EER studies and improvement projects have been transferred across settings, too. For example, Professional Learning Communities based on previous research in the UK (Bolam, McMahon, Stoll, Thomas, & Wallace, 2005) are being implemented in China as part of the broader Improving Teacher Development and Educational Quality in China (ITDEQC) project (Thomas et al., 2016). On one hand, such examples of the uptake of, and action based on, findings from EER and school improvement initiatives from one setting to another are promising in that they have involved input from researchers based in both the original and the new contexts, allowing for local insight to inform implementation. On the other hand, the long-term success of such initiatives merits further investigation.

More recent examples of comparative international efforts in EER have included a mixture of different approaches, each with unique contributions to make towards the EER knowledge base. In addition to the projects mentioned above that included countries from different regions of the globe, regional multi-country studies help to shed light on within-region variation and inform the design of future studies (see Kelcey & Shen, 2016, for an example investigating school and teacher effects across 15 Sub-Saharan African countries). Two-country comparative studies have fewer practical complications to implement, yet still contribute insight into cultural and contextual differences in educational factors and student outcomes (see van de Grift, Chun, Maulana, Lee, & Helms-Lorenz, 2017, for an example within teacher effectiveness comparing across South Korea and the Netherlands).

Despite the developments within and beyond EER towards an international perspective via comparative international studies highlighted in this section, further work is needed to more fully understand how national or system-level policies and cultural factors affect schools, teaching and students (see Chap. 3 of this volume by Kyriakides, Creemers, & Panayiotou, 2020) and thus support or hinder their effectiveness in promoting different student outcomes. Lessons learned from the successes and also from the limitations of past studies, as well as across the areas of EER and ILSA, have the potential to inform increasingly rigorous and comprehensive future comparative international research in EER. This in turn has the potential to inform policy and practice across the system, school and classroom levels in multiple contexts.

2.7 Reflections and Suggestions for Future Directions

There is no question that EER has become an increasingly international field since the earliest studies conducted in the 1970s. Recent EER studies in middle- and low-income countries have begun to demonstrate the use of more sophisticated and up-to-date research methods and models (both theoretical and statistical), and a

wider variety of settings have been considered to contribute to the international evidence base. International comparative studies have become more common, perhaps at least in part because of increased synergy between EER and ILSA, and researchers seem to have generally become more outward-looking in their reviews of the literature to contextualise single-country empirical studies.

EER researchers have also begun to employ or at least call for a greater diversity of approaches to research design and analysis. The increasing convergence and mutual learning between educational effectiveness and school improvement research has afforded opportunities to better account for and attend to context. Mixed methods studies (e.g. ISERP) have demonstrated how the combination of qualitative and quantitative approaches can provide more thorough and robust insight into contextual understandings of concepts and educational processes. New statistical approaches to investigating educational effects (e.g. Sammons & Luyten, 2009) and consideration of small-scale experimental studies to test innovations in particular school contexts (Muijs et al., 2014) have the potential to contribute to the international knowledge base while maintaining attention to context specificity and cultural features. If the purpose of EER is ultimately to inform not only more nuanced understandings of educational quality and equity (Kyriakides, Creemers, & Charalambous, 2018) but also to support the development of more effective and equitable education (Chapman et al., 2012), such insight is essential.

However, there are still a number of issues that we suggest are in need of further investigation with regard to international perspectives in EER. Critical changes in national and international politics have taken place within the last decade, and the implications of these for the international landscape of EER, not to mention broader consequences for education in general and conceptualisations of what and for whom schooling is meant to function, have yet to be seen. We live in an increasingly globalised world, but many localised problems persist while others emerge and pose unique challenges (e.g. in response to natural disasters and violent conflict). We know a great deal about educational effectiveness, but much still remains to be learned, and many problems remain to be solved.

Reflecting on the above narrative account of international perspectives in the field, then, and on the political and policy contexts in which this narrative is contextualised, future work is needed to develop and assess the state of the field with regard to international perspectives in four areas:

1. Previous reviews of the literature have drawn attention to themes and developments from research in "developing" or "emerging" contexts, but a comprehensive, up-to-date systematic review of the international literature in EER is due.
2. Long-term research is needed to provide robust evidence of the uptake and transfer of educational innovations and initiatives between diverse contexts, in order to inform deeper understandings of why, how and where between-country or between-culture practices and policies "travel" (or conversely fail to "travel").
3. More research is still needed to develop a meaningful account that explains how concepts and educational factors are perceived and operationalised in different

contexts, something Reynolds et al. (2016a) have raised an "a matter for urgent attention" (p. 412).

4. On-going capacity building remains an essential consideration in places where EER research has generally been conducted by external researchers, to ensure that EER is informed by appropriate local knowledge and understanding.

International perspectives in EER can inform, and should be informed by, critical debates "about what education is for, about how it should be organized and – indeed – about the kind of society in which we wish to live" (to echo the words of Mortimore, 2013, p. 236). Further, we agree that education is 'under siege' and that educational research – including EER – plays a critical role within a democratic discourse on education policy,

> ...Research findings frequently challenge the expectations of policymakers. This discord, awkward though it may be, is essential in a democracy. The 1950s Quaker adage applies aptly to educational research when it states the need "to speak truth to power". (Mortimore, 2010)

Managing productive relationships and facilitating dialogue across stakeholder groups (policymakers, practitioners and researchers) is an important part of that democratic discourse (see Chapman, Ainscow, & Hadfield, 2020).

Ultimately, the development of an internationally diverse body of EER knowledge can advance our understanding of education (from the system-level to the level of the classroom), it can inform evidence-based and appropriately contextualised educational policy, and it can inform meaningful educational improvements and reform. To help inform progress towards these goals, this chapter has provided a historic overview of the development of this body of knowledge to date. This helps signpost future directions of EER and it sets the scene for the other chapters in this volume.

References

Ainscow, M., Dyson, A., Goldrick, S., & West, M. (2012). *Developing equitable education systems*. London, UK: Routledge.

Bambanota, G., Van Damme, J., De Fraine, B., Masimango, P. V., Mabela, G. K., Tshite, A. M., ... Legono, J. P. B. (2020). Educational effectiveness research in Africa: The case of the Democratic Republic of the Congo (DRC). In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). *Creating and sustaining professional learning communities*. Bristol, UK: University of Bristol.

Bollen, R., & Hopkins, D. (1987). *School based research: Towards a praxis*. Leuven, Belgium: ACCO Publishing.

Brookover, W. B., Schweitzer, J. H., Schneider, J. M., Beady, C. H., Flood, P. K., & Wisenbaker, J. M. (1978). Elementary school social climate and school achievement. *American Educational Research Journal, 15*(2), 301–318.

Caldas, S. J., & Bankston, C. L. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research, 93*(2), 91–100. https://doi.org/10.1080/00220679909597633

Chapman, C., Ainscow, M., & Hadfield, M. (2020). *Changing education systems: A research based approach*. London, UK: Routledge.

Chapman, C., Armstrong, P., Harris, A., Muijs, D., Reynolds, D., & Sammons, P. (2012). *School effectiveness and improvement research, policy, and practice challenging the orthodoxy?* London, UK/New York, NY: Routledge.

Chapman, C., & Muijs, D. (2014). Does school-to-school collaboration promote school improvement? A study of the impact of school federations on student outcomes. *School Effectiveness and School Improvement, 25*(3), 351–393.

Chapman, C., Reynolds, D., Muijs, D., Sammons, P., Stringfield, S., & Teddlie, C. (2016). Educational effectiveness and improvement research and practice: The emergence of the discipline. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement* (pp. 1–24). London, UK: Routledge.

Charalambous, C. Y., & Praetorius, A.-K. (2018). Studying instructional quality in mathematics through different lenses: In search of common ground. *ZDM: The International Journal on Mathematics Education, 50*(3), 355–366.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: US Department of Health, Education, and Welfare.

Creemers, B. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B. (1999). Educational effectiveness and improvement in developing societies: Some experiences from the primary education quality improvement project in Indonesia. *Tertium Comparationis, 5*(1), 32–51.

Creemers, B. (2002). From school effectiveness and school improvement to effective school improvement: Background, theoretical analysis, and outline of the empirical study. *Educational Research and Evaluation, 8*(4), 343–362. https://doi.org/10.1076/edre.8.4.343.8814

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.

Creemers, B., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. Milton Park, UK/Abingdon, UK/New York, NY: Routledge.

Creemers, B., Peters, T., & Reynolds, D. (1989). *School effectiveness and school improvement*. Amsterdam, The Netherlands: Swets & Zeitlinger.

Creemers, B., Reynolds, D., & Swint, F. E. (1996). The International School Effectiveness Research Project quantitative and class study data. Washington, DC: US Department of Education, Office of Educational Research and Improvement

Crow, G., Day, C., & Møller, J. (2017). Framing research on school principals' identities. *International Journal of Leadership in Education, 20*(3), 265–277. https://doi.org/10.1080/13603124.2015.1123299

Day, C., Gu, Q., & Sammons, P. (2016). The impact of leadership on student outcomes: How successful school leaders use transformational and instructional strategies to make a difference. *Educational Administration Quarterly, 52*(2), 221–258. https://doi.org/10.1177/0013161X15616863

Day, C., Harris, A., Hadfield, M., Tolley, H., & Beresford, J. (2000). *Leading schools in times of change*. Maidenhead, UK: Open University Press.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education, 5*(1), 163–198. https://doi.org/10.3102/0091732X005001163

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–27.

Education Endowment Foundation. (2018). *Sutton trust-EEF teaching and learning toolkit & EEF early years toolkit: Technical appendix and process manual*. London, UK: Education Endowment Foundation.

Fleisch, B. (2007). History of the school effectiveness and improvement movement in Africa. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 341–350). Dordrecht, The Netherlands: Springer.

Foshay, A. W., Thorndike, R. L., Hotyaat, F., Pidgeon, D. A., & Walker, D. A. (Eds.). (1962). *Educational achievements of thirteen-year-olds in twelve countries*. Hamburg, Germany: UNESCO Institute for Education.

Gage, N. L. (1963). *Handbook of research on teaching* (N. L. Gage, Ed.). Chicago, IL: Rand McNally.

Gurr, D. (2014). Successful school leadership across contexts and cultures. *Leading and Managing, 20*(2), 75–88.

Hall, J., Lindorff, A., & Sammons, P. (2016). *Evaluation of the impact and implementation of inspire maths in year 1 classrooms in England*. Oxford, UK: University of Oxford. https://doi.org/10.13140/RG.2.2.23042.07369

Harris, A., & Chrispeels, J. (2006). Improving schools and educational systems : International perspectives. In *London*. New York, NY: Routledge.

Harrison, J., & Kuint, S. (1998). School culture and school effectiveness in emergencies: Lessons from Israeli experience during the Gulf war. *School Effectiveness and School Improvement, 9*(2), 192–217. https://doi.org/10.1080/0924345980090204

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge. https://doi.org/10.4324/9780203887332

Hopkins, D. (1990). The international school improvement project (ISIP) and effective schooling: Towards a synthesis. *School Organisation, 10*(2–3), 179–194. https://doi.org/10.1080/0260136900100204

Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: A narrative state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 257–281. https://doi.org/10.1080/09243453.2014.885452

Jaciw, A. P., Hegseth, W. M., Lin, L., Toby, M., Newman, D., Ma, B., & Zacamy, J. (2016). Assessing impacts of math in focus, a "Singapore Math" program. *Journal of Research on Educational Effectiveness, 9*(4), 473–502. https://doi.org/10.1080/19345747.2016.1164777

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., ... Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York, NY: Basic Books.

Jerrim, J., & Vignoles, A. (2015). *The causal effect of East Asian "mastery" teaching methods on English children's mathematical skills?* (No. 15–05). London, UK: University College London.

Johnson, L., Møller, J., Jacobson, S. L., & Wong, K. C. (2008). Cross-national comparisons in the international successful school principalship project (ISSPP): The USA, Norway and China. *Scandinavian Journal of Educational Research, 52*(4), 407–422. https://doi.org/10.1080/00313830802184582

Kelcey, B., & Shen, Z. (2016). Multilevel design of school effectiveness studies in sub-Saharan Africa. *School Effectiveness and School Improvement, 27*(4), 492–510. https://doi.org/10.1080/09243453.2016.1168855

Klieme, E. (2020). Policies and practices of assessment: A showcase for the use (and misuse) of international large scale assessments in educational effectiveness research. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Klieme, E., & Kuger, S. (2014). *PISA 2015 draft questionnaire framework*. Paris: OECD. Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA-2015-draft-questionnaire-framework.pdf

Kratz, H. E. (1896). Characteristics of the best teacher as recognized by children. *The Pedagogical Seminary, 3*(3), 413–460. https://doi.org/10.1080/08919402.1896.9943604

Kyriakides, L. (2006a). Introduction international studies on educational effectiveness. *Educational Research and Evaluation, 12*(6), 489–497. https://doi.org/10.1080/13803610600873960

Kyriakides, L. (2006b). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. *Educational Research and Evaluation, 12*(6), 513–534. https://doi.org/10.1080/13803610600873986

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152. https://doi.org/10.1016/j.tate.2013.07.010

Kyriakides, L., Creemers, B., & Charalambous, E. (2018). *Equity and quality dimensions in educational effectiveness*. Cham, Switzerland: Springer.

Kyriakides, L., Creemers, B., & Panayiotou, A. (2020). Developing and testing theories of educational effectiveness addressing the dynamic nature of education. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Lenkeit, J., & Caro, D. H. (2014). Performance status and change – Measuring education system effectiveness with data from PISA 2000–2009. *Educational Research and Evaluation, 20*(2), 146–174. https://doi.org/10.1080/13803611.2014.891462

Lietz, P., & Tobin, M. (2016). The impact of large-scale assessments in education on education policy: Evidence from around the world. *Research Papers in Education, 31*(5), 499–501. https://doi.org/10.1080/02671522.2016.1225918

Lindorff, A. (2016). *Networks for school support and improvement: A mixed methods study*. Oxford, UK: University of Oxford.

Lindorff, A., Hall, J., & Sammons, P. (2019). Investigating a Singapore-based mathematics textbook and teaching approach in classrooms in England. *Frontiers in Education, 4*(37). https://doi.org/10.3389/feduc.2019.00037

Miao, Z., Reynolds, D., Harris, A., & Jones, M. (2015). Comparing performance: A cross-national investigation into the teaching of mathematics in primary classrooms in England and China. *Asia Pacific Journal of Education, 35*(3), 392–403. https://doi.org/10.1080/02188791.2015.1056593

Mortimore, P. (1998). *The road to improvement: Reflections on school effectiveness*. Abington, UK: Taylor & Francis.

Mortimore, P. (2001). Globalisation, effectiveness and improvement. *School Effectiveness and School Improvement, 12*(2), 229–249. https://doi.org/10.1076/sesi.12.2.229.3453

Mortimore, P. (2010, June 1). The education ministers must base their policies on sound research. *The Guardian*. Retrieved from https://www.theguardian.com/education/2010/jun/01/education-research-sound-policy-base

Mortimore, P. (2013). *Education under siege: Why there is a better alternative*. Bristol, UK: Policy Press.

Mortimore, P., Field, S., & Pont, B. (2004). *Equity in education: Thematic review*. Paris, France: OECD.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Somerset, UK: Open Books.

Muijs. (2020). Extending educational effectiveness: The middle tier and network effectiveness. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Muijs, D., Ainscow, M., Chapman, C., & West, M. (2011). *Collaboration and networking in education*. Dordrecht, The Netherlands: Springer.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – Teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

Muijs, D., Reynolds, D., Sammons, P., Kyriakides, L., Creemers, B. P. M., & Teddlie, C. (2018). Assessing individual lessons using a generic teacher observation instrument: How useful is the

international system for teacher observation and feedback (ISTOF)? *ZDM, 50*(3), 395–406. https://doi.org/10.1007/s11858-018-0921-9

Murillo, J. (2007). School effectiveness research in Latin America. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 75–91). Dordrecht, The Netherlands: Springer.

Ning, B., Van Damme, J., Van Den Noortgate, W., Yang, X., & Gielen, S. (2015). The influence of classroom disciplinary climate of schools on reading achievement: A cross-country comparative study. *School Effectiveness and School Improvement, 26*(4), 586–611. https://doi.org/10.1080/09243453.2015.1025796

Nusche, D., Laveault, D., MacBeath, J., & Santiago, P. (2011). *OECD reviews of evaluation and assessment in education: New Zealand 2011*. Paris, France: OECD.

OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris, France: OECD. Retrieved from https://www.oecd.org/pisa/pisaproducts/44455820.pdf

OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris, France: OECD.

OECD. (2017). *TALIS 2018: Video study and global video library on teaching practices*. Retrieved October 30, 2018, from http://www.oecd.org/education/school/TALIS-2018-video-study-brochure-ENG.pdf

OECD. (2018). About – PISA. Retrieved March 1, 2019, from https://www.oecd.org/pisa/aboutpisa/pisa-2018-participants.htm

Opdenakker, M.-C. (2020). Three decades of educational effectiveness research in Belgium and the Netherlands: Key studies, main research topics and findings. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Ortega, L., Malmberg, L.-E., & Sammons, P. (2018a). School effects on Chilean children's achievement growth in language and mathematics: An accelerated growth curve model. *School Effectiveness and School Improvement, 29*(2), 308–337. https://doi.org/10.1080/09243453.2018.1443945

Ortega, L., Malmberg, L.-E., & Sammons, P. (2018b). Teacher effects on Chilean children's achievement growth: A cross-classified multiple membership accelerated growth curve model. *Educational Evaluation and Policy Analysis, 40*(3), 473–501. https://doi.org/10.3102/0162373718781960

Plowden, B., & Central Advisory Council for England. (1967). *Children and their primary schools*. London, UK: Her Majesty's Stationery Office.

Raudenbush, S. W., & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics, 10*(2), 75–98. https://doi.org/10.3102/10769986010002075

Reynolds, D. (1996). Introduction and overview. *School Effectiveness and School Improvement, 7*(2), 111–113. https://doi.org/10.1080/0924345960070201

Reynolds, D. (2000). School effectiveness: The international dimension. In C. Teddlie & D. Reynolds (Eds.), *International handbook of school effectiveness research* (pp. 232–256). London, UK: Falmer.

Reynolds, D. (2006). World class schools: Some methodological and substantive findings and implications of the International School Effectiveness Research Project (ISERP). *Educational Research and Evaluation, 12*(6), 535–560. https://doi.org/10.1080/13803610600874026

Reynolds, D., Caldwell, B., Cruz, R. M., Miao, Z., Murillo, J., Mugendawala, H., ... Ramon, M. R. R. (2016a). Comparative educational research. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement* (pp. 246–282). London, UK: Routledge.

Reynolds, D., Chapman, C., Clarke, P., Muijs, D., Sammons, P., & Teddlie, C. (2016b). The future of educational effectiveness and improvement research, and some suggestions and speculations. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge*

international handbook of educational effectiveness and improvement (pp. 408–439). London, UK: Routledge.

Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: International perspectives on school effectiveness*. London, UK: Routledge.

Reynolds, D., Kelly, A., Harris, A., Jones, M., Adams, D., Miao, Z., & Bokhove, C. (2020). Extending educational effectiveness: A critical review of research approaches in international effectiveness research, and proposals to improve them. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197–230. https://doi.org/10.1080/09243453.2014.885450

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.

Sammons, P., Anders, Y., & Hall, J. (2013). Educational effectiveness approaches in early childhood research across Europe. *School Effectiveness and School Improvement, 24*(2), 131–137. https://doi.org/10.1080/09243453.2012.749790

Sammons, P., Davis, S., & Gray, J. (2016). Methodological properties of school effectiveness research: Exploring the underpinnings, evolution and future directions of the field. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *Routledge international handbook of educational effectiveness and improvement* (pp. 25–76). London, UK: Routledge.

Sammons, P., & Luyten, H. (2009). Editorial article for special issue on alternative methods for assessing school effects and schooling effects. *School Effectiveness and School Improvement, 20*(2), 133–143. https://doi.org/10.1080/09243450902879753

Santiago, P., Gilmore, A., Nusche, D., & Sammons, P. (2012). *OECD reviews of evaluation and assessment in education: Czech Republic 2012*. Paris, France: OECD.

Scheerens, J. (2001). Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement, 12*(4), 359–384. https://doi.org/10.1076/sesi.12.4.359.3447

Scheerens, J. (2015). Theories on educational effectiveness and ineffectiveness. *School Effectiveness and School Improvement, 26*(1), 10–31. https://doi.org/10.1080/09243453.2013.858754

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness a critical review of the knowledge base*. Dordrecht: Springer. Retrieved from https://doi.org/10.1007/978-94-017-7459-8.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK/New York, NY: Pergamon.

Scheerens, J., Vermeulen, C. J. A. J., & Pelgrum, W. J. (1989). Generalizibility of instructional and school effectiveness indicators across nations. *International Journal of Educational Research, 13*(7), 789–799. https://doi.org/10.1016/0883-0355(89)90029-3

Scheerens, J., Witziers, B., & Steen, R. (2013). A meta-analysis of school effectiveness studies. *Revista de Educacion, 361*, 619–645.

Schildkamp, K. (2019). *President welcome*. Retrieved March 15, 2019, from https://www.icsei.net/president-welcome/

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Shewbridge, C., Ehren, M., Santiago, P., & Tamassia, C. (2012). *OECD reviews of evaluation and assessment in education: Luxembourg 2012*. Paris, France: OECD.

Society for Research on Educational Effectiveness. (2019). *Membership*. Retrieved from https://www.sree.org/membership

Soderlund, G., Sorlie, K., & Syse, I. (2015). Mestringsforventninger i matematikk. In *Finnut,* Lilyhammer.

Stevens, R. (1912). *The question as a measure of efficiency in instruction*. New York, NY: Teachers College, Columbia University. Retrieved from https://archive.org/details/questionasmeasur00stevuoft

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs 1. *Educational Research and Evaluation, 12*(6), 561–582.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *International handbook of school effectiveness research*. London, UK: Falmer.

Thomas, S., Kyriakides, L., & Townsend, T. (2016). Educational effectiveness in new, emerging, and traditional contexts. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement* (pp. 220–245). London, UK: Routledge.

Townsend, T. (2007). 20 years of ICSEI: The impact of school effectiveness and school improvement on school reform. In T. Townsend (Ed.), *Internationalhandbook of school effectiveness and school improvement* (pp. 3–26). Dordrecht, The Netherlands: Springer.

Townsend, T., Berryman, M., Gurr, D., & Drysdale, L. (2020). Leadership for learning in diverse settings: School leaders setting the agenda in Australia and New Zealand. In *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Townsend, T., Clarke, P., & Ainscow, M. (1999). *Third millennium schools: A world of difference in effectiveness and improvement*. Lisse, The Netherlands/Exton, PA: Swets & Zeitlinger.

Tymms, P., Merrell, C., Heron, T., Jones, P., Albone, S., & Henderson, B. (2008). The importance of districts. *School Effectiveness and School Improvement, 19*(3), 261–274. https://doi.org/10.1080/09243450802332069

Van Damme, J., Liu, H., Vanhee, L., & Pustjens, H. (2010). Longitudinal studies at the country level as a new approach to educational effectiveness: Explaining change in reading achievement (PIRLS) by change in age, socio-economic status and class size. *Effective Education, 2*(1), 53–84. https://doi.org/10.1080/19415531003616888

van de Grift, W. J. C. M., Chun, S., Maulana, R., Lee, O., & Helms-Lorenz, M. (2017). Measuring teaching quality and student engagement in South Korea and The Netherlands. *School Effectiveness and School Improvement, 28*(3), 337–349. https://doi.org/10.1080/09243453.2016.1263215

Weber, G. (1971). *Inner city children can be taught to read: Four successful schools*. Washington, DC: Council for Basic Education.

What Works Clearinghouse. (2017). *What works clearinghouse standards handbook (Version 4.0)*. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Yu, G. (2007). *Research evidence of school effectiveness in sub-Saharan Africa*. Bristol, UK: EdQual Research Programme Consortium.

Chapter 3
Developing and Testing Theories of Educational Effectiveness Addressing the Dynamic Nature of Education

Leonidas Kyriakides, Bert Creemers, and Anastasia Panayiotou

3.1 Introduction

Educational Effectiveness Research (EER) was originally concerned with the question of what works in education and why (Creemers, Kyriakides, & Sammons, 2010; Reynolds et al., 2014; Sammons, Davis, & Gray, 2016; Scheerens & Bosker, 1997). However, early effectiveness studies were only searching for factors associated with student achievement without attempting to explain why these factors could be seen as important for learning. As a consequence, a list of various effective characteristics (correlated with student achievement) was established during the first phase of EER. One could therefore argue that EER should not only identify factors associated with student achievement but also explain why these factors influence learning in order to establish relevant theoretical models. There are several reasons to argue for the need to develop and test models of educational effectiveness that could help us explain differences in student learning results by specifying the relationships between the components in the models and student outcomes. First, a model serves to explain previous empirical research parsimoniously. Second, the establishment and testing of models of educational effectiveness could help us generate a guide to the field to prevent new entrants from re-inventing the wheel by repeating existing research. It also maps a series of avenues for future research, which may help us expand the knowledge base of educational effectiveness. Finally, a model may provide a useful road map for practitioners (as first proposed by Creemers et al., 2010). Indeed, there

L. Kyriakides (✉) · A. Panayiotou
Department of Education, University of Cyprus, Nicosia, Cyprus
e-mail: kyriakid@ucy.ac.cy; panayiotou.anastasia@ucy.ac.cy

B. Creemers
Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands
e-mail: b.p.m.creemers@rug.nl

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_3

are hints that it has been partially the absence of educational effectiveness theory that has hindered the uptake of effectiveness knowledge by practitioners in schools (Creemers & Kyriakides, 2015).

In this chapter, we argue that EER should not only refer to factors associated with students' learning outcomes but also explain why, under which conditions, and for whom these factors promote learning. Specifically, research on differential effectiveness reveals the importance of searching for the extent to which a factor is more relevant for a specific group of students. As a consequence, the agenda of EER becomes broader and both the quality and equity dimensions of effectiveness can be addressed (see Kyriakides, Creemers, & Charalambous, 2018a). Thus, the first part of this chapter shows that during the early phases of EER, researchers were searching for the impact of individual factors expecting that these factors are generic in nature and can affect learning of all students in most contexts. The situational character of factors has gradually been examined and the recent theories presented in the second part of the chapter take into account the dynamic nature of effectiveness and can contribute to the establishment of stronger links between EER and improvement of practice. The third part of the chapter provides an overview of national and international studies conducted in order to test the validity of the dynamic model using empirical data and conducted in a range of country contexts. These empirical studies as well as relevant meta-analyses seem to provide support for the importance of specific factors included in the dynamic theory of educational effectiveness and their measurement dimensions. Empirical studies also reveal relations among factors operating at the classroom level which help us define stages of effective teaching. On the other hand, empirical studies reveal possibilities of establishing a more parsimonious model. Thus, in the last part of this chapter, we provide suggestions on how to improve further the dynamic model by taking into account that schools should not only search for improving student learning outcomes but also reducing differences in initial achievement of students with different socio-economic background to enhance equity as well as quality.

3.1.1 Educational Effectiveness Theories: Moving from Single Approaches to Integrated Models

In the literature of educational effectiveness modelling, three basic approaches have been used during the last three decades. First, the economic approach is focused on estimating the relationship between the "supply of selected purchased schooling inputs and educational outcomes controlling for the influence of various background features" (Monk, 1992, p. 308). Such research is focused on producing a function which could explain each pupil outcome at a given time. The function may be linear, consisting of main effects and interaction terms or non-linear (Brown & Saks, 1986). The emerging "education production" models (e.g., Brown & Saks, 1986; Elberts & Stone, 1988) are based on the assumption that increased inputs will lead to

increments in outcomes. In this context, researchers from the field of economics of education attempted not only to design their own studies but also to make use of secondary data analyses based on international evaluation studies such as PISA and TIMSS and test the validity of "education production" models by searching for the impact of input variables (especially resources) on final achievement at country level by using more sophisticated techniques to allow them to search for cause and effect relations (Creemers et al., 2010). However, the research conducted using these models revealed that the relation between input and outcomes is more complex than was assumed. For example, earlier studies from Hanushek and Hedges (e.g., Hanushek, 1986, 1989; Hedges, Laine, & Greenwald, 1994) show that reducing student – teacher ratio and/or increasing the amount of funding in education per student, does not necessarily result in higher student outcomes. Despite the use of more sophisticated techniques in analyzing data of international studies similar results seem to emerge.

The second approach to educational effectiveness modelling is similar to the economic approach but is focused on a different choice of antecedent conditions, since it is mainly focused on variables at the student level that are assumed to predict student outcomes. Some attention is also paid to school and teaching processes from two different perspectives concerning learning and schools as organizations. Within this approach, educational psychologists focused on student background factors such as "learning aptitudes", "personality", and "motivation", and on variables measuring the learning processes which take place in classrooms. As a consequence, different theories on effective teaching have been proposed (Creemers & Kyriakides, 2008; Klieme, Pauli, & Reusser, 2009; Muijs et al., 2014; Pianta & Hamre, 2009) which refer to both generic and domain specific teaching practices and different attempts were made by researchers to develop internationally valid instruments to assist the measurement of teacher effectiveness. Such an attempt was made during the International System for Teacher Observation and Feedback (ISTOF) project. This project aimed to produce an instrument for classroom observation which could not only be used for collecting quantitative, psychometrically reliable, and valid data across a variety of countries for cross-country comparisons, but also for providing meaningful feedback to teachers (Teddlie, Creemers, Kyriakides, Muijs, & Yu, 2006). The focus of ISTOF was therefore to develop a model of effective teaching that could be used in different countries to measure classroom factors associated with student learning. On the other hand, this sociological perspective is focused on factors that define the educational background of students such as SES, gender, social-capital, and peer group. This perspective does not only examine student outcomes, but also the extent to which schools manage to reduce the variance in student outcomes compared to prior achievement. Thus, two dimensions of measuring school effectiveness emerged from this perspective concerning quality and equity (Creemers & Kyriakides, 2008). Moreover, the sociological perspective raises attention for process variables that have emerged from organizational theories such as the school climate, culture and structure, and for contextual variables such as school size and school population including student average prior achievement and

socio-economic background (Hallinger & Heck, 1998; Opdenakker & Van Damme, 2006).

Finally, the models of the third approach emerged by researchers' attempt to integrate the findings of School Effectiveness Research (SER), Teacher Effectiveness Research (TER), and the early input-output studies. Thus, the models of this approach which were developed during the 1990s (e.g., Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) have a multilevel structure, where schools are nested in contexts, classrooms are nested in schools, and students are nested in classrooms or teachers. Although these models make use of both organizational theories and theories of learning and refer to multiple factors operating at different levels, each model is either focused on the classroom or school level. Depending on this, more emphasis is given either to theories of learning (e.g., Creemers, 1994) or to organizational theories (e.g., Scheerens, 1992).

In this chapter we also argue that most of the studies on educational effectiveness which were conducted during this period were atheoretical and were mainly concerned with the establishment of statistical relationships between variables rather than with the generation and testing of theories which could explain those relationships. During the first decade of the twenty-first century, one could however identify six national studies investigating the validity of the comprehensive model of educational effectiveness (Creemers, 1994) which was considered as one of the most influential theoretical constructs in the field (Teddlie & Reynolds, 2000). Specifically, these six studies investigated the main aspects of Creemers' model (i.e., De Jong, Westerhof, & Kruiter, 2004; Driessen & Sleegers, 2000; Kyriakides, 2005a; Kyriakides & Tsangaridou, 2008; Kyriakides, Campbell, & Gagatsis, 2000; Reezigt, Guldemond, & Creemers, 1999) in two European countries with more and less centralized educational systems (i.e., The Netherlands and Cyprus). A comparison of the results of these studies helps us identify the extent to which the model can be used to explain effectiveness in both centralized and less centralized educational systems (see Kyriakides, 2008). First, the six studies mentioned above demonstrate that the influences on student achievement are multilevel. Classrooms were found to have unique effects on student learning, independently of factors operating at the school and individual level. Moreover, by controlling for both student factors and classroom contextual factors, variables at the school level were also found to explain some variation in achievement at the school level (Kyriakides, 2008). This finding is in line with the findings of most studies on educational effectiveness conducted in various countries (Muijs et al., 2014; Reynolds et al., 2014; Teddlie & Reynolds, 2000) and provides support to the argument that models of EER should be multilevel in nature (Creemers et al., 2010).

Second, these six studies revealed that most of the student-level factors included in Creemers' model, such as aptitude, social background, and motivation, showed effects in the expected directions. However, the need to expand the model at the student level has also been pointed out since most of the unexplained variance was found to be located at this level. The importance of looking within studies in the field of psychology to identify student-level factors has been stressed (Kyriakides, 2005a). Nevertheless, researchers within the field of EER should be critical about

the extension of the current models and select only variables that have stable effects and can help us establish a model that is in line with the parsimony principle. In this context, searching for student level factors that may interact with factors operating at classroom and/or school level may be more important for establishing improvement strategies at school and classroom level. For example, one of the six studies mentioned above attempted to identify ways of expanding the comprehensive model (Kyriakides, 2005a). It was found that a factor concerned with the thinking styles of students which emerged from the theory of mental self-government (Sternberg, 1988) should be considered. Given that thinking styles are seen as dynamic, it can be claimed that it is possible to design projects attempting to help students develop "optimal" styles in order to improve their achievement. It can also be argued that high-quality teaching may help students develop optimal thinking styles. This argument reveals the importance of including this student-level factor in the models of educational effectiveness. Moreover, this study suggests that there is an interaction between measures of quality of teaching and measures of the personal characteristics of students. For example, generic teaching skills, found to be consistently correlated with student achievement, were found to have a general effect across all students but to a different degree with students of different thinking styles and personality traits. One could therefore claim that the analyses of the results of studies testing the validity of Creemers' model revealed that, next to the multilevel nature of effectiveness, the relationship between factors at different levels might be more complex than assumed in the integrated models. This may be especially important to tap into possible interaction effects among factors operating at classroom and student level which reveal the importance of investigating differentiated effectiveness (Campbell, Kyriakides, Muijs, & Robinson, 2004).

Third, in each of these six studies, the concept of quality of teaching was treated in a different way since each study was searching for the impact of different aspects of quality of teaching upon student achievement. This can be attributed to the fact that, although the comprehensive model gives more emphasis to the processes of teaching than the other integrated models (e.g., Scheerens, 1992; Stringfield & Slavin, 1992), the concept of quality of teaching is not defined precisely (Kyriakides, 2008). A comparison of the aspects of teaching taken into account by the six studies also reveals that the aspects of quality of teaching considered mainly refer to the direct and active teaching approach (Joyce, Weil, & Calhoun, 2000). However, in recent years, constructivists and others who support the "new learning" approach (e.g., Choi & Hannafin, 1995; Savery & Duffy, 1995; Simons, Van der Linden, & Duffy, 2000; Vermunt & Verschaffel, 2000) have developed a set of instructional techniques that are supposed to enhance the learning disposition of students such as modelling, coaching, scaffolding and fading, articulating, reflection, exploration, generalization, provision of anchors, goal orientation, and self-regulated learning. One could therefore claim that researchers attempting to develop models which illustrate the complexity of educational effectiveness at the classroom level should refer not only to skills associated with direct teaching and mastery learning but also to factors which are in line with new theories of learning. In the next section, it is

explained that the dynamic model of educational effectiveness took into account these findings in defining the quality of teaching (see Creemers & Kyriakides, 2008).

Fourth, in contrast with the classroom-level factors, not a huge variation in the aspects of school factors taken into account by these six studies can be identified. However, different measurement frameworks were used in order to define similar aspects of school-level factors. For example, one study investigated the frequency dimension of school evaluation policy to identify the effect of this factor on achievement and revealed both negative and positive effects, whereas another study was looking at the emphasis given to the formative aspect of evaluation and revealed positive effects. A similar problem can also be identified in the way the classroom-level factors are measured by these six studies. The reason that these six studies made use of different approaches to measure effectiveness factors can be attributed to the fact that Creemers' model (as well as all the other integrated models of educational effectiveness) does not explicitly refer to the measurement of each effectiveness factor. On the contrary, it is often assumed that these factors represent unidimensional constructs. The synthesis of studies testing Creemers' model revealed that the models of educational effectiveness should not only refer to factors operating at different levels, but also identify the dimensions upon which each factor can be measured. Considering effectiveness factors as multidimensional constructs provides a better picture of what makes teachers and schools effective and may help us develop specific strategies for improving educational practice (see Kyriakides & Creemers, 2008). In this context, the dynamic model, which is presented in the next part of this chapter, attempts to introduce a specific framework for measuring the functioning of factors operating at classroom, school and system level.

Finally, some support for the concept of differential effectiveness was provided by the studies testing Creemers' model. However, the importance of considering differentiation in developing the theoretical framework of EER does not only arise from the results of the studies testing Creemers' model, but also from recognition that students of any age and in any culture will differ from one another in various intellectual and psychomotor skills, in both generalized and specialized prior knowledge, in interests and motives, in their socioeconomic backgrounds, and in personal styles of thoughts and work during learning (Dowson & McInerney, 2003). Researchers in the area of educational effectiveness have shown that these differences are in turn related to differences in students' learning progress (e.g., Gustafsson, Nilsen, & Hansen, 2018; Kyriakides, 2005a; Kyriakides & Luyten, 2009; Lim, 2013; Sammons, Toth, & Sylva, 2018; Slavin, 1987; Teddlie & Reynolds, 2000). These relations imply individual predispositions that somehow condition student readiness to profit from the particular instructional environments provided. Despite the fact that educational practice has remained basically fixed and non-adaptive in most countries, research into differential effectiveness seems to reveal that teachers and schools may be differentially effective in promoting the learning of different groups of students (Campbell et al., 2004). Thus, generic models of educational effectiveness which are able to incorporate the results of research into differential teacher and school effectiveness should be developed. In

this way theory-driven and evidence-based school improvement strategies aiming to promote not only quality but also equity may emerge (see Kyriakides et al., 2018a).

3.1.2 Concluding Comments

In the first part of this chapter, we refer to three main approaches to educational effectiveness and identify their strengths and limitations by taking into account findings of studies searching for educational effectiveness factors coming from each of these approaches. The importance of establishing integrated and multilevel models is also acknowledged. This part also refers to studies testing the validity of one of the most influential integrated models; namely the comprehensive model of educational effectiveness. A synthesis of six studies investigating the validity of this model provided support to the main assumption of the integrated models, since it was demonstrated that factors associated with student learning operate at different levels. Direct and indirect relations between the levels and the outcomes are also identified. One could therefore argue that the findings of the studies conducted in order to test the validity of Creemers' model provide support to the importance of establishing a multilevel integrated model of educational effectiveness, such as Creemers' model. It has also been argued that the results of these studies reveal four weaknesses of the comprehensive model which have been taken into account for the establishment of the dynamic model, presented in the next section. First, we need a clear definition of quality of teaching which will take into account both the direct and active teaching approach, as well as the constructivist approach. Second, a framework for measuring the functioning of factors seems to be missing from the integrated models. As a consequence, most effectiveness studies consider factors as unidimensional and thereby the complex nature of educational effectiveness is not fully addressed. Considering effectiveness factors as multidimensional constructs not only provides a better picture of what makes teachers and schools effective, but also helps us develop specific strategies for improving educational practice. Third, the importance of treating differentiation as a separate dimension of measuring each effectiveness factor is also recognized. In this way, it can be acknowledged that the impact of effectiveness factors on different groups of students/teachers/schools may vary. Finally, the importance of investigating relations among factors operating at different levels is stressed. The next section refers to the dynamic model of educational effectiveness which was developed by taking into account the above limitations of the integrated models. This model was also developed in order to establish stronger links between EER and improvement of practice (see Creemers & Kyriakides, 2006). It should however be acknowledged that integrated models were not only influential in establishing the theoretical framework of EER, but also in developing educational effectiveness policies in various countries around the world. Their contribution in understanding the complex nature of educational effectiveness is pointed out. It is, however, acknowledged that these theories were

not in a position to establish links between research on educational effectiveness and school improvement efforts.

3.2 Theories of Educational Effectiveness Addressing the Dynamic Nature of Education

By considering the limitations of the integrated models of educational effectiveness, a theory that takes into account the dynamic nature of education was proposed (Creemers & Kyriakides, 2008). The rationale and major assumptions of this theory are presented in this part of the chapter. First, the dynamic model is multilevel in nature and refers to factors operating at the four levels shown in Fig. 3.1. Thus, the teaching and learning situation is emphasized and the roles of the two main actors (i.e. teacher and student) are analyzed. Above these two levels, the dynamic model also refers to school-level factors. It is expected that school-level factors influence the teaching and learning situation by developing and evaluating the school policy on teaching and the policy on creating a learning environment at the school. The system level refers to the influence of the educational system through more formal avenues, especially through the development and evaluation of educational policy at the national and/or regional level. The model also takes into account the fact that the teaching and learning situation is influenced by the wider educational context in which students, teachers and schools are expected to operate. For example, factors such as the societal values for learning and the level of social and political importance attached to education play important roles both in shaping teacher and student expectations, as well as in the opinion formation of various stakeholders about what constitutes effective teaching practices.

Second, the dynamic model suggests that factors at the school and system level have both direct and indirect effects on student achievement since they are able to influence not only student achievement but also the teaching and learning situations. Third, the model assumes that the impact of the school and system-level factors has to be defined and measured in a different way than the impact of classroom-level factors. Policy on teaching and actions taken to improve teaching practice must be measured over time and in relation to any identified weaknesses that may occur in a school. The assumption is that schools and educational systems, which are able to identify their weaknesses and develop a policy on aspects associated with teaching and their learning environment, are also able to improve the functioning of classroom-level factors and their effectiveness status. Only changes in those factors for which schools face significant problems are expected to be associated with the improvement of school effectiveness measured by improved student outcomes (Creemers & Kyriakides, 2012). This implies that the impact of school- and system-level factors depends on the current situation of the objects under investigation (i.e., students, teachers, schools, systems). Fourth, the model assumes that there is a need to carefully examine the relationships between the various effectiveness factors

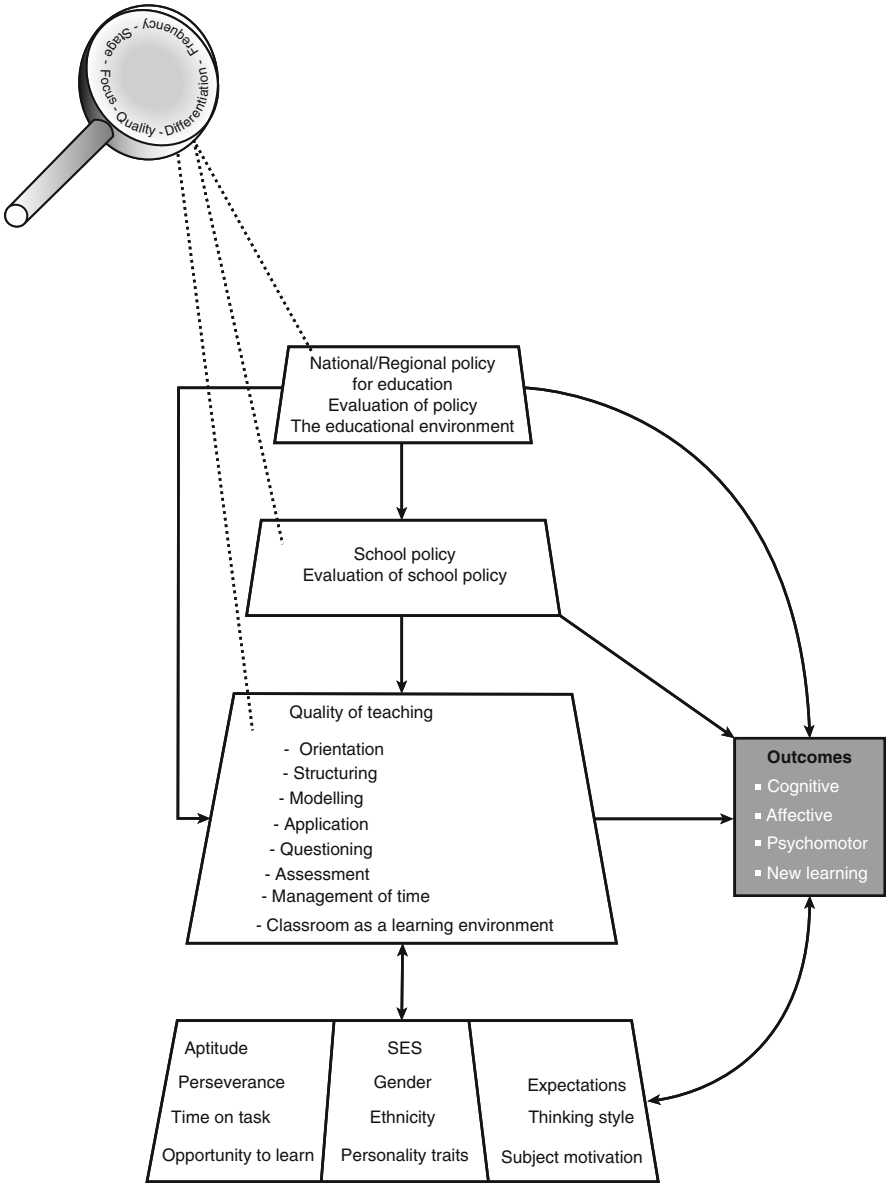

Fig. 3.1 The dynamic model of educational effectiveness

which operate at the same level. Walberg's (1984) model, which is one of the most significant educational productivity models, attempts to illustrate such relationships. Aptitude, instruction, and the psychological environment are seen as major direct causes of student learning. They also influence one another and are in turn influenced by feedback on the amount of learning that takes place. The Walberg model was

tested as a structural equation model on science achievement, indicating more complex, indirect relationships (Reynolds & Walberg, 1990). This implies that there is a need to refer to the relationships between the effectiveness factors which operate at the same level. Such an approach to modelling educational effectiveness may reveal grouping of factors that make teachers and schools more or less effective. Therefore, strategies for improving effectiveness which are comprehensive in nature may emerge.

Finally, the dynamic model is based on the assumption that each factor can be defined and measured by using five dimensions: frequency, focus, stage, quality, and differentiation. This can be considered as one of the main differences of the dynamic model from all the existing theoretical models in EER since the dynamic model attempts to show that effectiveness factors are multidimensional constructs and can be measured in relation to specific dimensions. More specifically, *frequency* is a quantitative way to measure the functioning of each effectiveness factor and most effectiveness studies made use of this dimension to measure factors and their relation in achievement, whereas the other four dimensions examine qualitative characteristics of the functioning of each effectiveness factor at the system/school/classroom level. Using this measurement framework implies that each factor should not only be examined by measuring how frequently the factor is present in a system/school/class (i.e., through a quantitative perspective) but also by investigating specific aspects of the way the factor is functioning (i.e., looking at qualitative characteristics of the functioning of the factor). We briefly describe below the importance of taking into account these five dimensions in measuring the functioning of each factor. The *frequency* dimension refers to the extent, in quantitative terms, to which an activity associated with an effectiveness factor is present in a system, school or classroom. This is probably the easiest way to measure the effect of a factor on student achievement.

The factors are also measured by taking into account the focus of the activities associated with a factor. For example, in the case of school policy on parental involvement, the policy could either be specific in terms of concrete activities that are expected to take place (e.g., it refers to specific hours that parents can visit the school) or more general (e.g., it informs parents that their presence is welcome in the school but without giving them specific information about what, how and when). Moreover, an activity may be expected to achieve a single or multiple purposes. In the case of school policy on parental involvement, the activities might be restricted to a single purpose (e.g., parents visit the school to get information about student progress). On the other hand, the activities might aim to fulfil more than one purpose (e.g., parents visit the school to exchange information about children's progress and to assist teachers within and outside the classroom). A balance between specific and general tasks should exist. For example, guidelines on parental involvement, which are very general, may not be helpful either for parents or teachers in establishing good relations to support student learning. On the other hand, a school policy which is very specific in defining activities may restrict the productive involvement of teachers and parents, preventing them from creating their own ways of implementing the school policy. Similarly, if all the activities are expected to achieve a single

purpose, then the likelihood of success is high, but the effect of the factor might be small due to the fact that other purposes are not achieved and synergy may not exist (Scheerens, 2013; Slater & Teddlie, 1992). On the other hand, if all the activities are expected to achieve multiple purposes, there is a danger that specific purposes are not addressed in such a way that they can be implemented successfully.

In addition, the activities associated with a factor can be measured by taking into account the *stage* at which they take place. We know from other research that the factors need to take place over a suitable period of time to ensure that they have a *continuous direct or indirect effect* on student learning (Creemers, 1994; Slater & Teddlie, 1992). For example, school policy on student absenteeism is expected to be implemented throughout the year and not only through specific regulations announced at a specific point in time (e.g., only at the beginning of the school year). It is also expected that continuity will be achieved when the school is flexible about redefining its own policy and adapting the activities related to the factor by taking into account the results of its own self-evaluation mechanism (Kyriakides & Campbell, 2004; Visscher & Coe, 2002).

The *quality* dimension can be determined in two different ways. The first one refers to the properties of the specific factor itself, as these are discussed in the literature. For instance, school policy on assessment can be measured by looking at the mechanisms which have been developed in order to establish instruments which meet psychometric standards (e.g., valid, reliable, representative of the content taught). At the same time, the impact of a factor on improving practice is taken into account. In the example given above, the policy on assessment is expected to ensure that teachers use information gathered from assessment in order to meet their students' needs. In this way, the school policy is expected to place more emphasis on the formative rather than the summative function of assessment (Christoforidou, Kyriakides, Antoniou, & Creemers, 2014).

Finally, although the dynamic model is expected to be a generic model, it takes into account the findings of research into differential educational effectiveness (Campbell, Kyriakides, Muijs, & Robinson, 2003). Specifically, effectiveness factors are seen as generic in nature, but it is also acknowledged that their impact on different groups of students/teachers/schools may vary. As a consequence, differentiation is treated as a measurement dimension and is concerned with the extent to which activities associated with a factor are implemented in the same way for all the actors involved with it at each relevant level (e.g., all the students, teachers, schools). It is expected that adaptation to the specific needs of each actor or group of actors will increase the successful implementation of a factor and ultimately maximize its potential effect on student learning outcomes. Although differentiation could be considered a property of an effectiveness factor, it was decided to treat differentiation as a separate dimension of measuring each effectiveness factor rather than incorporate it into the quality dimension. In this way, the importance of taking into account the special needs of each actor or group of actors is recognized. It is finally important to note that the dynamic model is based on the assumption that it is difficult to deny that persons of all ages learn, think, and process information differently. One way to differentiate instruction is for teachers to teach according

to individual student learning needs as these are defined by their background and personal characteristics such as gender, socioeconomic status (SES), ability, thinking style, and personality type (Kyriakides, 2007). For example, effective teachers provide more active instruction and feedback, more redundancy, and smaller steps with a higher success rate to their low-achieving students (Brophy, 1986). Warmth and support, in addition to good instruction, is provided to low-achieving students, who are more frequently encouraged for their efforts (Muijs, Campbell, Kyriakides, & Robinson, 2005). A similar argument can be made in relation to the way teachers should be treated by their school leaders. For example, instructional leadership should not be seen as equally important for all the teachers of a school. Effective principals are expected to adapt their leadership to the specific needs of the teachers by taking into account the extent to which they are ready to implement a task (Hersey & Blanchard, 1993). Similarly, policy-makers are expected to adapt their general policy to the specific needs of groups of schools and encourage teachers to differentiate their instruction. Research into differential educational effectiveness reveals that teachers' objectives, as well as organizational and cultural factors, should be taken into account when the dimension of differentiation is measured (Dowson & McInerney, 2003; Hayes & Deyhle, 2001). However, the differentiation dimension does not imply that the actors are not expected to achieve the same purposes. On the contrary, adapting the policy to the special needs of each group of schools/teachers/students may ensure that all of them will become able to achieve the same purposes. This argument is partly supported by research into adaptive teaching and the evaluation projects of innovations concerned with the use of adaptive teaching in classrooms (e.g., Houtveen, Van de Grift, & Creemers, 2004; Noble, 2004; Reusser, 2000). Therefore, policy-makers should make explicit to teachers what they are expected to achieve through differentiating their instruction and through responding to the different needs of their students. This is particularly crucial for establishing an effective policy on equal opportunities since research has shown that some existing educational practices are maladaptive (e.g., Kyriakides, 2004; Peterson, Wilkinson, & Hallinan, 1984). Therefore, the differentiation dimension helps policy-makers not only establish a policy on equal opportunities but also provide support to the schools where teaching practice is maladaptive and help them act in such a way that differentiation of instruction does not result in holding lower achievers back and increasing individual differences (Kyriakides, 2007).

3.2.1 Student-Level Factors Included in the Dynamic Model of Educational Effectiveness

Both the sociological and the psychological perspective of EER are taken into account in defining student level factors. Figure 3.2 refers to the student-level factors included in the dynamic model and reveals that this model classifies student factors into three categories: (a) sociocultural and economic background variables emerging

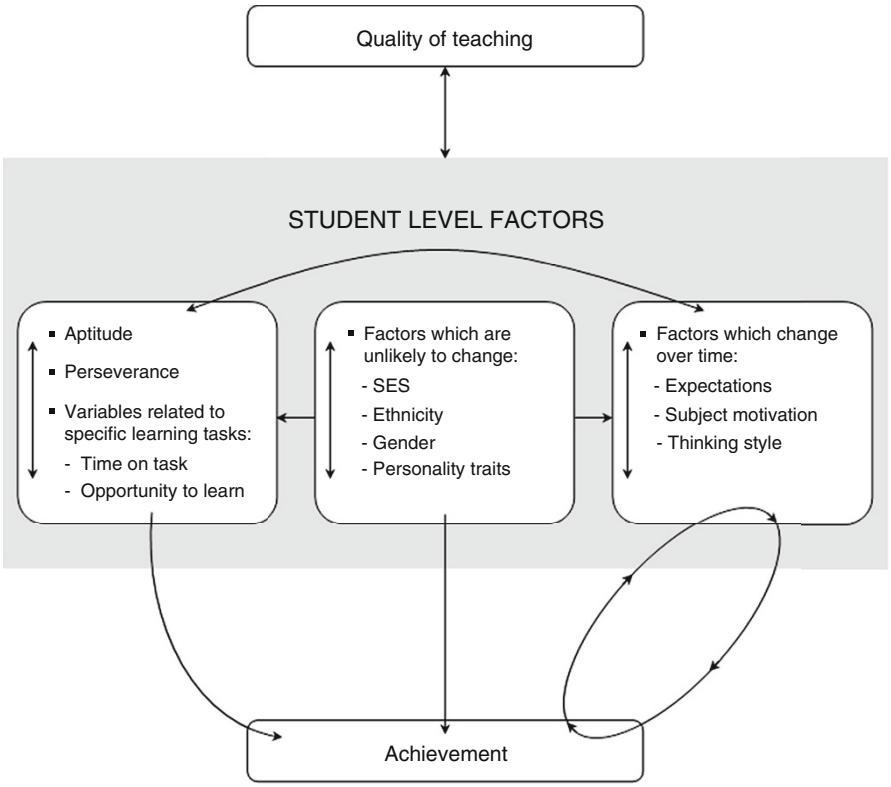

Fig. 3.2 Factors of the dynamic model operating at the student level

from the sociological perspective of EER, (b) background variables emerging from the psychological perspective of EER and (c) variables related to specific learning tasks associated with the learning outcomes used to measure effectiveness, such as prior achievement, time on task and opportunity to learn. Moreover, Fig. 3.2 shows that a distinction is made between the student-level factors, differentiating those factors which are unlikely to change (e.g., gender, SES, ethnicity, personality) from those that may be more likely to change over time (e.g., subject motivation, thinking styles). Factors that are unlikely to change are more closely related to the equity dimension of effectiveness, since teachers and other stakeholders are expected to reduce their impact on student learning outcomes in order to establish a fair class/school/educational system (see Kyriakides et al., 2018a).

3.2.1.1 Socio-Cultural and Economic Background Variables Emerging from the Sociological Perspective of EER

The first group of student-level factors in the dynamic model refers to sociocultural and economic background variables that emerge from the sociological perspective of EER, such as SES, ethnic background and gender. The treatment of these variables as student-level factors is strongly related to the history of EER and especially to the sociological perspective of educational effectiveness. Many studies have showed that the strongest predictors of student outcomes comprise student background characteristics, like SES, ethnicity and gender (Sirin, 2005). In this context, the integrated models of educational effectiveness (e.g., Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) have treated background variables as student-level factors. Moreover, coming from the history of research on inequality in education, it was clear that EER would look at the educational outcomes of disadvantaged children in particular and investigate equity in schools (Sammons, 1995, 1996; Sammons, Nuttall, & Cuttance, 1993; Sammons et al., 2013; Ortega, Malmberg, & Sammons, 2018).

Beyond indicating the importance of treating background variables as student-level factors and providing suggestions on how research into differential effectiveness could help teachers/schools/systems become more effective in terms of both quality and equity, the dynamic model also addresses the importance of looking at relations between these variables. For example, the majority of studies on equity in education have focused on the effects of either gender or SES on educational attainment, or on the effect of SES or ethnicity (Grant & Sleeter, 1986; Lenkeit, Caro, & Strand, 2015; Strand, 2011). Even in the few studies in which student background factors have been evaluated simultaneously (Anders et al., 2011; Hall et al., 2013; Ortega et al., 2018; Sammons, 1995, 1996; Sammons et al., 1993, 2008, 2013, 2018; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2010; Thomas, Sammons, Mortimore, & Smees, 1997), interactions between SES, ethnicity and gender have rarely been made explicit (Strand, 2012; Strand, 2014a, 2014b). Most studies examine the net effects of background factors on student outcomes and are not considering interactions between background factors. The Effective Pre-School and Primary Education 3–11 Project (EPPE3–11), comprises an example of such studies that simultaneously examined the effect of several student background factors (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004).The EPPE3–11 project foresaw a major longitudinal study of a national sample of young children in England and followed children's cognitive and social/behavioural development between the ages of 3 and 11. This project collected background data (i.e., child, family and home characteristics) from 3000 children coming from different backgrounds to examine their possible effects on their cognitive and social development at different ages. Namely, data on background factors such as birth weight, gender, parental qualification/occupations and the home learning environment were collected. The differentiated effect of background factors was also examined for students with different characteristics showing that different groups

of students portray different educational needs. Differential school effectiveness in relation to gender and social class was also examined in a study by Kyriakides (2004) where, it was investigated whether some primary schools are more or less effective for girls rather than boys, and for students from different social groups. The findings of this study demonstrated the complexity of schools and the need for further research into differential school effectiveness in order to identify cultural factors affecting their effectiveness (Kyriakides, Creemers, & Charalambous, 2018b). Further studies examining school characteristics moderating the relation between student SES and achievement are also needed to improve equity of educational outcomes (Gustafsson et al., 2018).

Thus, the dynamic model is defined at the student level in such a way that it raises the importance of a systematic investigation of differential school and teacher effectiveness in relation not only to SES, but also to the other student background characteristics, and especially gender and ethnicity. This argument is based on findings of studies in the field of equity in education. With regard to the impact of ethnicity, several national (e.g., Lehmann & Lenkeit, 2008; Ohinata & Van Ours, 2012; Strand, 2014a) and international studies (e.g., Mullis, Martin, Kennedy, & Foy, 2007; Retali, 2011) have shown that students with an immigrant background exhibit lower educational attainment than those without such a background (e.g., Dronkers & de Heus, 2013; Shapira, 2012). It is also important to note that only a part of this gap in attainment can be explained by students' family SES (Dustmann, Machin, & Schönberg, 2010; Wilson, Burgess, & Briggs, 2005). Nonetheless the situation is complicated in some countries such as England where white low SES (working class) boys are shown to have the poorest outcomes in the school system compared with other ethnic groups indicating the importance of considering interactions between different background factors especially gender, ethnicity and SES.

It is, however, important to acknowledge that at the level of the classroom, students should be treated as individuals rather than as representing stereotypical groupings so that the promotion of learning for all students is encouraged. Nevertheless, at the level of the school or the system, if groups of students are systematically being disadvantaged in terms of their rate of learning in comparison with other groups, as some effectiveness studies in different countries have shown (e.g., Beaton et al., 1996; Gorard, Rees, & Salisbury, 2001; Gray, Peng, Steward, & Thomas, 2004; Harskamp, 1988; Kyriakides, 2004; Reynolds et al., 2014; Sammons, 1995; Sammons et al., 1993, 2008, 2013; Thomas et al., 1997), stakeholders at the school and system level should develop more effective policies to promote greater equity in educational outcomes and to reduce equity gaps in achievement.

3.2.1.2 Background Variables that Emerged from the Psychological Perspective of EER

The dynamic model also refers to five background variables emerging from the psychological perspective of EER, which were found to be related to student achievement: aptitude, motivation, expectations, personality, and thinking style.

Aptitude is seen as one of the most critical background variables associated with student achievement. Aptitude embraces general intelligence and prior learning. Since the 1960s, this variable has been included in the early educational productivity models. Specifically, aptitude is considered to be the amount of time students need to learn in optimal instructional conditions (Creemers, 1994). This factor is taken into account in proposing methods to measure the effect of teachers and schools in promoting student learning outcomes (see Creemers et al., 2010).

Achievement is not only influenced by aptitude and social background but also by motivation (Brophy, 2013). Specifically, it is argued that motivation not only influences outcomes, but also that academic outcomes do have an effect on motivation (Marsh & Craven, 2006). Thus, the dynamic model treats motivation as an important student-level factor but not all dimensions of motivation have been included. The model gives emphasis to perseverance and subject motivation whereas other conceptions, like self-confidence and self-efficacy beliefs, have not been found to be related to achievement gains (e.g., De Jong et al., 2004; Kyriakides & Tsangaridou, 2008) and are not included in the model. At the same time, it is argued that teacher behavior in the classroom may influence subject motivation, either, positively or negatively (Baumert & Demmrich, 2001) and relevant interaction effects are included in the model.

The expectations that students believe that significant others (e.g., parents and friends) have of them is also taken into account. Such expectations could be perceived by students as a kind of external pressure that significant others may impose on them. A critical question is whether a linear or a non-linear relationship with student achievement exists, since after a certain point a negative relationship with achievement may develop (Creemers & Kyriakides, 2008). Given that there are individual differences with respect to prior achievement, policy-makers and practitioners should be aware that this factor implies that teachers should hold different types of expectations of each student. Moreover, the concept of 'expectations' should be seen as dynamic in nature. For example, as soon as a student makes progress, his/her expectations of him/herself may become higher. At the same time the demands of a series of lessons may induce different types of expectations in different students. It is therefore important to make sure that realistic expectations of and by each student are generated. As a consequence, the dynamic model treats this student-level factor as being strongly related to the classroom-level factors, and especially to self-regulation theories of learning. Treating expectations as a student-level factor is also in line with the new theories of learning which address the idea of self-regulation and contain information about learning processes and the content of learning processes (Flavell, 1979).

Finally, the dynamic model takes into account the results of more recent studies which claim that the early integrated models of educational effectiveness should be expanded to refer to students' personal characteristics, such as their personality traits and their thinking styles (Hartig, Klieme, & Leutner, 2008; Kyriakides, 2005b). These studies show that measures of both personality and thinking style can be treated as predictors of student achievement gains. This argument is also supported by the fact that studies within psychology reveal that types of personality and styles

of thinking are associated with student achievement (Bratko, Chamorro-Premuzic, & Saks, 2006; Demetriou, Kyriakides, & Avraamidou, 2003; Demetriou, Spanoudis, & Mouyi, 2011; Noftle & Robins, 2007).

3.2.1.3 Variables Related to Specific Learning Tasks Emerging from the Psychological Perspective of EER

The third category of variables included in the dynamic model comprises of the following two main student-level factors which are related to specific learning tasks and are briefly described below. First, the variable 'time on task' refers to the time students are willing to spend on learning and on educational tasks and is determined not only by motivation and expectations, but also by the time provided by the school/teacher and by processes taking place at the school and classroom levels. Thus, 'time on task' refers to the time in which students are really involved in learning, provided that this time is filled with opportunities to learn. The special attention paid to time on task stems directly from the early models of educational effectiveness (i.e., Carroll, 1963; Walberg, 1986), which attempted to relate time spent on learning with achievement. Second, the variable 'opportunity to learn' refers to the fact that in order to achieve educational outcomes, students should at least have some opportunity to acquire knowledge and skills. Lugthart, Roeders, Bosker, and Bos (1989) have presented an overview of studies looking at the relationship between opportunity to learn and student achievement. These studies reveal the importance of the variable 'opportunity to learn' in accounting for variations between schools and teachers. Variation increases when students are the unit of analysis since differences in curricula are not the only source of variation in opportunity to learn. Other overview studies and meta-analyses have confirmed the importance of opportunity to learn (Hattie, 2009; Hendriks, 2014) and in relation to educational policy.

3.2.2 Teacher Factors: An Integrated Approach to Effective Teaching Is Promoted

Based on the main findings of teacher effectiveness research (e.g., Brophy & Good, 1986; Fraser, Walberg, Welch, & Hattie, 1987; Muijs & Reynolds, 2001; Opdenakker & Van Damme, 2000; Rosenshine & Stevens, 1986), the dynamic model refers to factors which describe teachers' instructional role and are associated with student outcomes. These factors refer to observable instructional behavior of teachers in the classroom rather than to factors that may explain such behavior (e.g., teacher beliefs and knowledge and interpersonal competences). The eight factors included in the model are as follows: orientation, structuring, questioning, teaching modelling, application, management of time, teacher role in making classroom a learning environment, and classroom assessment. These eight factors do not refer

only to one approach of teaching such as structured or direct teaching (Joyce et al., 2000) or to approaches associated with constructivism (Schoenfeld, 1998). An integrated approach in defining quality of teaching is adopted. Specifically, the dynamic model does not refer only to skills associated with direct teaching and mastery learning such as structuring and questioning but also to orientation and teaching modelling which are in line with theories of teaching associated with constructivism (Brekelmans, Sleegers, & Fraser, 2000). Moreover, the collaboration technique is included under the overarching factor of teacher contribution to the establishment of the classroom learning environment. A short description of each factor follows.

A. Orientation: This factor refers to teacher behaviour in providing the students with opportunities to identify the reason(s) for which an activity or lesson or series of lessons occur and/or actively involving students to the identification of the reason(s) for which a lesson includes a specific task. Through this process it is expected that the activities that take place during lessons, become meaningful to students and consequently increase their motivation for participating actively in the classroom (e.g., De Corte, 2000; Paris & Paris, 2001). This factor may thus have an impact on increasing student motivation and through that, on increasing student learning outcomes.
B. Structuring: Student learning is positively influenced when teachers actively present materials and structure them by: (a) beginning with overviews and/or review of objectives; (b) outlining the content to be covered and signalling transitions between lesson parts; (c) calling attention to main ideas; and (d) reviewing main ideas at the end (Rosenshine & Stevens, 1986). Structuring activities aim at assisting students develop links between the different parts of lessons, instead of dealing with them in an isolated way (Creemers & Kyriakides, 2015).
C. Questioning: This factor is concerned with teacher ability in: (a) raising different types of questions (i.e., process and product) at appropriate difficulty level; (b) giving time for students to respond; and (c) dealing with student responses. Raising numerous questions in a lesson increases the active involvement of students in class discussion and promotes interactions, both with the teacher and among students. Questioning can also be used in order to assess students' understanding and help them clarify and verbalize their thinking in order to develop a sense of mastery (Muijs et al., 2014).
D. Teaching-modelling: Although there is a long tradition in research on teaching higher-order thinking skills and problem solving, these teaching and learning activities have received unprecedented attention during the last two decades, due to the policy emphasis on the achievement of new goals of education. Thus, the teaching-modeling factor is associated with findings of effectiveness studies revealing that effective teachers are expected to help pupils use strategies and/or develop their own strategies that can help them solve different types of problems (Grieve, 2010; Muijs et al., 2014). Consequently, students are expected to develop skills that help them organize their own learning (e.g.,

self-regulation and active learning). In defining this factor, the dynamic model also addresses the properties of teaching-modeling tasks, and the role that teachers are expected to play in order to help students devise problem-solving strategies. Teachers may either present students with a clear problem-solving strategy, or they may invite students to explain how they themselves would approach or resolve a particular problem and then use that information for promoting the idea of modeling. Recent research suggests that the latter approach may encourage students to not only use, but also develop their own problem-solving strategies (Aparicio & Moneo, 2005; Gijbels, Van de Watering, Dochy, & Van den Bossche, 2006).

E. Application: Providing students with practice and application opportunities can enhance learning outcomes. Learning new information cannot be a constant process, since according to the Cognitive Load Theory the working memory can only process a limited amount of information at each given time (Kirschner, 2002). Effective teachers may use seatwork or small-group tasks in order to provide necessary practice and application opportunities as starting points for the next step in teaching and learning.

F. The classroom as a learning environment: This factor as described in the dynamic model consists of five components which were shown to be the most important aspects of the classroom climate through studies and meta-analyses: (a) teacher-student interaction, (b) student-student interaction, (c) students' treatment by the teacher, (d) competition between students, and (e) classroom disorder. Classroom environment research has shown that the first two of these elements are important components of measuring classroom climate (for example, see Cazden, 1986; Den Brok, Brekelmans, & Wubbels, 2004; Harjunen, 2012). However, according to the dynamic model, what should be examined are the types of interactions that exist in a classroom, rather than how students perceive their teacher's interpersonal behavior. Specifically, the dynamic model is concerned with the immediate impact that teacher initiatives have on establishing relevant interactions in the classroom, and it investigates the extent to which teachers are able to establish on-task behavior through promotion of interactions. The other three elements refer to teachers' attempts to create an efficient and supportive environment for learning in the classroom (Walberg, 1986). These elements are measured by taking into account the teacher's behavior in establishing rules, persuading students to respect and use the rules, and the teacher's ability to maintain them in order to create and sustain an effective learning environment in the classroom.

G. Management of time: To address this factor the amount of time used per lesson for on-task behaviour is investigated. Teachers are expected to: (a) prioritize academic instruction and allocate available time to curriculum-related activities; and (b) maximize student engagement rates. Time management skills are not restricted solely to teachers' ability to avoid the loss of teaching time through minimizing external classroom disruptions, or through dealing effectively with organizational issues (e.g., moving between classes, organizing and distributing materials or giving instructions). Apart from the overall teaching time,

management of time skills also include teacher actions that increase the learning time for each individual student (i.e., the on-task time).

H. Assessment: Assessment is seen as an essential part of teaching (Stenmark, 1992). Especially formative assessment has been shown to be one of the most important factors associated with effectiveness at all levels, especially at the classroom level (Christoforidou et al., 2014). Effective teachers are therefore expected to: (a) Use appropriate techniques to collect data on student knowledge and skills; (b) analyse data in order to identify student needs; (c) report assessment results to students and parents; and (d) evaluate their own practices.

3.2.3 *School Factors: Promoting Quality and Equity by Taking Actions to Improve School Policy on Teaching and the Learning Environment*

The dynamic model is based on the assumption that factors at the school level influence classroom-level factors, particularly teaching practice. Since learning takes place both inside and outside the classroom, the model emphasises not only how to improve *teaching,* but also the *school learning environment (SLE).* As a consequence, the model refers to: (a) the school policy on teaching, and (b) the school policy on creating a learning environment at school. Based on the assumption that the essence of a successful organisation in the modern world is the search for improvement (Hopkins, 2001), the processes and the activities which take place in the school in order to improve the teaching practice and the SLE are also examined. For this reason, the processes which are used to evaluate the school policy for teaching and the SLE are investigated. Thus the following overarching factors at the school level are included in the model.

a) School policy on teaching and actions taken for improving teaching practice;
b) Policy on creating the SLE and actions taken for improving the SLE;
c) Evaluation of school policy on teaching and evaluation of the SLE.

Moreover, the model assumes that school policy has a situational effect on student achievement, implying that its impact may vary depending on the current situation of the school under investigation (Kyriakides, Creemers, Antoniou, Demetriou, & Charalambous, 2015). This situational character of school policy suggests that, in developing the school policy, school stakeholders should take into account the abilities and readiness of those who are expected to implement it. Three elements of school policy are therefore considered. Firstly, it is expected that school policy should clarify the role of all stakeholders in improving learning. When the school policy is clear, the stakeholders are more likely to consider its recommendations and decide whether it is worth making the effort to change their actions. Guidelines are seen as one of the main indications of school policy. In using the term ‘guidelines’, the dynamic model refers to a range of documents. These include staff meeting minutes, announcements and action plans which make the

policy of the school more concrete for school stakeholders. However, the factors concerned with the school policy do not imply that each school should simply develop formal documents to institute policy but mainly refer to the actions taken by the school to help teachers and other stakeholders have a clear understanding of what is expected of them. Secondly, the framework assumes that in introducing a school policy, the skills and the willingness of school stakeholders should be taken into account. If a certain policy expects stakeholders to undertake roles they do not have the skills to perform or to which they are strongly opposed, it is unlikely that the policy will be implemented effectively. This element of the dynamic model is taken into account in proposing the Dynamic Approach to School Improvement (DASI) which is based on the assumption that school stakeholders should develop their own strategies and action plans by adapting the knowledge-base of EER to the context of their schools (see the second part of this section). The third element of school policy is concerned with the support that the school management team should provide to stakeholders in order to help them change their actions. Introducing a policy on teaching and/or the SLE that addresses these three elements is likely to influence stakeholders' actions (Kyriakides et al., 2015). The elements of the school factors are presented below to clarify the concepts upon which school stakeholders' actions could be based.

3.2.3.1 School Policy on Teaching and Actions Taken to Improve Teaching

Meta-analyses of factors associated with student achievement show that concepts such as teaching quality, time on task and opportunity to learn are key factors that explain variation in student outcomes (Hattie, 2009; Scheerens, Seidel, Witziers, Hendriks, & Doornekamp, 2005; Kyriakides, Creemers, Antoniou, & Demetriou, 2010). In this context, the definition of the dynamic model at the school level refers to factors related to the key concepts of *quality*, *time on task*, and *opportunity to learn*. Therefore, the model attempts to investigate aspects of school policy on teaching associated with (a) the quantity of teaching, (b) provision of learning opportunities, and (c) quality of teaching. Actions taken to improve the above three aspects of teaching, such as the provision of support for teachers to improve their teaching skills, are also taken into account. Therefore, an examination of school policy on teaching reveals that effective schools take decisions about maximising the use of teaching time and the learning opportunities offered to their students. In addition, effective schools support their teachers in their attempt to help students learn by using effective teaching practices (Hallinger & Heck, 2011; Heck & Moriyama, 2010).

3.2.3.2 School Policy on Creating the SLE and Actions Taken to Improve the SLE

Over the past four decades, the work on the SLE has rapidly expanded to cover issues such as interpersonal relationships between the school personnel and the management team, as well as the support provided for students (e.g., Lüftenegger et al., 2012; Mainhard, Brekelmans, & Wubbels, 2011). From this array of elements, we focus here on policy initiatives only if they aim to improve stakeholders' learning, and through that student learning. This is accomplished by focusing on the following four aspects of school policy on improving the SLE, which have been systematically found to be associated with student learning outcomes:

1. Student behaviour outside the classroom;
2. Collaboration and interaction between teachers;
3. Partnership policy (i.e., relations of school with community, parents, and advisors);
4. Provision of sufficient learning resources for students and teachers.

The first three aspects refer to the practices which the school has developed to establish a learning environment inside and outside the classroom. It is important to note that the term learning does not refer exclusively to student learning. For example, collaboration and interaction between teachers may contribute to their professional development (i.e., learning of teachers) and may also have an effect on teaching practice, with the possibility of improving student learning outcomes (Goddard, Goddard, & Tschannen-Moran, 2007). Similarly, by involving parents in the functioning of schools and also providing them with opportunities for learning, the school facilitates learning on two fronts: through the classroom learning environment (e.g., when parents provide teachers with information regarding their children or bring human and other resources to the school) and the home learning environment (e.g., when parents are informed about how to support/supervise their children when doing their homework) (Fan & Chen, 2001; Kyriakides, 2005b). The fourth aspect refers to the policy on providing resources for learning. The availability of learning resources in schools may not only have an effect on student learning, but may also encourage the learning of teachers (Hattie, 2009). Actions taken to improve the SLE beyond the establishment of policy guidelines are also taken into account. Such actions can be directed at changing the school rules and providing educational resources (e.g., teaching aids and educational assistance).

3.2.3.3 Evaluation of School Policy on Teaching and the SLE

Finally, the dynamic model refers to the mechanisms used to evaluate school policy on teaching and the SLE. To measure the functioning of this factor, it is taken into account that previous research tells us that effective schools use various sources from which to collect evaluative data, and that these data are collected periodically during

the school year, not just at the beginning and the end of the school year (Kyriakides et al., 2010; Reynolds et al., 2014; Scheerens, 2013). Moreover, school evaluation mechanisms are expected to measure the properties of school policy (e.g., clear, concrete, in line with the research literature), its relevance to the problems which teachers and students have to face, and its impact on school practice and student outcomes. Furthermore, the quality of the evaluation mechanisms is examined by looking at the psychometric properties (i.e., reliability, validity and use) of the instruments schools use to collect data as well as to the extent to which the formative rather than the summative purpose of evaluation is served, especially since school evaluation is seen as closely related to the school improvement process.

3.2.4 System-Level Factors Included in the Dynamic Model of Educational Effectiveness

In this section, we refer to the most important factors operating at the system level (see Fig. 3.2) included in the dynamic model that may affect student achievement gains. The model does not refer to all of the characteristics of an educational system which reveal variations in the operation of the educational systems around the world. Specifically, it does not refer to the structure of the system, but to aspects of the national policy that affect learning inside and outside the classroom. This assumption is supported by the fact that international studies and meta-analyses of comparative studies reveal that the effectiveness of an educational system cannot be attributed to the fact that it is a centralised or a decentralised system, since neither a centralised nor a decentralised system can promote curriculum changes that may improve the effectiveness of the system (Fullan, 1991; Kyriakides & Charalambous, 2005; Schmidt, Jakwerth, & McKnight, 1998; Schmidt & Valverde, 1995). Moreover, the finding that the few effectiveness studies investigating the impact of middle-level factors (such as the effect of the LEAs) revealed that the impact of this level is very small (e.g., Tymms et al., 2006) can be attributed to the fact that in the countries where these studies were conducted, these levels are not in a position to directly influence school policy on teaching or the learning environment of the schools. For example, the curriculum is defined at the national level, and the schools which belong to different authorities were expected to use the same curriculum. Therefore, the dynamic model argues that those authorities who are only responsible for solving administrative problems that are faced at the school and/or the system level may not have any significant effect on student achievement gains.

Given that not many studies have been conducted in order to identify factors operating at the context level, in our attempt to define context-level factors we take into account the two major overarching factors operating at the school level which may directly affect: a) student learning through improving teaching practice (i.e., school policy for teaching) and b) learning which takes place outside the classroom and is addressed to all the school stakeholders (i.e., policy for the school learning

environment). As a consequence, a similar overarching factor at the national level is included in the dynamic model. This factor refers to the national educational policy in relation to the teaching practice and the learning environment of the school and is expected to directly affect teaching practice and the SLE or even indirectly by providing support to the schools to develop their own policies on teaching and their SLE. As in the case of the school level, actions taken for improving national policy in relation to teaching and the learning environment of the schools are also taken into account. Moreover, the term guidelines is used in a broader way to indicate all kind of documents sent to schools by the context level which try to make sure that teachers and other stakeholders understand the meaning of the national/regional policy and what they are expected to do. Finally, the evaluation mechanism of the national educational policy that may contribute to the improvement of the national policy and, through that, to the improvement of educational effectiveness is also treated as an overarching factor operating at the system level.

An essential difference of the factors operating at the system and the school level is concerned with the fact that not only policy on teaching and the learning environment is taken into account in the case of the context level, but also the wider environment of education. Specifically, the wider educational environment of a country and its ability to increase opportunities for learning and develop positive values for learning is considered as an important context level factor. This is because it is acknowledged that student learning is not expected to take place only inside schools, but also in the wider school community. However, instead of referring in a very general way to the context of education, the dynamic model concentrates on two aspects of the wider educational environment which are expected to influence learning. First, we examine the support provided to schools from different stakeholders (e.g., church, companies, universities, educational researchers, institutions responsible for providing support/advice/in-service training to schools). The model is not only concerned with the financial support which different stakeholders provide to schools. Support provided to schools may refer to strategies/advice offered to schools which may help them improve their teaching practice or establish better learning environments (e.g., help them establish better relations among teachers and/or between teachers and students; help them identify ways to treat student misbehaviour outside and inside the classroom; support their attempts to undertake extra-curricular activities that are related to the official aims of the curriculum). In regard to the second aspect of this overarching factor, it refers to the expectations of different stakeholders (e.g., employers, policy-makers, parents, and public) from schools about learning and learning outcomes. These expectations may result in achievement press and, through that, in student achievement gains (Valverde & Schmidt, 2000). This implies that the schools of most effective countries are driven by a quest for academic excellence.

3.3 Developing and Testing Theories of Educational Effectiveness: From a Synthesis of Effectiveness Studies to Improving the Dynamic Model

Some material supporting the validity of the dynamic model has been produced since 2003, when the model was first developed (Creemers & Kyriakides, 2015). Specifically, the model has received empirical support from national studies (e.g., Antoniou & Kyriakides, 2011, 2013; Azigwe, Kyriakides, Panayiotou, & Creemers, 2016; Azkiyah, Doolaard, Creemers, & Van Der Werf, 2014; Creemers & Kyriakides, 2010a; Kyriakides & Creemers, 2008, 2009; Kyriakides, Anthimou, & Panayiotou, 2020), international studies (e.g., Kyriakides, Archambault, & Janosz, 2013; Panayiotou et al., 2014), and two meta-analyses (quantitative syntheses) of studies investigating the impact of teacher and school factors (i.e., Kyriakides et al., 2010; Kyriakides, Christoforou, & Charalambous, 2013) as well as from empirical and theoretical reviews (see Heck & Moriyama, 2010; Hofman, Hofman, & Gray, 2010; Sammons, 2009; Scheerens, 2013). Table 3.1 refers to the studies and meta-analyses which have been carried out and the type of support that each assumption of the model has received from these studies. The following observations arise from this table.

Studies

1. A longitudinal study measuring teacher and school effectiveness in different subjects (Kyriakides & Creemers, 2008).
2. A study investigating the impact of teacher factors on achievement of Cypriot students at the end of pre-primary school (Kyriakides & Creemers, 2009).

Table 3.1 Empirical evidence supporting the main assumptions of the dynamic model emerging from empirical studies and meta-analyses

Assumptions of the dynamic model	Studies	Meta-analyses
1. Multilevel in nature	All	All
2. Five dimensions can be used to measure		
a) Teacher factors	1, 2, 4, 5, 7, 11, 12	2
b) School factors	1, 3, 4	1
3. Impact of teacher factors on learning outcomes	1, 2, 4, 5, 6, 7, 11, 12	2
4. Impact of school factors on learning outcomes	1, 3, 4, 6	1
5. Situational character of school factors	1	
6. Relationships between factors operating at the same level: Stages of effective teaching	1, 5, 6, 7, 8, 9, 10	2
7. Changes in the functioning of school factors predict changes in the effectiveness status of schools	3	
Negative results in relation to any assumption	None	None

3. A follow-up study testing the validity of the model at the school level (Creemers & Kyriakides, 2010b).
4. A European study testing the validity of the dynamic model (Panayiotou et al., 2014).
5. A study in Canada searching for grouping of teacher factors: stages of effective teaching (Kyriakides, Archambault, & Janosz, 2013).
6. An experimental study investigating the impact upon student achievement of a teacher professional development approach based on DASI (Antoniou & Kyriakides, 2011).
7. Searching for the impact and sustainability of the dynamic approach on improving teacher behaviour and student outcomes (Antoniou & Kyriakides, 2013).
8. Searching for stages of teacher's skills in assessment (Christoforidou et al., 2014).
9. The effects of two intervention programs on teaching quality and student achievement (Azkiyah et al., 2014).
10. Using the dynamic model to identify stages of teacher skills in assessment in different countries (Christoforidou & Xirafidou, 2014).
11. Using observation and student questionnaire data to measure the impact of teaching factors on mathematics achievement of primary students in Ghana (Azigwe et al., 2016).
12. Searching for the impact of teacher behavior on promoting students' cognitive and metacognitive skills (Kyriakides et al., 2016).

Meta-Analyses

1. A quantitative synthesis of 67 studies exploring the impact of school factors on student achievement (Kyriakides et al., 2010).
2. A quantitative synthesis of 167 studies searching for the impact of generic teaching skills on student achievement (Kyriakides et al., 2013).

First, it is clear that none of these studies and meta-analyses has provided negative results in relation to any assumption of the dynamic model. Moreover, all studies have provided support for the multilevel nature of the model since factors operating at different levels were found to be associated with student achievement gains. These studies have also revealed that teacher and school factors included in the dynamic model are associated with and predictive of student achievement gains. It is important to note that different learning outcomes were used to measure the impact of factors, and thereby some support for the assumption that teacher and school factors are associated with student achievement gains in different learning outcomes has been provided. Second, the two meta-analyses have provided support for the assumption that teacher and school factors have an impact on student achievement and have also revealed that the great majority of effectiveness studies conducted during the last four decades have only been concerned with the impact of the quantitative characteristics of a given factor upon student achievement. In addition, the empirical studies which have been conducted in order to test the validity of the dynamic model have revealed that all five dimensions used to measure quantitative and qualitative characteristics of the functioning of factors should be used to explain

variation in student achievement gains. Particularly, only examining the number of activities related with a factor (i.e., quantitative characteristics) is not sufficient to determine the quality of teaching offered, as the relation of some factors with student achievement may not be linear but curvilinear (Creemers & Kyriakides, 2008). For example, providing students with opportunities to apply new knowledge was found to have a positive impact on their outcomes. However, spending too much teaching time on application activities may not allow sufficient time for teaching new content; which in turn may have a negative effect on student outcomes. Therefore, when measuring the functioning of a factor one should also take into consideration its qualitative characteristics. Third, with regard to the attempt of the model to search for relationships among factors operating at the same level, seven studies have revealed that the teacher factors of the dynamic model and their dimensions can be grouped into specific stages of effective teaching. In this way, more comprehensive strategies for teacher professional development can be developed (see Creemers, Kyriakides, & Antoniou, 2013). Fourth, one of the studies has managed to examine the situational character of school factors, and empirical support for this assumption has been provided. Specifically, the development of a school policy for teaching and evaluation of the policy have been found to have stronger effects in schools where the quality of teaching at classroom level is low (Creemers & Kyriakides, 2009). Finally, a follow-up study testing the validity of the dynamic model was conducted (Creemers & Kyriakides, 2010b). The methods used were identical to those followed by the first study, which had been conducted 5 years earlier. Since the follow-up study had been conducted in the same schools where the original study took place, changes in the effectiveness status of schools and in the functioning of effectiveness factors were identified. Discriminant function analysis reveals that changes in the functioning of school factors can help us classify the schools into those which have improved their effectiveness status, those which have remained equally effective, and those which have even reduced their effectiveness status (see Creemers & Kyriakides, 2010b). Thus, this study was able to test one of the essential differences between the dynamic model and the integrated models developed in the 1990s, which has to do with its attempt to relate changes in the effectiveness status of schools to changes in the functioning of school factors.

Although the studies mentioned above have provided support for the main characteristics and assumptions of the dynamic model, we need further research to test the generalizability of the findings of these studies. Moreover, comparative studies should be conducted in order to find out whether the factors of the model are associated with student achievement in a range of different countries. In this context, a longitudinal study has been conducted in six European countries and revealed that the teacher and school factors of the dynamic model are associated with student achievement gains in mathematics and science (Panayiotou et al., 2014; Panayiotou, Kyriakides, & Creemers, 2016). Given the fact that this study has been conducted in only six countries within Europe, comparative studies should also be undertaken to see if the teacher and school factors can explain variation in student achievement in countries with more diverse educational contexts. Such comparative studies may also be used to develop the dynamic model at system level further,

especially since this international study provided some empirical support for the importance of system level factors in explaining variation on student learning outcomes (Kyriakides, Georgiou, Creemers, Panayiotou, & Reynolds, 2018). Further comparative studies searching for the impact of specific national policies on outcomes in different sociocultural contexts are needed. Such studies may eventually contribute to the establishment of the international dimension of EER (Reynolds, 2006).

It is also important to acknowledge that the empirical studies which have been conducted so far in order to test the validity of the model took place only in primary and pre-primary schools. One could argue that the model is not necessarily able to describe the nature of educational effectiveness in secondary schools. However, the two meta-analyses which looked at the impact of teacher and school factors on student achievement by drawing on data from studies conducted in different countries at the primary and secondary school levels revealed that the factors included in the dynamic model are relevant for both phases of education (see Kyriakides et al., 2010, Kyriakides, Christoforou, & Charalambous, 2013). Clearly, further research on testing the impact of the teacher and school factors included in the dynamic model on learning outcomes of secondary school students would be desirable and may help us to develop further this model and investigate the impact of using the model for school improvement purposes in secondary education.

It is finally important to mention that one of the main aims of establishing the dynamic model was to contribute to the improvement of practice. As a consequence, an approach that makes use of this model has been developed (Creemers & Kyriakides, 2012). The proposed dynamic approach to school improvement (DASI) is based on the assumption that the knowledge base of EER can be used in working with schools in establishing strategies and action plans for improving the quality of education. It is beyond the scope of this chapter to present the main features and steps of this evidence-based and theory-driven approach to school improvement. We however argue that further research is needed for not only developing and testing the dynamic theory of effectiveness but also searching for the impact that an approach based on this theory can have on promoting quality and equity in education.

3.4 Conclusions and Suggestions for Further Research

In this chapter, it is demonstrated that useful contributions were already being made in the theoretical area of EER in the 1990s but there was a need to take them further. Studies testing the validity of the comprehensive model revealed that the variation in ‘what worked’, if it could be explained and theoretically modelled, would encourage the EER field towards the development of more complex and multifaceted accounts than the ‘one size fits all’ mentality that had hitherto existed in the field during the 1990s. In this context, the dynamic model of educational effectiveness has been developed. This model attempts to illustrate the complex and dynamic nature of

educational effectiveness and takes into account the important findings of research on differential effectiveness (Scheerens, 2013). In this chapter, the dynamic model is outlined and factors included in the model at student, teacher, school and system level are presented. In the final part of this chapter, studies supporting the validity of the model are briefly presented. Suggestions for further research on testing the dynamic model by considering both the quality and equity dimensions of teacher, school and system effectiveness are also provided. Such research may help us understand not only why and under which conditions factors included in the model promote learning and so better outcomes for students, but also if these factors have differential effects that can contribute to the promotion of both greater quality and greater equity in education (see Kyriakides et al., 2018a). It is finally argued that since the dynamic model was designed in order to establish stronger links between EER and improvement of practice, we also need experimental studies and/or case studies to identify the extent to which teachers and schools can make use of the dynamic model for improvement purposes. These studies may help us identify when and under what conditions schools can make use of the dynamic model and establish an evidence-based and theory driven approach to school improvement.

References

Anders, Y., Sammons, P., Taggart, B., Sylva, K., Melhuish, E., & Siraj-Blatchford, I. (2011). The influence of child, family, home factors and pre-school education on the identification of special educational needs at age 10. *British Educational Research Journal, 37*(3), 421–441.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: Results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291–311.

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behavior and student outcomes. *Teaching and Teacher Education, 29*(1), 1–12.

Aparicio, J. J., & Moneo, M. R. (2005). Constructivism, the so-called semantic learning theories, and situated cognition versus the psychological learning theories. *Spanish Journal of Psychology, 8*(2), 180–198.

Azigwe, J. B., Kyriakides, L., Panayiotou, A., & Creemers, B. P. M. (2016). The impact of effective teaching characteristics in promoting student achievement in Ghana. *International Journal of Educational Development, 51*, 51–61.

Azkiyah, S. N., Doolaard, S., Creemers, B. P. M., & Van Der Werf, M. P. C. (2014). The effects of two intervention programs on teaching quality and student achievement. *Journal of Classroom Interaction, 49*(1), 4–11.

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*(3), 441–462.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years. IEA's third international mathematics and science study*. Chestnut Hill, MA: Boston College, TIMSS International Study Center.

Bratko, D., Chamorro-Premuzic, T., & Saks, Z. (2006). Personality and school performance: Incremental validity of self-and peer-ratings over intelligence. *Personality and Individual Differences, 41*(1), 131–142.

Brekelmans, M., Sleegers, P., & Fraser, B. J. (2000). Teaching for active learning. In R. J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 227–242). Dordrecht, The Netherlands: Kluwer Academic.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41*(10), 1069–1077.

Brophy, J. (2013). *Motivating students to learn*. New York, NY: Routledge.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York, NY: Macmillan.

Brown, B. W., & Saks, D. H. (1986). Measuring the effects of instructional time on student learning: Evidence from the beginning teacher evaluation study. *American Journal of Education, 94*, 480–500.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*(3), 347–362.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2004). Effective teaching and values: Some implications for research and teacher appraisal. *Oxford Review of Education, 30*(4), 451–465.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 722–733.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432–463). New York, NY: Macmillan.

Choi, J. I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology Research and Development, 43*(2), 53–69.

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher skills in assessment. *Studies in Educational Evaluation, 40*, 1–11.

Christoforidou, M., & Xirafidou, E. (2014). Using the dynamic model to identify stages of teacher skills in assessment. *Journal of Classroom Interaction, 49*(1), 12–25. https://doi.org/10.13140/2.1.4511.2969

Creemers, B. P. M. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2009). Situational effects of the school factors included in the dynamic model of educational effectiveness. *South African Journal of Education, 29*(3), 293–315.

Creemers, B. P. M., & Kyriakides, L. (2010a). Using the dynamic model to develop an evidence-based and theory-driven approach to school improvement. *Irish Educational Studies, 29*, 5–23.

Creemers, B. P. M., & Kyriakides, L. (2010b). Explaining stability and changes in school effectiveness by looking at changes in the functioning of school factors. *School Effectiveness and School Improvement, 21*(4), 409–427.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education: Dynamic approaches to school improvement*. London, UK/New York, NY: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2015). Process-product research: A cornerstone in educational effectiveness research. *Journal of Classroom Interaction, 50*(2), 107–119.

Creemers, B. P. M., Kyriakides, L., & Antoniou, P. (2013). *Teacher professional development for improving quality of teaching*. Dordrecht, The Netherlands: Springer.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London, UK/New York, NY: Routledge.

De Corte, E. (2000). Marrying theory building and the improvement of school practice: A permanent challenge for instructional psychology. *Learning and Instruction, 10*(3), 249–266.

De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the first year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3–31.

Demetriou, A., Kyriakides, L., & Avraamidou, C. (2003). The missing link in the relations between intelligence and personality. *Journal of Research in Personality, 37*(6), 547–581.

Demetriou, A., Spanoudis, G., & Mouyi, A. (2011). Educating the developing mind: Towards an overarching paradigm. *Educational Psychology Review, 23*(4), 601–663.

Den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3–4), 407–442.

Dowson, M., & McInerney, D. M. (2003). What do students say about their motivational goals: Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology, 28*(1), 91–113.

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement, 11*(1), 57–79.

Dronkers, J., & de Heus, M. (2013). The educational performance of children of immigrants in sixteen OECD countries. In M. H. Lopez & M. Siegel (Eds.), *A world in motion: Trends in migration and migration policy*. Oxford, UK: Oxford University Press.

Dustmann, C., Machin, S., & Schönberg, U. (2010). Ethnicity and educational achievement in compulsory schooling. *The Economic Journal, 120*(546), F272–F297.

Elberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principles make a difference? *Economics Education Review, 7*, 291–299.

Fan, X., & Chen, M. (2001). Parent involvement and students academic achievement: A meta-analysis. *Educational Psychology Review, 13*(1), 1–22.

Flavell, J. H. (1979). Meta-cognition and cognitive monitoring -new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145–252.

Fullan, M. (1991). *The new meaning of educational change*. New York, NY: Cassell.

Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2006). New learning environments and constructivism: The students' perspective. *Instructional Science, 34*(3), 213–226.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

Gorard, S., Rees, G., & Salisbury, J. (2001). Investigating the patterns of differential attainment of boys and girls at school. *British Educational Research Journal, 27*(2), 125–139.

Grant, C. A., & Sleeter, C. F. (1986). Race, class and gender effects. *Review of Educational Research, 56*, 219–230.

Gray, J., Peng, W. J., Steward, S., & Thomas, S. (2004). Towards a typology of gender-related school effects: Some new perspectives on a familiar problem. *Oxford Review of Education, 30*(4), 529–550.

Grieve, A. M. (2010). Exploring the characteristics of "teachers for excellence": Teachers' own perceptions. *European Journal of Teacher Education, 33*(3), 265–277.

Gustafsson, J. E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation, 57*, 16–30.

Hall, J., Sylva, K., Sammons, P., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013). Can preschool protect young children's cognitive and social development? Variation by center quality and duration of attendance. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 24*(92), 155–176.

Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement, 9*(2), 157–191.

Hallinger, P., & Heck, R. H. (2011). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. In *International handbook of leadership for learning* (pp. 469–485). Dordrecht, The Netherlands: Springer.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–62.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141–1177.

Harjunen, E. (2012). Patterns of control over the teaching–studying–learning process and classrooms as complex dynamic environments: A theoretical framework. *European Journal of Teacher Education, 34*(2), 139–161.

Harskamp, E.G. (1988). *Een evaluatie van rekenmethoden* [An evaluation of arithmetic curricula] (Dissertation). Groningen, The Netherlands: RION.

Hartig, J., Klieme, E., & Leutner, D. (Eds.). (2008). *Assessment of competencies in educational contexts*. New York, NY: Hogrefe Publishing.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

Hayes, M. T., & Deyhle, D. (2001). Constructing difference: A comparative study of elementary science curriculum differentiation. *Science Education, 85*(3), 239–262.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). An exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5–14.

Hendriks, M. (2014). The influence of school size, leadership, evaluation and time on student outcomes: four reviews and meta-analyses. *Unpublished PhD thesis, University of Twente*, Enschede, The Netherlands.

Hersey, P. B., & Blanchard, K. K. H. (1993). *Management of Organizational Behaviour: Utilizing human resources* (6th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hofman, W. A., Hofman, R. H., & Gray, J. M. (2010). Institutional contexts and international performances in schooling: Comparing patterns and trends over time in international surveys. *European Journal of Education, 45*(1), 153–173.

Hopkins, D. (2001). *School improvement for real*. London, UK: RoutldegeFalmer.

Houtveen, A. A. M., Van de Grift, W. J. C. M., & Creemers, B. P. M. (2004). Effective school improvement in mathematics. *School Effectiveness and School Improvement, 15*(3–4), 337–376.

Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching*. Boston, MA: Allyn & Bacon.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1–10.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Münster, Germany: Waxmann.

Kyriakides, L. (2004). Differential school effectiveness in relation to sex and social class: Some implications for policy evaluation. *Educational Research and Evaluation, 10*(2), 141–161.

Kyriakides, L. (2005a). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103–152.

Kyriakides, L. (2005b). Evaluating school policy on parents working with their children in class. *The Journal of Educational Research, 98*(5), 281–298.

Kyriakides, L. (2007). Generic and differentiated models of educational effectiveness: Implications for the improvement of educational practice. In *International handbook of school effectiveness and improvement* (pp. 41–56). Dordrecht, The Netherlands: Springer.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: A step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*(4), 429–446.

Kyriakides, L., Anthimou, M., & Charalambous, C.Y. (2016). *Searching for the impact of teacher behavior on promoting students' cognitive and metacognitive skills*. Paper presented at the American Educational Research Association (AERA) 2016 conference "Public Scholarship to Educate Diverse Democracies". Washington, DC, April 8 – April 12.

Kyriakides, L., Archambault, I., & Janosz, M. (2013). Searching for stages of effective teaching: A study testing the validity of the dynamic model in Canada. *Journal of Classroom Interaction, 48*(2), 11–24.

Kyriakides, L., & Campbell, R. J. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation, 30*(1), 23–36.

Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501–529.

Kyriakides, L., & Charalambous, C. (2005). Using educational effectiveness research to design international comparative studies: Turning limitations into new perspectives. *Research Papers in Education, 20*(4), 391–412.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143–152.

Kyriakides, L., & Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183–205.

Kyriakides, L., & Creemers, B. P. M. (2009). The effects of teacher factors on different outcomes: Two studies testing the validity of the dynamic model. *Effective Education, 1*(1), 61–86.

Kyriakides, L., Creemers, B. P. M., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal, 36*(5), 807–830.

Kyriakides, L., Creemers, B. P. M., Antoniou, P., Demetriou, D., & Charalambous, C. (2015). The impact of school policy and stakeholders' actions on student learning: A longitudinal study. *Learning and Instruction, 36*, 113–124.

Kyriakides, L., Creemers, B. P. M., & Charalambous, E. (2018a). *Equity and quality dimensions in educational effectiveness*. Dordrecht, The Netherlands: Springer.

Kyriakides, L., Creemers, B. P. M., & Charalambous, E. (2018b). Searching for differential teacher and school effectiveness in terms of student socioeconomic status and gender: Implications for promoting equity. *School Effectiveness and School Improvement., 30*, 286–308. https://doi.org/10.1080/09243453.2018.1511603

Kyriakides, L., Georgiou, M. P., Creemers, B. P. M., Panayiotou, A., & Reynolds, D. (2018). The impact of national educational policies on student achievement: A European study. *School Effectiveness and School Improvement, 29*(2), 171–203.

Kyriakides, L., & Luyten, H. (2009). The contribution of schooling to the cognitive development of secondary education students in Cyprus: An application of regression-discontinuity with multiple cut-off points. *School Effectiveness and School Improvement, 20*(2), 167–186.

Kyriakides, L., & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: A study on school and teacher effectiveness in physical education. *British Educational Research Journal, 34*(6), 807–838.

Kyriakides, L., Anthimou, M., & Panayiotou, A. (2020). *Searching for the impact of teacher behavior on promoting students' cognitive and metacognitive skills. Studies in Educational Evaluation*, 64. https://doi.org/10.1016/j.stueduc.2019.100810.

Lehmann, R., & Lenkeit, J. (2008). *ELEMENT. Erhebung zum Lese- und Mathematikverständnis. Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin* [ELEMENT. Survey for reading and

mathematics literacy. Development in Grades 4 to 6 in Berlin]. Berlin, Germany: Humboldt University of Berlin.

Lenkeit, J., Caro, D. H., & Strand, S. (2015). Tackling the remaining attainment gap between students with and without immigrant background: An investigation into the equivalence of SES constructs. *Educational Research and Evaluation, 21*(1), 60–83.

Lim, L. (2013). Meritocracy, elitism, and egalitarianism: A preliminary and provisional assessment of Singapore's primary education review. *Asia Pacific Journal of Education, 33*(1), 1–14.

Lüftenegger, M., Schober, B., Van de Schoot, R., Wagner, P., Finsterwald, M., & Spiel, C. (2012). Lifelong learning as a goal–do autonomy and self-regulation in school result in well prepared pupils? *Learning and Instruction, 22*(1), 27–36.

Lugthart, E., Roeders, P.J.B., Bosker, R.J., & Bos, K.T. (1989). *Effectieve schoolkenmerken in het voortgezet onderwijs. Deel 1: Literatuurstudie* [Effective schools characteristics in secondary education. Part I: Literature review]. Groningen, The Netherlands: RION.

Mainhard, M. T., Brekelmans, M., & Wubbels, T. (2011). Coercive and supportive teacher behaviour: Within-and across-lesson associations with the classroom social climate. *Learning and Instruction, 21*(3), 345–354.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*(2), 133–163.

Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis, 14*(4), 307–332.

Muijs, D., Campbell, R. J., Kyriakides, L., & Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement, 16*(1), 51–70.

Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice*. London, UK: Sage.

Muijs, R. D., Kyriakides, L., van der Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art-teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256.

Mullis, I. V., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report*. Boston, MA: IEA.

Noble, T. (2004). Integrating the revised Bloom's taxonomy with multiple intelligences: A planning tool for curriculum differentiation. *Teachers College Record, 106*(1), 193–211.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116–130.

Ohinata, A., & van Ours, J. C. (2012). Young immigrant children and their educational attainment. *Economics Letters, 116*(3), 288–290.

Opdenakker, M. C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and Well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*(2), 65–196.

Opdenakker, M. C., & Van Damme, J. (2006). Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools. *School Effectiveness and School Improvement, 17*(1), 87–117.

Ortega, L., Malmberg, L., & Sammons, P. (2018). School effects on Chilean children's achievement growth in language and mathematics: An accelerated growth curve model. *School Effectiveness and School Improvement.* Published online: https://www.tandfonline.com/doi/full/10.1080/09243453.2018.1443945

Panayiotou, A., Kyriakides, L., & Creemers, B. P. (2016). Testing the validity of the dynamic model at school level: A European study. *School Leadership and Management, 36*(1), 1–20.

Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., … Bren, M. (2014). Teacher behavior and student outcomes: Results of a European study. *Educational Assessment, Evaluation and Accountability, 26*, 73–93.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.

Peterson, P., Wilkinson, L. C., & Hallinan, M. (Eds.). (1984). *The social context of instruction: Group organization and group processes*. New York, NY: Academic.

Pianta, R., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Reezigt, G. J., Guldemond, H., & Creemers, B. P. (1999). Empirical validity for a comprehensive model on educational effectiveness. *School Effectiveness and School Improvement, 10*(2), 193–216.

Retali, A. K. (2011). *Students' science achievement, self-beliefs and aspirations in Greece, with a focus on immigrants: An analysis of Greek PISA 2006 data and a cross-national comparison* (Doctoral dissertation, Oxford University).

Reusser, K. (2000). Success and failure in school mathematics: Effects of instruction and school environment. *European Child & Adolescent Psychiatry, 9*(2), S17–S26.

Reynolds, A. J., & Walberg, H. J. (1990). *A structural model of educational productivity*. DeKalb, IL: Northern Illinois University.

Reynolds, D. (2006). World class schools: Some methodological and substantive findings and implications of the international school effectiveness research project (ISERP). *Educational Research and Evaluation, 12*(6), 535–560.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197–230.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York, NY: Macmillan.

Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and Progress: A longitudinal analysis of student achievement over nine years. *British Educational Research Journal, 21*(4), 465–485.

Sammons, P. (1996). Complexities in judging. *School Effectiveness, Educational Research and Evaluation, 2*(1), 113–149.

Sammons, P. (2009). The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement, 20*(1), 123–129.

Sammons, P., Anders, Y., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B. and Barreau, S. (2008). *Children's cognitive attainment and progress in English primary schools during key stage 2: Investigating the potential continuing influences of pre-school education*, Zeitschrift für Erziehungswissenschaften, 10. Jahrg., Special Issue (Sonderheft) 11/2008, pp. 179–198.

Sammons, P., Davis, S., & Gray, J. (2016). Methodological and scientific properties of school effectiveness research. In *The routledge international handbook of educational effectiveness and improvement* (pp. 58–109). New York, NY: Routledge.

Sammons, P., Hall, J., Sylva, K., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013). Protecting the development of 5–11-year-olds from the impacts of early disadvantage: The role of primary school academic effectiveness. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 24*(2), 251–268.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a reanalysis of the inner London education Authority's junior school project data. *British Educational Research Journal, 19*(4), 381–405.

Sammons, P., Toth, K., & Sylva, K. (2018). The drivers of academic success for 'bright' but disadvantaged students: A longitudinal study of AS and A-level outcomes in England. *Studies in Educational Evaluation, 57*, 31–41.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology, 35*(5), 31–38.

Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London, UK: Cassell.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*(1), 1–38.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. New York, NY: Pergamon.

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). *Positioning and validating the supervision framework*. Enschede, The Netherlands: University of Twente.

Schmidt, W., Jakwerth, P., & McKnight, C. C. (1998). Curriculum sensitive assessment: Content *does* make a difference. *International Journal of Educational Research, 29*, 503–527.

Schmidt, W., & Valverde, G. A. (1995). *National Policy and cross-National Research: United States participation in the third international and science study*. East Lansing, MI: Michigan State University, Third International Mathematics and Science Study.

Schoenfeld, A. H. (1998). Toward a theory of teaching in context. *Issues in Education, 4*(1), 1–94.

Shapira, M. (2012). An exploration of differences in mathematics attainment among immigrant pupils in 18 OECD countries. *European Educational Research Journal, 11*(1), 68–95.

Simons, R. J., Van der Linden, J., & Duffy, T. (2000). New learning: Three ways to learn in a new balance. In R. J. Simons, J. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 1–20). Dordrecht, The Netherlands: Springer.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Slater, R. O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement, 3*(4), 247–257.

Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research, 57*(2), 175–213.

Stenmark, J. K. (1992). *Mathematics assessment: Myths, models, Good questions and practical suggestions*. Reston, Virginia: NCTM.

Sternberg, R. J. (1988). Mental self-government: A theory of intellectual styles and their development. *Human Development, 31*(4), 197–224.

Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal, 37*(2), 197–229.

Strand, S. (2012). The White British-Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal, 38*(1), 75–101.

Strand, S. (2014a). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it'for the white working class. *Research Papers in Education, 29*(2), 131–171.

Strand, S. (2014b). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education, 40*(2), 223–245.

Stringfield, S. C., & Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. In B. P. M. Creemers & G. J. Reezigt (Eds.), *Evaluation of effectiveness*. ICO: Groningen, The Netherlands.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *EPPE technical paper 12: Effective provision of pre-school education, the final report*. London, UK: DfES/ Institute of Education.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (Eds.). (2010). *Early childhood matters: Evidence from the effective pre-school and primary education project*. Oxford, NY: Routledge.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation, 12*(6), 561–582.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London, UK: Falmer Press.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Differential secondary school effectiveness. *British Educational Research Journal, 23*(4), 451–469.

Tymms, P., Merrell, C., Heron, T., Jones, P., Albone, S. & Henderson, B. (2006). The importance of districts. *Paper presented at the AERA Annual Conference.* San Francisco, USA.

Valverde, G. A., & Schmidt, W. H. (2000). Greater expectations: learning from other nations in the quest for 'world-class standards' in US school mathematics and science. *Journal of Curriculum Studies, 32*(5), 651–687.

Vermunt, J., & Verschaffel, L. (2000). Process-oriented teaching. In R. J. Simons, J. Van der Linden, & T. Duffy (Eds.), *New learning* (pp. 209–225). Dordrecht, The Netherlands: Springer.

Visscher, A. J., & Coe, R. (2002). *School improvement through performance feedback*. Rotterdam, The Netherlands: Swets & Zeitlinger.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19–27.

Walberg, H. J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 214–229). New York, NY: Macmillan.

Wilson, D., Burgess, S., & Briggs, A. (2005). *The dynamics of school attainment of England's ethnic minorities* (CMPO Working Paper Series No. 05/130). Retrieved from http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp130.pdf

Chapter 4
The Fifth Phase of Educational Effectiveness Research: The Philosophy and Measurement of Equity

Anthony Kelly

4.1 Introduction

Educational Effectiveness Research (EER) investigates, and seeks to explain, the causal relationships in formal school settings between on the one hand *outcomes* (both academic-cognitive and educational-affective) and on the other, *inputs* (student intake characteristics such as social background and prior attainment). Of course, ratios of output-to-input are simply measures of *efficiency*, not effectiveness, but EER goes beyond this to look at the educative *processes* between inputs and outputs[1] and the *policy context* in which these processes are set, and this is what justifies 'effectiveness' in the nomenclature. Recently there has been a certain nervousness within the field about the formality of the institutional setting within the EER paradigm - usually schools but more recently pre-schools and colleges of further education – and this has manifested itself in a desire to be more inclusive in how the field is defined, as if the mere fact of blurring the boundaries of the discipline will meet some higher social aspiration. The practical reality remains that usually only *formal* systems collect the kind of empirical data that underpins EER methodologies and enables modelling to take place at the level of the pupil, the teacher, the school and the system, although some EER studies have also added measures of out-of-school learning (e.g. engagement in private tutoring, time spent on homework and other home learning environment measures) to provide additional controls of

[1]We distinguish between an 'outcome' and an 'output' although the two are conflated in the literature. An outcome (e.g. a satisfying job or a happy life) is the benefit/consequence of an output (e.g. good examination grades or a university degree).

A. Kelly (✉)
Southampton Education School, University of Southampton, Southampton, UK
e-mail: a.kelly@soton.ac.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_4

intake differences between schools or cohorts and to provide insights into the potential importance of such out-of-school learning opportunities.

The educative processes (broadly defined) within EER constitute the sub-field of School Improvement Research (SIR), which investigates factors such as classroom teaching, curriculum, school and classroom behaviour, learning climate, organisation and leadership, and their effect on student outcomes /outputs. These processes are closely linked to policy and system-level factors, just as inputs are intrinsically linked to the social and cultural milieu of the school, but in its 40-year-old morphology, EER has tended to overlook the impact of context on *outcomes*: the field measures *what society deems valuable*. This is not 'context' in the usual sense, but it is a context nevertheless - a *philosophical* one – and it is something that should not be ignored, especially as EER moves into a new phase where it is applied to the issue of equity.

4.2 How the Phases of EER Have Developed

EER accepts that differences in pupil attainment are largely determined by 'ability', however that is defined, and by socio-economic and other factors linked to family, school and neighbourhood, but that schools can and do make a significant difference. It attempts to explain why, how and to what extent, some schools and teachers are more effective than others in achieving desirable outcomes for pupils. It is a quantitative, institutional-focused approach that began in earnest in 1979 as a critique of earlier US research on equity and opportunity carried out by Coleman et al. (1966), who found that only an *insignificant* amount of variance in pupil outcomes was explained statistically by school-level factors and that after taking into account pupil background characteristics such as ability and socio-economic status, only a small proportion of the variation in pupil attainment could be attributed to differences between individual schools. EER was born of a desire to counteract this pessimism in an era that was marked by social frivolity and the dysfunction of schooling for socially disadvantaged pupils, and of not knowing what, if anything, we could do to about it.

The first studies, undertaken by Edmonds (1979) in the US and Rutter, Maughan, Mortimore, and Ouston (1979) in the UK, showed that schools *could* make a difference. Both studies used similar quantitative methodologies in looking at effectiveness between (rather than within) schools. The unit of assessment was the institution, the approach was scientific and the findings (like the initial motivation) were optimistic. EER became the thing it wanted to be, and as time passed and more sophisticated approaches to data modelling were developed, it drove the science upon which many national policies and international comparisons were based. Put simply, it involved measuring a school's output in terms of pupil attainment,

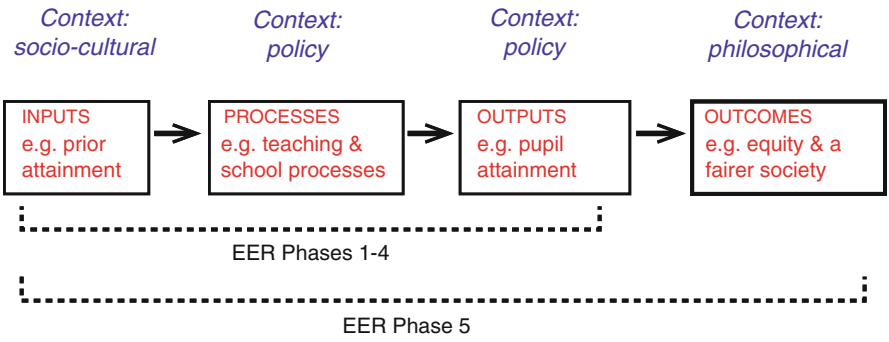

Fig. 4.1 Phases of EER development

correcting for input, circumstance and context, and assigning a scalar to the school effect.[2]

Creemers and Kyriakides (n.d.) have identified four historical phases in the development of the field since 1979, each addressing different research questions and adopting different theoretical stances (see Fig. 4.1). Phase One, from 1979 to the mid-1980s, focused on the size of the school effect, showing (as it was designed to show) that the effect was big enough to matter. Studies looked at differences in the impact that particular teachers and schools had on pupil attainment (ignoring possible confirmation bias) and finding that school and teacher effects were larger for disadvantaged groups (Scheerens & Bosker, 1997). Phase Two, from the mid-1980s to the early 1990s, focussed on the characteristics of effective schools and the factors associated with good teaching and better pupil outputs. In the manner of leadership research within SIR, it produced lists of factors (and lists of lists) characteristic of effective teachers and schools. Phase Three, from the 1990s to the early 2000s, developed theoretical models to explain why some factors, but not others, were important in explaining variations in pupil attainment. Various integrated models of educational effectiveness were developed (Creemers, 1994; Scheerens, 1992; Stringfield & Slavin, 1992) and according to Creemers and Kyriakides (n.d.) these 'guided not only the theoretical development of EER but also the design of empirical studies in the field' (Kyriakides, Campbell, & Gagatsis, 2000; de Jong, Westerhof, & Kruiter, 2004).

Phase Four, from the mid-2000s to the present day, seeks to take better account of context and re-engage with the processes of SIR, especially the kind of improvement that is sustained over time. It acknowledges that effectiveness is not a *stable* trait of schools or classrooms, but is 'dynamic and complex' (Creemers & Kyriakides, 2008), that it can vary from pupil to pupil and can be differentially effective in respect of different outputs. Historians of EER claim that this fourth phase places 'change' at the heart of the field, but this is perhaps to overstate the metamorphosis.

[2]For a detailed discussion of the origin and development of the field, see Reynolds et al. (2014) and Sammons, Davis, and Gray (2016).

Change is merely the absence of certainty and is to EER what the equator is to geography; the *product* of a methodology rather than an object in and of itself. Nevertheless, Phase Four has moved EER into an era of profitable, large-scale, systematic evaluations of policy effects where it has formed an intellectual alliance with transnational organisations like the Organisation for Economic Cooperation and Development (OECD). The dangers of such associations have perhaps been downplayed - as Chekov said in *The Cherry Orchard* "If you run with the hounds you are entitled to bark, but you are also expected to wag your tail' – but the problem of 'context' continues as a motif across the field as context at the level of the school is replaced by context at the level of the nation state, without either being fully satisfied.

The unsatisfactory nature of how context is accommodated methodologically and philosophically is a stone-in-the-shoe of EER, particularly when it comes to equity and fairness, an area where 'context' (in every sense of the word) is critical. This suggests that EER will need to move into a new fifth phase of development as it shifts its main focus to effectiveness-for-equity (see Fig. 4.1). This will require a fundamental consideration of the field's underpinning philosophy. EER has a long tradition of addressing the attainment gap between the 'haves' and the 'have-nots' (Sammons, 2007) and it has always looked at variations in practice between schools and between classrooms with a view to making educational outcomes for disadvantaged pupils more equitable and fairer. The challenge for Phase Five will be to debate and define what is meant by 'fairness' and 'equity' and in what circumstances a degree of inequality is acceptable. Phase Five is akin in some ways to the way theoretical physics moved a century ago from the methodology of the laboratory to the metaphysical paradigm of quantum physics, where previously accepted certainties like time and position had to be defined. So it is with EER. Like Dirac's early work on quantum theory where there were too many 'infinities' (Kragh, 1990), EER has left too much undefined since 1979 in the field's rush to action. What do we mean by effectiveness and equity? How much of what we (think we) know is itself the product of how it has been investigated? And when we talk about education reducing inequality in society, who do we have in mind and why? To date, EER has not articulated a coherent philosophical underpinning (Kelly & Elliott-Kelly, 2018). This may be the result of paying so much attention to methodology over the first four phases of its development and to the rich datasets available to which that methodology has been applied. Phase Five will define generally what is meant by 'equity' in terms of justice and fairness; specifically, to address the problem that schools can improve dynamically in aggregate but adversely affect disadvantaged pupils, and that more effective schooling can sometimes exacerbate achievement differentials in society. This chapter proposes a philosophical underpinning for equity and goes on to look at its measurement.[3] It suggests that the utilitarianism into which EER defaulted *faute de mieux* is no longer appropriate for the moral imperatives at the

[3]Philosophical elements of this chapter have appeared previously in Kelly and Elliott-Kelly (2018) and methodological elements in Kelly (2015).

heart of the paradigm, and that an alternative philosophy based on Rawls's theory of 'justice as fairness' is one way forward, especially in progressing the dynamic theoretical model of Phase Four (Creemers & Kyriakides, 2006, 2008), which treats equity, measured crudely (and wrongly, as an output) by the size of the attainment gap between disadvantaged and advantaged pupils, as a critical internal component.

4.3 The Philosophy of Equity

4.3.1 EER and a Creeping Utilitarianism

Utilitarianism is a philosophy that developed over the course of the Eighteenth Century. In its original form, as developed in 1726 by Francis Hutcheson and more famously in 1780 by Jeremy Bentham, it holds that the best action is the one that maximises utility for the greatest number of people; that is to say, that virtue is in proportion to the number of people that derive benefit from it, and that the best course of action is therefore the one that secures the greatest happiness for the greatest number. Utility is defined as the sum of all the benefit, minus the detriment, that results from an action. In an educational setting, we can equate it with attainment and achievement.

The two defining characteristics of utilitarianism are the aggregation (or averaging) of benefit, and a reliance on the measurement of proxy outcomes. In utilitarianism, as with EER and SIR, actions are judged by their efficacy, convenience and advantage. The outcome of any action is the sole measure of whether it is right or wrong. Utilitarianism comes in several forms determined by their preferred proxy methods for measuring utility, like EER and SIR. Total Utilitarianism, for example, calculates utility as an aggregate, whereas Average Utilitarianism calculates it as an average. Henry Sidgwick, in his 1874 book *The Methods of Ethics*,[4] discusses the question of *aggregate* versus *average* utility, concluding that what should be maximised is the average utility multiplied by the number of people in the population. This means that if average utility remains constant, utilitarianism demands that we make the number of people benefitting as great as possible. Both aggregate and average utility might seem uncontroversial, but problems of moral alignment emerge when we apply either to education and schooling. Maximising *average* utility can mean ignoring all pupils whose educative benefit is below average, which might not be what a community wants at any given time; and maximising *aggregated* utility can lead to the situation where large numbers of pupils having very small educative benefit is regarded as a better outcome than a smaller number of pupils having larger benefit, which again might not be what a society needs (economically or socially) at any given time.

[4]Interestingly, the seventh edition, Sidgwick, H. (1981) *Methods of Ethics* (New York: Hackett Publishing), has a preface written by John Rawls.

Bentham (1780) stated the Principle of Utility as the principle that 'approves or disapproves of every action according to the tendency' it has to 'augment or diminish happiness'. In his so-called 'hedonic calculus', he suggested that the value of a pleasure is measured by the number of people affected. Hutcheson eventually rejected this algorithm as 'useless and disagreeable', but Bentham could see 'nothing unwarranted' in it. Philosophically, this echoes the fixation among policy makers in education with measuring the achievement of intangibles, or failing that, concentrating on what is measurable and ignoring other factors that are important but difficult to gauge (Kelly & Elliott-Kelly, 2018). Despite the advocacy of Creemers and Kyriakides (2008) whose Phase Four dynamic model proposes including non-traditional metrics like well-being (see also Opdenakker & Van Damme, 2000), EER has, over the course of its four phases of development, been driven (or at least, encouraged) unwittingly into its own spurious calculus by policy-makers for whom a utilitarian view of schooling is convenient. This imperative, officially sanctioned, has been secured through targeted funding from quasi-government sources like research councils, and the privileging of certain lines of research inquiry and certain econometric methods at the expense of other approaches.

Bentham's view that in promoting greater utility, 'governments should punish' in proportion to the extent to which certain actions are 'pernicious'[5] also has echoes in today's education policy landscape of accountability and naming-and-shaming schools that do not come up to some notional aggregate water mark. This is not the fault of EER *per se* – the field cannot be blamed for the uses to which its findings are put - but it does show how well-aligned the field has become with the utilitarianism of western government policy and that of transnational bodies like the OECD. *'Natural justice demands that schools are held accountable only for those things they can influence (for good or ill) and not for all the existing differences between their intakes'* (Nuttall, 1990, p. 25), so for EER, exploring the impact of different intake factors is seen as crucial to the field's attempt to promote social inclusion and widen the social distribution of achievement. There was early recognition by many in EER that there were limits to how much schools could compensate for/ameliorate wider disadvantage in society (Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Mortimore & Whitty, 1997), but in the absence of any clearly formulated and asserted philosophy to the contrary, policy-makers unhampered by any pedantic adherence to the evidence have steered some EER (especially by bodies such as the OECD and TIMSS) towards an outlook where utilitarianism seems to have become the default paradigm (Kelly & Elliott-Kelly, 2018).

Modern utilitarianism began with John Stuart Mill. He was a follower and promoter of Bentham's ideas, but his 1863 book *Utilitarianism* rejected as absurd

[5]In his book *Theory of Legislation*, Bentham distinguishes between 'evils of the first and second orders' First-order evils have immediate consequences; second-order evils occur when consequences spread through society causing disruption, and it is the latter that 'makes punishment necessary'.

the latter's purely quantitative measurement of utility. Although Mill asserted the 'pleasures of the intellect' over the more inclusive hedonism of Bentham, the two were agreed that the welfare of the majority was always to be paramount. In Chap. 4 of his book, Mill offered his famous proof for the Principle of Utility: that greater utility is desirable and is 'a good to the aggregate of all persons' because people make it so, in the same way that the proof that an object is visible is that people can see it. This 'notorious' (as Alican, 1994, called it) argument is fallacious on several counts. Firstly, Mill is inferring what people *ought* to do from what they *actually* do; a naturalistic fallacy. Secondly, he is inferring that something *ought* to be desired from the fact that it is *capable* of being desired; a fallacy of equivocation. And thirdly he is inferring that because people desire greater utility *for themselves*, that the aggregate of all persons will desire greater utility *generally*, and that this will be the *only* thing they desire.

The Twentieth Century saw the development of new types of utilitarianism: Act Utilitarianism and Rule Utilitarianism, both of which emphasise the central role of regulation in helping people chose the course of action that maximises utility. The difference between Act and Rule lies in how the action is judged to be the right one: Act Utilitarianism holds that an action is right if *that action* maximises utility; Rule Utilitarianism maintains that an action is right if it conforms *to a rule that maximises utility*. The difference is really about the general versus the specific, and Hare (1973, 1981) developed this into a two-level theory to bring forth the distinction between the two:

- Specific-rule Utilitarianism, which Hare suggests we use when we are deciding what principles to follow. Specific-rule Utilitarianism reduces to Act Utilitarianism.
- General-rule Utilitarianism, which Hare suggests we use this when we are in a situation where natural bias is likely to prevent us from calculating the best course of action. General-rule Utilitarianism does *not* collapse into Act Utilitarianism.

Hare (1981) illustrates his two-level theory using a thought experiment similar to that of John Rawls (1971) a decade earlier, which we will discuss in the next section. Hare conjures up two archetypes to represent the two extremes of 'general' and 'specific'. The Archangel is a hypothetical person who has perfect knowledge of every situation, has no personal bias and always uses critical thinking to do the right thing. The Prole, on the other hand, is completely incapable of critical thinking, uses only intuition and from necessity blindly follows general rules. Hare is not suggesting that people are either Archangels or Proles, but rather that everyone has the characteristics of both to varying degrees, in different contexts and at different times (Kelly & Elliott-Kelly, 2018). Hare did not specify when and under what conditions people act as Archangels and when and under what conditions they act as Proles, but his 'dramatic device' is important in illustrating the primacy in Hare's utilitarianism of formal *critical thinking* (the Archangel's modus operandi) over experiential *intuition* (the Prole's modus operandi). This has parallels in the development of EER, which has come to regard its own modeling as having primacy over the intuition of teachers. In fact, this is acknowledged implicitly in the successful

dynamic model of Creemers and Kyriakides (2008). It is quite proper that autonomous practitioners in any profession should default to critical thinking when working in unusual situations, but in the context of professional practice, whether in schools by teachers or in hospitals by doctors, it is a mistake to regard professional intuition as being devoid of criticality. The dichotomy between critical thinking and intuition is a false one because it privileges thought over action in all situations. The Prole cannot be both robot and trained professional, whether the context is medical triage or classroom practice, and conversely the Archangel cannot be devoid of bias and uncertainty (Kelly & Elliott-Kelly, 2018). That much is obvious from the decades of claim and counter-claim in EER (see, for example, Muijs, Kelly, Sammons, Reynolds, & Chapman, 2011). Utilitarianism also ignores emotional motivations such as jealousy and generosity (Harsanyi, 1975, 1977). It demands only that aggregate benefit be maximized; everything else is disregarded. This is reflected in the development of EER in that peer effects like bullying, friendship and altruistic willingness to help others have been largely ignored, with some exceptions (e.g. Kyriakides, Creemers, Papastylianou, & Papadatou-Pastou, 2014). There is no recognition given to the impact that these factors have on outputs or outcomes nor of the impact of the act of measurement on the factors in the first place. As Kelly & Elliott-Kelly (2018) put it: the veins of utilitarianism run through the body of EER both in terms of how data is selected, collected and treated, and in terms of what is ignored.

In summary then, utilitarianism in education implies a willingness to disadvantage some pupils for the greater good. The question for EER in Phase Five is whether it can accept such an approach to schooling; namely, that some pupils are treated unfairly because the greater good is served by ignoring their plight. Few educationalists would accept that one pupil's deprivation should be weighed against another pupil's achievement or that fairness can be transferred from one pupil to another like a zero-sum financial transaction. Some commentators like Karl Popper (1945) have suggested a workaround – namely, that instead of 'the greatest happiness for the greatest number', we should talk instead about 'the least amount of suffering for the greatest number' – but this just rephrases, rather than solves, the problem of accepting a policy of aggregation. EER already has a proud history of carrying out robust empirical research so that educationalists can choose the most effective course of action towards an end (quite properly, in a democracy) determined by policymakers, but in its next iteration EER needs to articulate the moral impulse behind those policies and decisions (Kelly & Elliott-Kelly, 2018). After all, education is a moral not an economic endeavor, so it needs to identify an immutable ethical basis for equity beyond the current vague notion of having more of it.

4.3.2 EER and John Rawls's Theory of Justice

John Rawls published *A Theory of Justice* in 1971 as an alternative to utilitarianism. It is a practical attempt to address the tension between freedom and equity in a

democratic society, and in that sense it speaks to the underpinning objective of EER. In contrast to utilitarianism, which holds to the single universal principle of maximising aggregated utility, Rawls offers no equivalent universal principle because he regarded 'the correct principle for anything' as depending on 'the nature of that thing', on the actors and on the context (Rawls, 1971, pp. 29–30). Rawls acknowledges that in a democratic society, while people will have different opinions and competing priorities, there can be only one set of laws, and that this poses two challenges:

- The difficulty of having the state exercise coercive power to force everyone to follow the same set of laws. This is Rawls's Principle of Legitimacy, the test for which is that the law is enacted and enforced in ways that all stakeholders can endorse and has 'reciprocity'; that is to say, everyone believes that everyone else will also accept enforcement. Everyone still has their own set of beliefs and values – Rawls called these Comprehensive Doctrines – but they are unwilling to impose their own doctrines on others. Instead, they seek out and agree to mutually acceptable rules as long as everyone else does the same and provided no one group imposes its own Comprehensive Doctrine.
- The difficulty of having people willingly obey the law when that law is devised and implemented by a ruling group whose members probably have different beliefs and values. This is Rawls's Principle of Stability, the test for which is based on his idea of 'overlapping consensus'. In this concept everyone endorses the same core set of laws, *but for different reasons*. People support their own ideas of equity and justice consistent with their own Comprehensive Doctrines, but the core set of laws is common to each doctrine and is therefore supported by everyone. Rawls regards 'overlapping consensus' as a better basis for social stability than 'balance of power', but stability is impossible to achieve when there is insufficient overlap between different sections of society or when there is no convergence on what is meant by equity and fairness.

As far as EER's education policy context is concerned, it is important that these two challenges identified by Rawls are recognised when find their way into the formal schooling system; for example, by supporting schools that promote illiberal ideologies that seek to impose one group's Comprehensive Doctrine on everyone else. EER needs to be aware philosophically of the dangers and ensure that its effectiveness metrics – for example, using ethnicity to contextualise schools' value-added – do not offer perverse incentives towards an undermining divergence and intolerance. Rawls's theory of justice is predicated on what he called a 'spirit of public reason'; namely, the belief that people will justify their political decisions to one another in a respectful manner and only by referencing *publicly accepted* (and not *personal*) values. A basic Constitution will act as an overarching guide to and guarantor of these publicly accepted values - the right to vote, the right to own property, and so forth – but of course there will always be a tension as people (respectfully) express their preferences between the aspiration to create a fair society and the rights of the individual. As Kelly and Elliott-Kelly (2018) point out, this is readily applicable to public schooling; for example, the right to maintain good

schools for everyone, as part of what Rawls calls the Basic Structure, can sometimes be at loggerheads with the right of individual parents to raise their children and to spend their money as they see fit, even (or especially) if doing so benefits their own children at the expense of the system by going to private schools or paying for extra tutoring. For this reason, Rawls sets out, in a very original way, some fundamental principles for the Basic Structure of social institutions like schools and these are discussed below.

4.3.3 Rawls's Veil of Ignorance and the Original Position

Rawls devised a thought experiment in which the principles for his Basic Structure are chosen in a way that forces everyone to choose only those rules that are fair and justifiable to everyone. He proposed a Veil of Ignorance behind which individuals do not know anything about themselves or about their society, so they do not know which choices will affect them positively and which will affect them negatively. Behind this so-called veil, nothing is known about ethnicity, social class, natural ability, intelligence, age, the structure of society or current affairs. Rawls called this baseline of ignorance the Original Position and he saw it as the best method for reaching a 'reflective equilibrium' (Mandle, 2009, p. 17). Everyone in the Original Position knows that their own interests are at risk from the choices they make, so everyone has the incentive to choose principles that protect *everyone's* fundamental interests. It is similar to the maximin strategy in game theory where a player chooses the action that produces the *best of the worst* possible outcomes. The Veil of Ignorance deprives people of all knowledge about themselves and about society, but everyone is allowed to know 'scientific facts'; for example, everyone in the Original Position choosing principles for how society runs schools and educates children is allowed to know the scientific findings of EER and the factors that impact on pupil attainment.

Since the actors do not know their own natural inclinations or circumstances, they do not advocate for any one set of abilities or skills over another and this approach is useful in addressing some issues within EER and its policy context. If parents did not know the intellectual ability of their own children, would they favour the expansion of academically selective schools? If parents were unaware of their own financial situation or social status, would they be in favour of fee-paying schools? If they had no knowledge of whether or not their children had learning difficulties, would they favour or oppose more resources being spent on remedial education? Without knowing whether or not their own children would be affected, what would their preferences be in areas like discipline, streaming by ability and the provision of extracurricular activities (Kelly & Elliott-Kelly, 2018)? These are all questions that the utilitarianism of the earlier phases of EER cannot address beyond demanding that *most* people should benefit in aggregate, but which a Rawlsian approach within an 'equity phase' *can* address.

4.3.4 Rawls's Principles of Justice

Rawls suggests that under the conditions of the Original Position, the following two principles will emerge consensually from behind the Veil of Ignorance:

- People will agree to guarantee basic democratic freedom for everyone; in other words, that everyone will have an equal right to the largest set of basic liberties compatible with everyone else having a similar entitlement. This principle – Rawls's so-called Liberty Principle - will be realised by the *political* institutions of society's Basic Structure and cannot be traded off against other rights or social goods or against economic prosperity.
- People will agree that in order to allow *any* social or economic inequality, the following two conditions *must both* be satisfied:
  - Everyone must have a fair chance of getting the best positions in society. Equally gifted people with the same willingness to apply those skills and work diligently should have equal opportunity, regardless of social status.
  - Inequality in the distribution of primary goods (defined as those things that everyone would want, whatever else they would want) is only justified if it works for the benefit of everyone and for the most disadvantaged; in other words, for inequality to be accepted, everyone, but especially the disadvantaged, must benefit, though perhaps not to the same extent. This is called the Difference Principle and is based on the premise that the distribution of natural skills and abilities is 'undeserved'. A pupil does not *deserve* greater advantage simply because he or she was lucky enough to be born with certain academic talents. This is not to say that everyone must get the same share of society's goods or of nature's bounty, but it means that the distribution of natural ability should be treated as a common asset that should benefit everyone. Those lucky few who are better endowed by nature can use their innate gifts to make themselves better off, as long as they *also* make the disadvantaged better off (Kelly & Elliott-Kelly, 2018).

According to Rawls, both parts of the second principle will be actualised by the *social and economic* institutions of the Basic Structure, which includes schools (although they were not mentioned specifically by Rawls). It is widely accepted that schools should act to increase social mobility and remediate on behalf of pupils from disadvantaged backgrounds, but the position of schools is more nuanced in Rawls's theory. Inequality is something that should benefit *everyone*, especially the most disadvantaged, so the challenge for educationalists in a Rawlsian paradigm, ironically, is to examine how *advantaged* pupils benefit from remediating on behalf of disadvantaged ones! After decades of policy striving to correct the social injustices of a world where disadvantaged children are largely left behind, this feels counterintuitive, but Rawls demands that we think about our inherited assumptions and prejudices so that we can better understand how to create a more just society; in this case, to think about how 'bright' pupils (say) benefit from having less

academically gifted pupils receive additional resources and compensations (like getting extra time in examinations, say).[6] This issue has not been addressed to date in EER because the problem is not recognised in utilitarianism, but it does have a philosophical 'solution' in a Rawlsian paradigm. Benefits accrue to 'advantaged' pupils (and to advantaged sections of society) from the social coherence generated and secured by the fact that disadvantaged pupils get extra help. Cynics might suggest that disadvantaged pupils only receive enough support to maintain them in their subordinate roles, but a more harmonious, less turbulent society is to *everyone's* advantage educationally, culturally and economically. The same logic applies, say, to mixed-ability teaching: what bright pupils supposedly lose in not being 'streamed' they gain from the harmonious atmosphere of the school's learning environment, which is why it is important to take account of peer-to-peer effects in measuring effectiveness (Kelly & Elliott-Kelly, 2018).

4.3.5 A Response to Critiques of Rawls's Theory

Some critics regard Rawls's theory as an attempt to treat effort as morally arbitrary, although effort is the one feature of working-class life that provides self-respect for the disadvantaged. Galston (1991), for example, claims that Rawls's Difference Principle severs the link between the 'willingness to produce' and the 'right to consume', replacing claims made on the basis of achievement with those based merely on existence (Mandle, 2009, p. 31). This is a distortion of Rawls's theory - the Difference Principle is not there to evaluate *individual* shares, but to evaluate *institutional* and *structural* inequalities – but Kelly and Elliott-Kelly (2018) embrace Galston's critique and suggest that Rawlsianism can be suitably modified by his criticism before being applied to EER. When Rawls suggests that society's social and economic institutions, like schools, actualise his second principle, EER might add codicils about the *maturity* and *cultural context* of the education system in question, particularly when making international comparisons. For example, social mobility might be a political obsession in the UK, but it might not be a priority in developing countries where citizens might accept greater inequality (say) as long as it created jobs or alleviated famine (say), even if that inequality did not benefit the disadvantaged *most*, as Rawls requires. And in adapting Rawls, EER also needs to discuss how it relates to stability, which is an issue considered by the dynamic model

[6]It might be useful to distinguish Rawlsianism from traditional egalitarianism at this point. The latter is known for its negative attitude to regulation and its positive attitude towards collective decision-making, so an egalitarian society is motivated by cooperation and peer pressure rather than by competition and regulation. However, modern egalitarianism rejects this, holding that if everyone had the same opportunity cost, there would be no relative advancement and no one would derive any benefit from dealing with others in society. In the egalitarian view, the benefits that people get from dealing with each other arise *because* they are unequal, whether that inequality is from natural or from nurture.

of Creemers and Kyriakides (2008) and in earlier EER research (Bosker & Scheerens, 1994; Sammons, Thomas & Mortimore, 1997; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000). Rawls suggests that his two principles make societies more inherently stable, but he assumes that the societies in question are democratic and it is not clear how this plays out in undemocratic societies like China, which are included (and often lauded) in OECD international educational league tables. Stable dictatorships may not be more desirable that unstable ones for people whose basic freedoms have been traded-off against transient and ecologically catastrophic prosperity.

Other critiques of Rawlsianism mistake its counter-intuitiveness for weakness – for example, his premise that under certain conditions, inequality is 'acceptable' – but this is to lose sight of the fact that opposing philosophies, such as utilitarianism, create inequality under *all* conditions without any preconditions or in the case of communism, strip people of the freedom that Rawls is seeking *a priori* to extract and establish in society as a fundamental human entitlement. The modern utilitarianism of transnational organisations that encourage international competition in education as both participant and referee is a powerful mixture of erudition and hypocrisy designed to obfuscate its own shortcomings, as we will see below with the Restricted Utility Principle in utilitarianism. It survives and has acquired the appeal of a religious sect because it privileges the liberal metropolitan elite who give thanks for their inherited freedoms by whining about them, but have the cultural and economic capital to survive the confusion. With Rawlsianism, on the other hand, what you see is what you get. Rawls simply argues that under his principles people would prefer to maximise the *minimum* amount of benefit that everyone gets, instead of maximising the average amount of primary goods that they receive under utilitarian principles (Freeman, 2003). This is an important point for EER because there is no measure or approach that would be acceptable to international comparison tests which treats success against such a principle; that is to say, that the most successful school or schooling system is the one that maximises minimum (rather than aggregated) pupil achievement. Rawls's principles secure equal rights for everyone; utilitarianism restricts the basic rights of some for the sake of benefit to the many. Within the sphere of education, the latter permits us to restrict a weak minority or deny them access to schooling – for example, by sending home weak or troublesome pupils during an inspection – if it produces greater utility. This is unacceptable in Rawls's Original Position. In Rawlsianism, where everyone can see that everyone else has equal basic liberties, pupils are incentivised by the prospect of cooperation based on transparency and mutual respect.

Kelly and Elliott-Kelly (2018) point to another interesting contrast between Rawlsianism and utilitarianism; specifically, between the former's Difference Principle and the latter's Restricted Utility Principle, which allows a society to maximise wealth with the only constraint being that the worst-off have a minimum income threshold. In education, most nations guarantee minimum provision for children up to a stated legal school-leaving age and afford extra or compensatory provision for pupils with learning difficulties and for those from poor socio-economic circumstances, while it simultaneously encourages *advantaged* pupils to maximise their

attainment at all costs. This is the Restricted Utility Principle in action, but what utilitarianism fails to acknowledge is that those being 'schooled on support' will eventually realise that they are being sacrificed to benefit more advantaged pupil and as a consequence will disrupt everyone's active participation. And in any event, it is not clear that there is any advantage in Restricted Utility; there is no reason why it should deliver greater aggregate utility than Rawls's Difference Principle. Under Rawlsianism, people can still pursue their own advancement. Rawls's principles are congruent with self-interest without disadvantaging others, but it is a congruence 'of the right and the good' (Mandle, 2009), which requires a sufficient number of people to affirm the same principles of equity in the overlapping consensus. The extent to which this is the case today – i.e. that there is any significant overlapping consensus - in education policy in developed countries is a moot point. The incessant tweaking and upheaval, producing no sustained improvement worth the chaos, is evidence enough of Rawls's framing of the problem.

It is worth noting that although the Difference Principle depends on the moral claim that it is unfair for people to benefit differentially because of differences between them *that are not their fault*, Rawls does not think that all arbitrary inequalities are unjust. He regarded the 'natural distribution of talent' as 'neither just nor unjust' (Rawls cited in Mandle, 2009, p. 24), but requires that those who are lucky enough to be born with greater talents – or more accurately, with talents that are in greater demand at the time – are not profiting at the expense of those less fortunate, while still being congruent with self-interest. As Rawls (1971, p. 102) himself says of his own theory, stakeholders 'agree to share one another's fate'. This acknowledges the reality of schooling *as a means of social advancement*, but only provided there is sufficient overlapping agreement between different sections of society to affirm the same principles of equity (Kelly & Elliott-Kelly, 2018). This has implications for the adaptation of Rawls to EER and to education generally: how to remediate for those who are born with less academic talent; how to deal with those who can pay for private education; and how to structure learning in schools given natural imbalances in ability (Kelly & Elliott-Kelly, 2018). While Rawls advocates that social institutions like schools should transform the pool of talent into a fair distribution of outcomes, he does not share the same understanding of distributive justice as those who simply believe that fairness requires us to correct all arbitrary inequalities. Rawls's principles themselves do not require society to even-out handicaps 'as if it were a horse race' (Mandle, 2009, p. 25), especially those inequalities that come from natural endowment. Instead, Rawls demands that those who have the same ability and talent, and the same willingness to work diligently to use those talents, should have the same prospect of success. In EER, the narrow definition of 'success' that has come from utilitarianism means that within the aggregation of outcomes we have failed to check *who exactly* is achieving *what* in schools, just as we have sometimes failed to check whether or not the school system benefits the most disadvantaged as much as it benefits the well-off (Kelly & Elliott-Kelly, 2018). Part Two of this chapter addresses those issues and presents new alternative metrics for measuring equity to add to the existing suite of methodologies within EER.

4.4 The Measurement of Equity

Policy-makers, and transnational organisations like the OECD and the European Union (EU), have encouraged governments to look at educational equity through an economic lens, treating prosperity and consumer choice as desirable outcomes of schooling. Their definitions of educational equity are twofold and can be summarised as: (i) the extent to which pupils can take advantage of education in terms of *opportunities*; and (ii) the extent to which pupils can take advantage of education in terms of *outcomes*. This is a perfectly reasonable view, although there is an intellectual dichotomy contained within it: the former suggests that we look at school effectiveness through the lens of Amartya Sen's capability theory (Sen, 1982, 1984, 1985a, 1985b); the latter suggests that we look at school effectiveness in terms of attainment and its measurement. While this chapter focuses on measurement, it is worth taking a brief detour to consider the part that Sen's capability theory might play in the new paradigm and how it dovetails a Rawls's approach to equity as justice.

In capability theory, equity is about opportunity and taking advantage of it – in fact advantage is a way of viewing relative opportunity – but this is not to be judged solely by pupil attainment. It is possible for a student to have real advantages but not to make good use of them, and it is possible to have opportunity but not to achieve. Opportunity is not simply whether, say, a pupil can get into an over-subscribed school, but whether the pupil can benefit from the curriculum on offer and the learning atmosphere there (see Kelly, 2012a). And acquiring a particular education does not predetermine what a pupil can *do* with it, so as we search for greater equity, some cognisance must be given to what students will actually succeed in *doing* with that greater equity; that is to say, we need to look at pupil 'functionings' and personal achievements. Therefore defining equity in education as the extent to which pupils can take advantage of school in terms of *opportunities* captures the freedom of pupils (and parents) to choose functionings *that they value*, and this is in contrast to the classic utilitarian EER paradigm, which defines effectiveness in terms of outcomes and attainment. The utilitarian paradigm, as discussed in Part One of this chapter, does not distinguish between functioning and capability. For example, entitlement to Free School Meals (FSM) is the most frequently used proxy metric for socio-economic deprivation in school effectiveness research, but being entitled to FSM, which is a functioning, is not the same as the capability of being able to eat FSM without feeling stigmatised. And having a more equitable school system might increase social mobility and the ability of pupils to live a better life, but it is more complicated than that: greater equity needs to be accompanied by (among other things) a raising of expectation among students. Otherwise, those from poor backgrounds, who through fate rather than design occupy the margins of society, become reconciled to under-achievement, as do their teachers; they become habituated to the adverse conditions that induce them to accept and endure their lot.

Returning now to the topic of measuring equity. International studies like the OECD's Programme for International Student Assessment (PISA) attempt to gauge success *at system level* and to facilitate comparisons between nations. To do this, the definition of an equitable system must be narrowed to one in which pupil attainment is independent of those factors that lead to disadvantage, like gender, ethnic minority status, disability, and so forth (see EU, 2006, p. 2). However, they do not use metrics that allow between-school or within-school comparisons to be made, and ultimately they ignore those aspects of equity - like capability, opportunity and democratic access – that go to the heart of well-being. For example, Shanghai is lauded by OECD for its high attainment in mathematics without acknowledging that China is a totalitarian state on the verge of ecological meltdown whose citizens are subject to continuous political repression. Singapore is also singled out for praise, but Singaporeans live without many of the liberal democratic rights that citizens in Western countries take for granted. The criterion for educational success in OECD terms is clearly driven by *economic* imperatives, which is what China and Singapore have in common, but there should be more to educational equity than this 'at-least-the-trains-run-on-time' mentality. Education is more than the training of compliant units of production and there is more to equity than the instrumentalist view that a lack of it inhibits free trade and economic growth. There are *normative* reasons why equity matters, which involve value judgments regarding democratic entitlement and moral purpose, and which go to the heart of the different notions of justice that exist across nations. For Western nations, it *may* be that the price of political freedom is a less efficient schooling system that produces vigilant citizens capable of selecting and de-selecting their own governments. For developing economies, the imperatives may be otherwise. It is important to state this contextual and philosophical limitation clearly here, in advance of describing methods for measuring equity. As this author has noted elsewhere (Kelly, 2015), historic attempts to increase equity have drawn heavily on compensation for disadvantaged groups and have been hampered by the perceived need of policy-makers to link greater equity (for pupils) to greater accountability (for teachers). So after decades of focused policy in this area, it is *still* not clear what the outcome targets should be or how attainment should be spread across the range of factors that impact on it. For example, in England, there is not one single reference to 'equity' or 'equitability' or 'equality' in official school improvement reports like those from the London Challenge (DfES 2003, 2006a, 2006b; Ofsted 2010). It is not even clear what policy-makers mean by 'social justice', although their desire for it is what drives the equity agenda (Sammons, 2008). How does this relate to utilitarianism? Well, the utilitarian approach to EER happily dodges these issues by concentrating on the measurement of attainment without regard for what it means for the student in terms of leading a 'better life' – perhaps this is why the approach is so popular with policy-makers - but neither Sen's capability approach nor Rawls's theory of justice can avoid such consideration. Rawls views a just society as a mixed-motive game where individuals are neither totally selfish nor totally unselfish, but where the dominant inclination is to advance one's own aims through cooperation and agreement rather than through competition and conflict. And capability theory has formal links to Rawlsianism. In fact, Sen's

2009 book *The Idea of Justice* is a revision of Rawls's basic ideas, although it criticises the latter for assuming that social justice is a binary, rather than a continuous, variable. According to Sen, and the approach suggested here for a fifth phase of EER, is that social justice and equity are not things that either exist or do not exist. Multiple conflicting but fair principles can co-exist in an equity paradigm, but the main focus for EER should be that aspect of equity that relates to taking advantage of education in terms of outputs and outcomes. So we will now look at six metrics that have been developed over recent decades to do just that.[7]

4.4.1 The Range Ratio and Its Variations

Range is the difference between the highest and lowest values of a given variable. It is the simplest measure of dispersion, but it is limited by the fact that it uses only two values from the data set and is greatly affected by extreme outliers. Range Ratio is an improvement on simple range because although it still uses only two data points, at least it ignores outliers. It is calculated by dividing the value at a certain percentile above the median, by the value at a certain percentile below the median.

The Range Ratio is most often used for measuring equity in terms of financial expenditure. A version of it commonly used in the US is the Federal Range Ratio, which divides the difference between spending on the pupil at the 95th percentile and the pupil at the 5th percentile, by the spending on the pupil at the 95th percentile.

$$\left(\text{Spending at } 95^{\text{th}} - \text{Spending at } 5^{\text{th}}\right)/\text{Spending at } 95^{\text{th}}$$

Another variation is the Inter-Quartile Range Ratio, which is obtained by dividing the spending on the pupil at the 75th percentile by that on the pupil at the 25th percentile.

$$\text{Spend at } 75^{\text{th}}/\text{Spend at } 25^{\text{th}}$$

Both of these measures could be adapted for use with pupil attainment data, but it would require examination *grades* to be converted to *points* or kept as raw percentages.

In the case of the Federal Range Ratio and the Inter-Quartile Range Ratio, and the Palma Index which we will discuss later, the larger the ratio, the *lower* the equity. The lower limit of 1 occurs when the numerator and the denominator are equal. i.e. when there is zero disparity between the cohorts being measured.

[7] A more detailed mathematical treatment, with worked examples, can be found in Kelly (2015).

4.4.2 The Coefficient of Variation

The Coefficient of Variation (CoV) is calculated by dividing the standard deviation by the mean.

$$CoV = \sigma/\mu$$

Unlike the various range ratios described above, the Coefficient of Variation does not depend on just two data points but takes into account *all* areas of a distribution. The higher the CoV, the less equitable the distribution. The lower limit, which represents perfect equity, is zero.

The Coefficient of Variation represents the spread in the data: when the data is not spread out, the peak is high and CoV is small; a distribution that is more dispersed with a lower peak has a higher CoV representing a less equitable distribution. It should be noted that CoV tends to be biased on the low side (i.e. CoV tends to indicate more equity) when *sample* data is used.

4.4.3 The McLoone Index

Like the Range Ratio, the McLoone Index is most often used for measuring equity in terms of financial expenditure and it is the preferred metric when the lower part of the distribution is of interest (Kelly, 2015). It is calculated using the formula:

$$\sum \frac{(spending \leq the\ median)}{[(Number \leq the\ median) \times (the\ median\ spend)]}$$

The McLoone Index increases as the distribution becomes *more* equitable. Data above the median is not used, but the formula does use a relatively large amount of data and not just two values. The lower limit is zero when the distribution is very *in*equitable and the population below the median receives nothing. The upper limit is 1 when the distribution is perfectly equitable and everyone receives the median amount.

One point to note on the use of the McLoone Index is that the above formula assumes that the disadvantaged group has a *below*-the-median amount of whatever is being measured – usually financial expenditure – so when the disadvantaged group has *above*-the-median amount, it is necessary to invert the McLoone Index ..... or invert the variable (Kelly, 2015); for example, from *pupil-teacher* ratio, say, where lower is better, to *teacher-pupil* ratio where higher is better.

4.4.4 Theil's T

Like the various Range Ratios, the Coefficient of Variation and the McLoone Index, Theil's T has been used traditionally to measure equity in financial expenditure. It is calculated using the equation:

$$T_{indiv} = \sum_{i=1}^{n} [(1/n).(v_i/\mu).\ln(v_i/\mu)]$$

where n is the number of individuals in the population, v_i is the value of the variable (usually financial expenditure) for person i, and μ is the population mean (Theil, 1967).

v_i/μ is the ratio of individual amount to population-average amount, and its natural logarithm determines whether that individual Theil element is positive (when the individual's amount is greater than the mean), negative (when the individual's amount is less than the mean) or zero (when the individual's amount is equal to the mean) in the case of perfect equity (Kelly, 2015). When there is perfect equity and every individual's amount is equal to the mean, Theil's T is at its lower limit of zero. When one person has everything and the distribution is totally inequitable, the above formula reduces to

$$T = \{0\} + \{0\} \ldots + \left\{ \left(\frac{1}{n}\right).\left[\frac{v}{\left(\frac{v}{n}\right)}\right].\ln\left[\frac{v}{\left(\frac{v}{n}\right)}\right] \right\} = \ln n$$

It is possible to compare Theils for two or more schools or groups of schools, but even with everything else equal, schools (or groups of schools) with more pupils have higher Theil upper limits, so it is difficult to draw any firm conclusions. Theil's T is at its best when we are looking at trends over time (Kelly, 2015).

Theil's T can have both within-school and between-school components. For n schools (rather than n individuals), T measures equity *between schools* using the equation:

$$T_{bet-sch} = \sum_{i=1}^{n} [p_i(\mu_i/\mu).\ln(\mu_i/\mu)]$$

where μ_i is the arithmetic mean of group i, p_i is the fraction of the population in school i (the equivalent of 1/n in the 'individual' T equation above) and μ is the population mean as before. This is the '*between-school'* Theil's T, which is the T that would be obtained if every pupil in each school had that school's average share (Kelly, 2015).

The '*within-school'* Theil's T is given by:

$$T_{in-sch} = \sum_{i=1}^{n} (p_i \mu_i / \mu).T_i$$

where T_i is the T of school *i*.

The *overall* T for the whole distribution is given by the between-school and within-school terms added together (and for groups of schools, the 'between-school' component will be the lower limit of the overall Theil's T).

$$T_{overall} = \sum_{i=1}^{n} [p_i(\mu_i/\mu).\ln(\mu_i/\mu)] + \sum_{i=1}^{n} (p_i \mu_i/\mu).T_i$$

4.4.5 The Attainment Equity Index

Unlike the previous four measures, which have their provenance in the fair allocation of financial resources, this equity metric was developed specifically for use with pupil attainment data (Kelly, 2012b). It is a Gini-type index based on the assumption that each percentile range of the population of a given school, as measured by prior attainment, achieves the equivalent percentile range of the school's examination success; that is to say, that a given proportion of a school's examination grades is attributable to an equal proportion of the pupil population. The evidence from the literature suggests that this is how equity is understood by policy-makers in the UK, the US and Europe (Baker & O'Neil, 1994; NCLB, 2001; EU, 2006).

Figure 4.2 explains the basis for Gini-type measures. The straight line $y = x$ represents 'perfect' equity – in other words, the *ideal* distribution of variable y over the population x - and the curve, called a Lorenz curve, represents the *actual* distribution. A Gini Coefficient is defined as:

$$\frac{A}{(A+B)} = 1 - 2B \text{ for normalised axes}$$

where B is the area under the curve and A is the area between the straight line and the curve.

The Lorenz curve, which defines all Gini-based metrics, represents the proportion of a variable y that is cumulatively attributable to the population x. If the Lorenz curve is represented by the function $y = L(x)$, the Gini can also be given by the formula:

$$1 - 2\int_0^1 L(x)dx$$

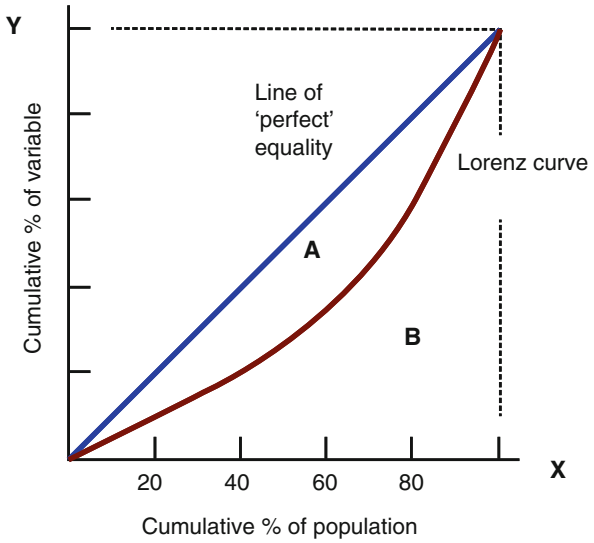

Fig. 4.2 The Lorenz curve and the Gini 'area'

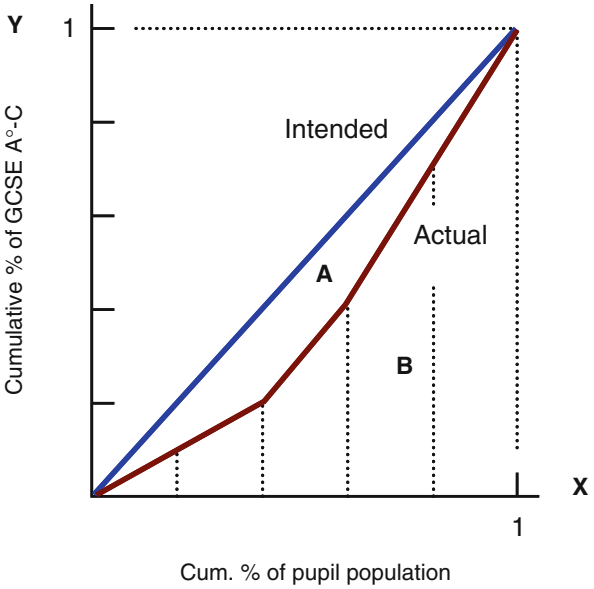

Fig. 4.3 The Lorenz curve approximated

To integrate the Lorenz function, $L(x)$, in the above formula, the equation must be known, which is seldom if ever the case. So for an Attainment Equity (*Æ*) index trapezoids can be used to approximate Area B (see Fig. 4.3) and use the formula:

$$Æ = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})(Y_k - Y_{k-1})$$

where (X_k,Y_k) are the points on the Lorenz, with $X_0 = Y_0 = 0$ and $X_n = Y_n = 1$.

In the example shown on Fig. 4.3, the X_k points on the horizontal axis are every 20%, so that $X_k - X_{k-1} = 0.2$, but it does not appreciably affect the calculation of the *Æ* Index if the intervals are every 10% (say) instead of every 20 per cent, although as Kelly (2015) has noted, most measurements of this kind are lowered by lower 'granularity'.

The Attainment Equity Index has certain advantages over the previous metrics: it uses 'raw' data; it incorporates all data and not just the extremes; it is based on the well-respected Gini concept; it is easily interpreted and can track changes over time; it is well-suited for use with ratio data like examination results; and as Kelly (2015) has suggested, it can be combined with existing value-added measures to categorise schools in terms of both equity *and* other school effects.

4.4.6 The Palma Index

Recently, Cobham and Sumner (2013) developed an equity metric for income inequality that could easily be adapted for use with pupil attainment data, and which they claim is superior to Gini-based metrics like Kelly's Attainment Equity Index. They called it the 'Palma' in honour of the Chilean economist Gabriel Palma (2011) who discovered that in nearly all countries in nearly all circumstances, that the 50% of people lying between the 40th and 90th deciles earn approximately half of all national income, and that this proportion is exceptionally stable.

The Palma Index is a type of Range Ratio, similar to the Federal Range Ratio and the Inter-Quartile Range Ratio discussed at the start of this section. Specifically, it is the ratio of income of the top 10% to the bottom 40%. For example, if the richest 10% earn 45% of a country's income and the poorest 40% earn 20% of the national income, the Palma Index is:

$$P = 0.45 \div 0.2 = 2.25$$

which means that the richest 10% of the country earn two-and-a-quarter times the income of the poorest 40%.

Like all range ratios, the larger the Palma the greater the *in*equity. The lower limit occurs when the poorest 40% earn virtually all of the 'other' 50%; the upper limit is $+\infty$ and occurs when the poorest 40% earn virtually nothing.

Cobham and Sumner (2014) suggest that the Palma Index has advantages over Gini-based indices like the Attainment Equity Index. Gini-based indices do not isolate *where* inequality lies – whether it is in the middle of a distribution or in its tails – and this makes it less than helpful for policy-makers. The Palma, on the other

hand, focuses precisely on the tails (as it defines them) and regards inequality as a question of how the 'other half of the variable' is proportioned between the richest and the poorest. This makes clear to policy-makers what needs to change; namely, to raise the share earned by the bottom 40% at the expense of the top 10%. Gini-based indices like the Attainment Equity Index are also most sensitive to changes around the *middle* of a distribution and are least sensitive to changes in the tails, but the tails are usually the areas of greatest concern economically and educationally, and this is where the Palma is focussed.

Political policy battles are usually fought along the fault-lines of the 'haves' and the 'have-nots' and the Palma fits well into this context, but there are several shortcomings with the Palma as far as educational effectiveness research is concerned. The most critical of these is the question of whether or not the Palma's underpinning assertion holds true that the five deciles from five through nine account for half of all examination success (however that is defined). Given that most assessment in western countries is criterion-referenced rather than norm-referenced, this seems unlikely. Secondly, although the Palma makes clear that what needs to change is the share earned by the bottom 40% at the expense of the top 10%, it is not clear that this can be imposed by policy-makers in schools, since every other government policy seems to incentivise teachers to triage not the bottom 10% but those on the cusp of certain pass grades at whatever decile they occur.

4.5 Technical Properties of Equity Metrics

Choice of metric should depend on policy criteria as much as on technical criteria, and on the view taken of the marginal utility value of attainment. At the start of this chapter, utility was defined as the sum of all the benefit (minus the detriment) that results from an action, but it can also refer to the satisfaction obtained from consuming a good or a service. *Marginal* utility, by extension, is the satisfaction a consumer gains from consuming *more* of the good or service. It is not clear whether educational attainment has a positive or negative marginal utility - that is to say, whether the satisfaction gained from greater examination success is decreasing or increasing – but we can say that when attainment has positive marginal utility, the Coefficient of Variation should be used because it has a 'flat' response to transfer; and when attainment has negative marginal utility, Theil's T should be used because transfer among low achievers is more important and Theil's T is more sensitive at that end of the spectrum. The Attainment Equity Index should be used when one is concerned with *changes* in equity or for middle-ranking schools and schools with comprehensive intakes.

Kelly (2015) has listed two other important and desirable characteristics of equity metrics. Firstly, they should be scale invariant so that multiplying by a constant or changing units should leave the results unchanged. Fortunately, the Coefficient of Variation, Theil's T and the Attainment Equity Index can be made scale invariant simply by dividing by the mean. Secondly, equity metrics should be sensitive to

changes in the data and they should be transferable; that is to say, equity should be shown to decrease when attainment is transferred from someone with less of it to someone with more of it. The Coefficient of Variation is *equally* sensitive to all transfers, which means it is very *in*sensitive as a metric; whereas Theil's T is more sensitive at the lower end of attainment, which means that it tends to *underestimate* inequality in underperforming schools because they are more likely to have larger homogeneous populations of 'have-nots'.

4.6 Conclusions

'Effectiveness' is not a neutral term. Defining it for a particular school will 'always require choices among competing values' and an acknowledgement that 'the criteria will be the subject of political debate' (Firestone, 1991, p. 2). This chapter argues that EER faces challenges regarding its lack of a coherent underpinning philosophy and standing up to those challenges takes the field into a new welcome fifth phase of development. There is little in the early literature to suggest that EER ever felt the need for a formal philosophy as such, so today there is little or no shared understanding within the field of what is meant philosophically or methodologically by 'fairness', 'justice' and 'equity'. The two parts of this chapter have addressed these – philosophy and methodology – in turn.

In many respects, EER has been corralled by the defining characteristics of utilitarianism; namely, the aggregation of utility (which has meant discounting pupils whose benefit is below average or accepting that one pupil's deprivation can be ignored because of another's achievement) and the primacy of a spurious calculus (which has encouraged the field to measure intangibles in an inappropriate fashion or to ignore factors that are difficult to measure). Part One argued that this utilitarian paradigm is outdated in the era of dynamic models (Phase Four) and effectiveness-for-equity (Phase Five) because it ignores the evolutionary basis of the empathy that people feel for each other. Those who have worked in schools know that pupils often sacrifice self-interest for comradeship, helping across the cognitive, conative and affective domains with academic work, socialisation and acceptance, although a review by Gray (2004) indicated that school effects were typically weaker for affective outcomes than for academic, attendance and behavioural outcomes. Utilitarianism is a credible philosophy with a distinguished provenance, but its shortcomings make it *unsuitable* for EER today because of the moral nature of educational equity. It fails to uphold the intrinsic value of the individual above and beyond the collective, and as such it undermines democracy, social justice and the educative imperative. In its stead, this chapter has argued for Rawls's theory of justice as the underpinning philosophy of EER in Phase Five and we have seen how Rawls's Veil of Ignorance approach could help the field theorise about issues like the expansion of academically selective (in the UK, 'Grammar') schools (Asthana & Campbell, 2017), the fair allocation of increasingly scarce resources, streaming by ability and the welfare of high-achieving pupils in the state sector. These are issues

that utilitarianism has failed to address properly because they are *philosophical* rather than *evidential* issues, and although EER has first-class methodological equipment, it does not yet have the philosophical equipment to tackle them. Emerging from this, Kelly and Elliott-Kelly (2018) have suggested a Rawlsian manifesto for EER to realign it without losing any of its methodological advantages. Such a manifesto would serve to underpin the dynamic approach of Creemers and Kyriakides (2008) by adopting the following five principles for Phase Five:

1. *The Benefit Principle:* Educational effectiveness, which is multi-level in nature and dynamic in how it changes over time, is that which increases educative benefit for all pupils, but increases the benefit for disadvantaged pupils more. Inequality is permitted and accepted only if it benefits everyone and especially the most disadvantaged.
2. *The Redress Principle:* Schools – even effectives ones – cannot correct for the fact that some children get a smaller share of society's goods or of nature's bounty, but schools can, and should, make a significant contribution towards redressing undeserved imbalances; that is to say, effective schools have an equity imperative.
3. *The Immutability Principle and the Within-School Imperative:* An effective school does not trade-off one pupil's failure for another's achievement. The metrics to measure effectiveness may include aggregate measures at the level of the school, but they should look primarily at specific educative benefit at the level of the pupil and the classroom. The primary objective is to gauge '*who'* is achieving '*what'*. The metrics to gauge *systemic* effectiveness underpin a secondary objective in looking at aggregated benefit and between-school performance, and this should take account of grouping at the meso-level between the institution and the system, such as with 'chains' of schools and academy 'trusts'.
4. *The Veil of Ignorance Principle applied to Policy:* Controversial national policies should be considered using Rawls's Veil of Ignorance approach, with new experimental methodologies as required, to enable stakeholders to give preferences free from bias and self-interest.
5. *The No-Harm Principle.* All methodologies, including those that enable benevolent and malevolent peer effects to be included, must pass a 'no-harm test' to ensure that they do not adversely affect the overlapping consensus.

Of course, declaring one school or system more equitable than another school or system depends not just on measurement, but on one's definition of equity. Sammons (2007, p. 20) has suggested that 'judgements about school effectiveness need to address three key questions essential to the consideration' and promotion of social justice: the 'what' of effectiveness (which outcomes?); the 'who' of effectiveness effective (which student groups?); the 'when' of effectiveness (over what time period?). These questions provide a sound basis for monitoring both an education system and an individual school's success in promoting equity and equal opportunities for all its students. They can also provide a clear focus for school development and improvement, planning and evaluation. More generally, policy-makers in education see equity – more specifically, *attainment* equity – as a mechanism for

creating greater social mobility and by inference, a more just society. Rawls's contribution in this respect fits well within the EER paradigm, although his theory of justice is not without its detractors. The assumed causal link between equity and opportunity is problematic too. Equality of opportunity does not necessarily result in equality of outcomes - in some contexts, equality of opportunity may actually accentuate *in*equality – and in any case, having more opportunity does not always result in greater achievement. Tristram Hunt (2015), the well-respected former Labour Party shadow Secretary of State for Education in the UK, warned that 'if Labour were ever to abandon equality, there would be very little left to distinguish it from the Conservative Party'. This was a view originally put forward by Hugh Gaitskell, the former (1955–1963) Labour Party leader, but the truth of the matter, as Hunt acknowledges, is that we have already arrived at a policy juncture where the Right is thought not to care about equity and the Left doesn't talk about it. The UK is the most unequal OECD country in Europe: a country where, despite decades of what Hunt calls 'a dizzying array of micro-targeted' education policies, there is a generation growing up with less opportunity and lower achievement than their parents.

> The barriers that bright, wonderful, earnest five year-olds face in modern Britain are wretched to behold. . . . That's the truth about inequality the statistics don't always show. The political philosopher Roberto Unger has described the modern centre-Left as 'content to appear . . . as humanisers of the inevitable'. Sadly, it is not an altogether unfair description. (Hunt, 2015)

Education policies in relation to equity, especially those of western democracies, have failed and if 'the truth about inequality' doesn't always show up in the statistics as Tristram Hunt says, that is a measurement problem. Whatever our philosophical position on equity and justice, and whatever its cause, measuring and monitoring equity can help gauge the effectiveness of policies aimed at increasing it (Kelly, 2015). This is the motivation behind this chapter: we need to know which pupils are achieving the appropriate grades, why there are gaps in achievement and whether, for example, equity might more easily be increased in norm-referenced examination systems. It is an urgent problem: the trend is towards ever-greater inequality. As Hunt (2015) says, it is 'staggering how much schools have to swim against the current when it comes to helping poorer children fulfill their potential'.

References

Alican, N. (1994). *Mill's principle of utility: A defense of John Stuart Mill's notorious proof*. Atlanta, GA/Amsterdam, The Netherlands: Rodopi.

Asthana, A., Campbell, D. (2017, March 6). Theresa May paves way for new generation of grammar schools. *The Guardian*. Retrieved from https://www.theguardian.com/uk-news/2017/mar/06/theresa-may-paves-way-for-new-generation-of-grammar-schools. Accessed 30 March 2017.

Baker, E., & O'Neil, H. (1994). Performance assessment and equity: A view from the USA. *Assessment in Education: Principles, Policy & Practice, 1*(1), 11–26.

Bentham, J. (1780). *An introduction to the principles of morals and legislation* (2007 edition, which is an unabridged republication of the Clarendon Press 1907 edition.) Mineola, NY: Dover Publications.

Bosker, R., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research, 21*, 159–180.

Cobham, A., & Sumner, A. (2013). *Is it all about the tails? The Palma measure of income inequality* (CGD working paper 343). Washington, DC: Center for Global Development.

Cobham, A., & Sumner, A. (2014, February). Is inequality all about the tails? The Palma measure of income inequality. *Significance, 11*(1), 10–13.

Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Creemers, B. (1994). *The effective classroom*. London: Cassell.

Creemers, B., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366.

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Abingdon: Routledge.

Creemers, B., & Kyriakides, L. (n.d.). *Using educational effectiveness research to design theory-driven evaluation aiming to improve quality of education*. Available online at: http://www.ivalua.cat/documents/1/19_10_2011_15_49_19_Ponencia_Taula1_Creemers_Kiriakidis.pdf

De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3–31.

DfES. (2003). *The London challenge: Transforming London secondary schools*. Nottingham, UK: DfES Publications.

DfES. (2006a). *London challenge: Families of schools, May 2006, secondary schools*. London, UK: Department for Education and Skills.

DfES. (2006b). *London challenge: Families of schools, May 2006, primary schools* [in 4 volumes]. London, UK: Department for Education and Skills.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–24.

EU (Commission of the European Communities). (2006). *Communication from the Commission to the Council and to the European Parliament: efficiency and equity in European education and training systems* (Brussels, 8.9.2006, COM2006 481 final).

Firestone, W. (1991). Introduction: Chapter 1. In J. R. Bliss, W. Firestone, & C. E. Richards (Eds.), *Rethinking effective schools: Research and practice*. Englewood Cliffs, NJ: Prentice Hall.

Freeman, S. (Ed.). (2003). *The Cambridge companion to Rawls*. Cambridge, UK: Cambridge University Press.

Galston, W. (1991). *Liberal purposes: Goods, virtues and diversity in the liberal state*. Cambridge, UK: Cambridge University Press.

Gray, J. (2004). School effectiveness and the 'other' outcomes of secondary schooling: A reassessment of three decades of British research. *Improving Schools, 7*(2), 185–198.

Hare, R. M. (1973). The presidential address: Principles, *Proceedings of the Aristotelian Society*, New Series*, 73*, 1–18.

Hare, R. M. (1981). *Moral thinking: Its levels, method and point*. New York: Oxford University Press.

Harsanyi, J. (1975). Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory a theory of justice by John Rawls. *The American Political Science Review, 69*(2), 594–606.

Harsanyi, J. (1977). Morality and the theory of rational behavior. *Social Research, 44*(3), 623–656.

Hunt, T. (2015). *Labour's moral mission: Inequality and social mobility in modern Britain*. Speech to the Fabian Society, 7 December 2015. Private communication with the author.

Hutcheson, F. (1726). *An inquiry into the original of our ideas of beauty and virtue*. (Liberty Fund 2004 edition, edited by W. Leidhold). Indianapolis, IN: Liberty Fund. Retrieved from http://oll.libertyfund.org/titles/hutcheson-an-inquiry-into-the-original-of-our-ideas-of-beauty-and-virtue-1726-2004. Accessed 30 Mar 2017.

Kelly, A. (2012a). Sen and the art of educational maintenance: Evidencing a capability, as opposed to an effectiveness, approach to schooling. *Cambridge Journal of Education, 43*(3), 283–296.

Kelly, A. (2012b). Measuring 'equity' and 'equitability' in school effectiveness research. *British Educational Research Journal, 38*(6), 977–1002.

Kelly, A. (2015). Measuring equity in educational effectiveness research: The properties and possibilities of quantitative indicators. *International Journal of Research & Method in Education, 38*(2), 115–136.

Kelly, A., & Elliott-Kelly, C. (2018). Towards a philosophy of equity in educational effectiveness research: Moving from utilitarianism to a Rawlsian paradigm. *School Effectiveness and School Improvement, 29*(4), 529–544.

Kragh, H. (1990). *Dirac: A scientific biography*. Cambridge, UK: Cambridge University Press.

Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501–529.

Kyriakides, L., Creemers, B., Papastylianou, D., & Papadatou-Pastou, M. (2014). Improving the school learning environment to reduce bullying: An experimental study. *Scandinavian Journal of Educational Research, 58*(4), 453–478.

Mandle, J. (2009). *Rawls's a theory of justice: An introduction*. Cambridge, UK: Cambridge University Press.

Mill, J. S. (1863). *Utilitarianism*. Retrieved from https://www.utilitarianism.com/mill2.htm

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Wells, UK: Open Books.

Mortimore, P., & Whitty, G. (1997). *Can school improvement overcome the effects of disadvantage?* London, UK: Institute of Education University of London.

Muijs, D., Kelly, A., Sammons, P., Reynolds, D., & Chapman, C. (2011). The value of educational effectiveness research: A response to recent criticism. *Research Intelligence, 114*, 24–25.

NCLB (No Child Left Behind) Act. (2001). Public law print of PL 107–110. Available online at: http://www2.ed.gov/policy/elsec/leg/esea02/index.html. Accessed 10 Oct 2014.

Nuttall, D. (1990). *Differences in examination performance* RS 1277/90. London, UK: Research and Statistics Branch, ILEA.

Ofsted. (2010). *London challenge* – Reference No. 100192. London, UK: Ofsted.

Opdenakker, M.-C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*(2), 165–196.

Palma, J. G. (2011). Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-U': It's all about the share of the rich. *Development and Change, 42*(1), 87–153.

Popper, K. (1945). *The open society and its enemies*. London, UK: Routledge. In two volumes. [Reissued in one volume as Popper, K., Ryan, A., & Gombrich, E. (2013). *The open society and its enemies*. Princeton, NJ: Princeton University Press.].

Rawls, J. (1971). *A theory of justice* (1999 edition). Cambridge, MA: Harvard University Press.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197–230.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London, UK: Open Books. [Reprinted 1995. London, UK: Paul Chapman].

Sammons, P. (2007). *School effectiveness and equity: Making connections*. Reading, MA: CfBT.

Sammons, P. (2008). Zero tolerance of failure and new labour approaches to school improvement in England. *Oxford Review of Education, 34*(6), 651–664.

Sammons, P., Davis, S., & Gray, J. (2016). Methodological and scientific properties of school effectiveness research: Exploring the underpinnings, evolution and future directions of the field. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement: Research, policy, and practice* (pp. 25–76). London, UK: Routledge.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective schools and effective departments*. London: Paul Chapman.

Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London, UK: Cassell.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon Press.

Sen, A. (1982). Rights and agency. *Philosophy and Public Affairs, 11*, 3–39.

Sen, A. (1984). *Resources, values and development*. Oxford, UK: Basil Blackwell.

Sen, A. (1985a). *Commodities and capabilities*. Amsterdam, The Netherlands: North-Holland.

Sen, A. (1985b). Well-being, agency and freedom: The Dewey lectures. *Journal of Philosophy, 82*(4), 169–221.

Sen, A. (2009). *The idea of justice*. Cambridge, MA: Harvard University Press & Allen Lane.

Sidgwick, H. (1874). *Methods of ethics* (7th/1981 edition). New York, NY: Hackett Publishing.

Stringfield, S., & Slavin, B. (1992). A hierarchical longitudinal model for elementary school effects. In B. Creemers & G. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35–69). Groningen: ICO.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. Lewes: Falmer Press.

Theil, H. (1967). *Economics and information theory*. Chicago, IL: Rand McNally.

Chapter 5
Extending Educational Effectiveness: The Middle Tier and Network Effectiveness

Daniel Muijs

5.1 Introduction

Educational effectiveness research has traditionally concentrated primarily on the school and classroom levels. More recently, under the impulse of large-scale international assessments such as PISA, there has also been an increased emphasis on the country level. Some effectiveness research has looked at the school district/local authority level, but typically not found this to explain a lot of variance. The largest-scale study, by Tymms et al. (2008), was conducted using seven English databases. The study found that Local Education Authority attended did not affect pupil progress in any meaningful way. Some US studies do appear to show some district-level effects. Caldas and Blankston (1999) did find a very significant district effect on attainment (42%), but this effect almost completely disappeared once family structures were entered into their multilevel model. Lee, Seashore Louis, and Anderson (2012) found positive indirect effects of school districts' use of data and creation of networked communities on pupil attainment. Some studies in the US found a significant relationship between increased spending in low income school districts and improved pupil attainment, suggesting that inequal distribution of resources may impact district effectiveness (Lafortune, Rothstein, & Schanzenbach, 2018; Unnever, Kerckhoff, & Robinson, 2000). Some studies have looked at specific characteristics of 'effective' districts, without however measuring impact on pupil attainment (e.g. De Witte & Schiltz, 2018). Most studies in this area consist of case studies which, while interesting and illuminative regarding processes, do not actually answer the question as to whether districts make a difference to pupil attainment.

These somewhat unpromising findings have led educational effectiveness researchers to somewhat shy away from studying the effects of districts and other

D. Muijs (✉)
Office for Standards in Education, Children's Services and Skills, London, UK
e-mail: daniel.muijs@ofsted.gov.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_5

intermediate structures, focussing on the school and classroom on the one hand, and the national country level on the other hand; the latter given a strong impulse by the growth of large-Scale international student assessment programmes such as PISA. However, it may be time to revisit this. The traditional role of intermediate structures has changed in many countries, where a greater emphasis on school autonomy has led to a reduction in the influence of districts and local education authorities, and concurrently to the emergence of new structures such as Charter Schools in the US and Academies and Free Schools in England and Sweden, which has in turn led to the emergence of new forms of intermediation as it has become apparent that autonomous individual schools often lack school improvement capacity (Simkins, Coldron, Crawford, & Maxwell, 2018).

This issue has led to an emphasis on collaboration and networking, whether emerging spontaneously or as a result of deliberate government policies. The latter is occurring both in systems in which new forms of governance (emphasising competition and school autonomy) have been most emphasised, and in those where more traditional structures continue to prevail. New forms of governance, for example in England and the US, include Academies and Charter schools increasingly forming part of overarching structures such as Multi-Academy Trusts (MATs) and Charter Management Organisations (CMOs); in England in particular, MATs form part of official government policy. The popularity of networks and collaboration as a school improvement model has, however, meant that jurisdictions that have not seen the radical changes present in many English-speaking countries, nevertheless often officially encourage the formation of school networks, for example through financial incentives or legal mandates (as is the case in Belgium and Spain, see Azorin & Muijs, 2017). At the same time, a more bottom-up movement to form networks and collaboratives has grown in many settings, encouraged by perspectives on school improvement that stress the value of schools learning from one another and building professional networks (Hargreaves & Fullan, 2012).

A range of reasons have been put forward as to why networking and collaboration could be a potentially valuable school improvement strategy. Muijs, West, and Ainscow (2010) outlined four key theoretical foundations for networking:

- *Constructivist organisational theory* posits that organisations are sense-making systems creating shared perceptions and interpretations of reality. This sense-making function is essential for organisations to function effectively, but runs the risk of becoming myopic, in that this shared perception of reality may be closed to external influences leading to a disconnection with alternative realities and the organisation's environment. It is this myopia that can be addressed through networking with other organisations or other external partners that can provide access to a *Social capital theory* also posits a potential positive role for networking and collaborationcomplementary cognition (Weick, 1995).
- *Social capital theory* also posits a potential positive role for networking and collaboration, with the value of networking in this perspective seen as lying in its ability to harness resources held by other actors and increase the flow of information in a network.

- School networks have also been compared to *New Social Movements*, fluid, bottom up social networks that coalesce, often for a relatively brief period of time, around shared values and interests (Hadfield, Jopling, Noden, O'Leary, & Stott, 2006).
- Another perspective on the importance of collaboration is provided by looking at Durkheimian notions of *anomie*, which can be defined as malaise in individuals, characterized by an absence or diminution of standards, and an associated feeling of alienation and purposelessness. This is seen as at times characterising schools in particular in disadvantaged communities, and can be alleviated by developing strong ties with other schools (Segre, 2004).

This growing importance of and support for networks of schools begs the question of their impact and effectiveness, and a still small but growing body of work in the field of educational effectiveness research has attempted to look at this question. In this chapter we will review the evidence, focussing on a range of sources both from within and outside of the field.

There is some research on the new intermediary structures such as MATs in the UK and CMOs in the US. To date, as these are relatively new structures, there is limited evidence on the effectiveness of MATs. Most recent studies point to highly variable estimates of impact. Some MATs appear to be highly successful, but the overall picture is mixed, with little evidence that schools that are part of a MAT outperform those that are not (e.g. Greany & Higham, 2018; Finch, Dobson, Fischer, & Riggs, 2016). A similarly differentiated picture emerges from the also relatively limited body of research on CMO's in the US. Again, there is evidence of some highly successful practise, especially in some of the 'no excuses' Charter school chains, but the overall picture is one of variable impact (Finn et al., 2016). There is also more general research on networks and collaborations between schools, and the school improvement field has a long history of research in this area (e.g. Fullan & Hargreaves, 2016) which generally points to positive impacts on professional development and building school capacity. This body of work cannot, however, provide definitive evidence of the impact of networking, and has as such been subject to criticism regarding the inferences made in light of the methodologies used, and the extent to which the theoretical underpinnings of the studies may override a truly empirical approach (Croft, 2015).

The mixed findings from the existing evidence suggest that while such structures can have a significant impact on outcomes, any overall effects of networks and middle-tier structures is overshadowed by significant variance between them, with both more- and less-formal structures differing greatly in effectiveness. This leaves the key question of what makes some collaborations and networks more effective than others. Here, we have even less evidence. One study reviewing the evidence, while acknowledging the dearth of studies linking networking and collaboration to pupil outcomes, suggested the following eight factors as characterising effective networks (Rincon-Gallardo & Fullan, 2016, p. 10):

1. focussing on ambitious student learning outcomes linked to effective pedagogy;
2. developing strong relationships of trust and internal accountability;

3. continuously improving practice and systems through cycles of collaborative inquiry;
4. using deliberate leadership and skilled facilitation within flat power structures;
5. frequently interacting and learning inwards;
6. connecting outwards to learn from others;
7. forming new partnership among students, teachers, families, and communities; and.
8. securing adequate resources to sustain the work.

In an overview of research produced for the English education ministry (Department for Education; DfE), Armstrong (2015) identified strong leadership, well defined structures and processes, a history of collaboration, and clear communication as factors that enhanced collaboration, while threats to school autonomy, perceived power imbalances, additional workload associated with collaborative activity and difficulties in establishing shared objectives and common frameworks were identified as barriers. Muijs (2015a), looking at successful networks of rural school in England, identified building trust, establishing clear goals and building in wins, and focussing on specific activities as factors that characterised effective networks.

All of these studies, however, identified significant gaps in knowledge, not least, as in Rincon-Gallardo and Fullan (2016), the lack of a relationship with pupil attainment. This is a major weakness of extant research, which makes conclusions on both the impact of networking in general and explanations for differential impact problematic. This of course mirrors the question of what makes certain schools more effective than others, the foundational question in school effectiveness research.

In this chapter we will therefore attempt to look more closely at whether or not being part of a school-to-school network is positively related to pupil attainment, using quantitative evidence from studies that do link networking to pupil attainment outcomes. While not in most cases directly addressing processes, differences between the networks studied may also shed some light on reasons for differential effectiveness reported in previous research. As such, this chapter seeks to address at least some of the criticisms made by Croft (2015).

5.2 Methodology

To start to address the above question, we will look at four datasets from the UK, all of which measured pupil outcomes and compared schools engaged in particular forms of collaboration and networking with schools that were not.

5.2.1 *Federations of Schools*

The first dataset is from a study of Federations of schools (Chapman & Muijs, 2014). This project looked at the impact of a particular form of collaboration supported by the government at the time. The term "Federation" encompasses a broad spectrum of collaborative arrangements used to loosely describe a range of partnerships, clusters and collaborations. In general, groups of schools agreed to work together to raise standards, promote inclusion, find new ways of approaching teaching and learning, and build capacity between schools in a coherent manner. This was to be brought about in part through structural changes in leadership and management, in many instances making use of the joint governance arrangements invoked in the 2002 Education Act (https://www.legislation.gov.uk/ukpga/2002/32/contents), which allowed for the creation of a single governing body or joint governing body committee to operate across two or more schools. Federations therefore constitute one particular form of school collaborative, and their expected benefits are due to that collaborative aspect.

A quasi-experimental design was used to explore the question of the impact of Federations on pupil attainment. National pupil- and school-level datasets were collected from the DfE to allow us to look at performance measures controlling for student background over time. Data from the National Pupil Database (NPD) and the Pupil Level Annual Schools Census (PLASC) were requested from and provided by the DfE for this purpose. Data were collected for each year from 2001 to 2008.

As no definitive list of Federations existed, a random sample of 50 Local Authorities was selected. Each local authority was contacted by the members of the research team with the request to identify Federations in their authorities and the schools that were a part of them. A total of 264 schools and 122 Federations were identified in this way.

Each Federation school was matched to a school as similar as possible on key characteristics prior to Federating using Propensity Score Matching methods. Matching variables were:

- Phase (e.g. primary, middle, secondary);
- Type of school (e.g. Academy, Faith School, Grammar School);
- Gender intake (co-educational, single sex boys, single sex girls);
- Performance levels (e.g. % achieving threshold levels in English and maths);
- Pupil intake characteristics (% pupils identified as having Special Educational Needs, % pupils eligible for Free School Meals; FSM);
- Location (this measure went beyond traditional rural/urban identification, and attempted to match areas that were as similar as possible on socio-demographic characteristics);
- School size.

Obviously no schools could be matched identically on these criteria. However, the PSM procedure showed that good matching had been achieved, with no

significant differences between Federation and comparison schools at baseline (3 years before a given Federation was formed).

A range of quantitative methodologies were used to analyse the data, including univariate and multivariate statistics and multilevel modelling. Stata 12 and MLWin 2.18 software packages were used for these analyses.

5.2.2 *Evaluation of a School-to-School Support Programme*

The second dataset comes from a project that was specifically designed to improve primary schools perceived as ineffective or failing in one school district in England (Muijs, 2015b). The district in question was an urban area, and was in the top quartile of most deprived Local Authorities in England. Approximately three quarters of the population was White British. Educational attainment was below the national average, and a relatively large proportion of local primary schools were deemed to be failing or underperforming.

In order to help combat this perceived educational underperformance, the district's school improvement service instigated a programme in which low and underperforming primary, junior and first schools would be partnered with high performing schools, with a view to the latter supporting the former. Each underperforming school was partnered with one support school, though the latter sometimes supported more than one partner. The partnerships could last from 6 months onwards, but did not last beyond 2 years. Supporting schools had to be high performing relative to intake, and show strong leadership, with the Headteacher being either a Local or National Leader of Education (NLE; a designation of principal quality). In total, 37 schools were involved in a total of 17 networks.

To look at the impact of the networks on attainment we used a mixed methods approach involving two phases of data collection and analysis. The first phase used a quasi-experimental quantitative design to look at impact, similar to that of the Federations project mentioned above. Each partnership school was matched to a school as similar as possible on key characteristics in the 2 years prior to joining the partnership. Schools were matched on prior attainment at Key Stage 1 (KS1; end of Year 2 of primary education), ethnicity, gender and social disadvantage (Free School Meal eligibility) and Special Educational Needs status, as well as location (institutions were selected from authorities that are reasonable statistical neighbours to the Local Authority in question). The comparison groups were constructed so that they contained an approximately equal number of students to the sample schools. PSM was used to match the schools, and multilevel modelling to analyse the data.

The second phase used qualitative case studies of both supporting and supported schools within the partnerships to study processes. Nine partnerships and 18 schools were randomly selected to be cases for the qualitative component of the study. In these schools, semi-structured interviews were conducted with senior leaders, middle-level leaders, governors and teachers in both supporting and supported schools, to provide the in-depth information needed to gain an understanding of

the processes involved. Data from the case studies consisted of 49 interviews. Interviews were between 23 and 68 min in length.

5.2.3 *Impact Evaluation of Teaching School Alliances*

The third dataset comes from an impact evaluation of Teaching School Alliances (TSA) for the DfE (Gu, Smethem, Rea, Dunford, & Varley, 2016). Teaching Schools were established as one of the main policy levers designed to fulfil the Government's vision for a self-improving school system. In 2010, the White Paper, 'The Importance of Teaching', set forth a firm commitment to the roll out of a national Teaching Schools programme across the country. This policy initiative was to give 'outstanding' schools the role of leading and developing 'sustainable approaches' to teacher and leadership development across the country. The 2016 White Paper, 'Educational Excellence Everywhere', reinforces the central role that teaching schools are expected to continue to play in the development of a self-improving school system. There are, at the time of writing, over 700 Teaching School Alliances in England.

Teaching School Alliances have six main goals:

1. Initial teacher training;
2. Continuing professional development;
3. Leadership & succession planning;
4. School-to-school support;
5. Specialist leaders of education; and
6. Research & development.

The school-to-school support component is the one we are primarily interested in here, and is specifically meant to improve pupil attainment through the support provided primarily by the outstanding school(s).

Again, propensity score matching was used to match schools in TSA's to a comparator sample using PSM. Free School Meal eligibility, gender, neighbourhood Income Deprivation Affecting Children Index (IDACI) score, SEN and language spoken at home were used as matching variables in the analyses. A range of attainment data were used as outcome measures, including the extent to which the gap in attainment between more and less disadvantaged students was narrowing over the years considered.

Following the successful matching of TSA schools to comparator schools, we used multilevel modelling to look at the relationship between being part of a TSA and pupil outcomes. Two-level multilevel models, with pupils nested in schools, were used in the absence of classroom level data in the NPD.

A team led by Gu et al. (2016) conducted a thorough qualitative analysis alongside this quantitative analysis.

5.2.4 Evaluation of a System Leadership Intervention

The fourth dataset comes from an evaluation of a system leadership intervention for the DfE. The project, funded by the English government, aimed to get headteachers or principals from good or outstanding schools to (with their staff) use their skills and experience to support schools in challenging circumstances. In addition to leading their own schools, they therefore worked to increase the leadership capacity of other schools to help raise standards. They were therefore deployed to support schools that were typically low performing, usually for a relatively short period of time, in order to affect school turnaround and help develop the leadership capacity in the supported school.

A similar quasi-experimental methodology was used to look at the impact of the programme on the supported schools, but also on the supporting schools. PSM was used to match schools to comparators. Free School Meal eligibility, gender, IDACI score, SEN and language spoken at home were used as matching variables in the analyses. Schools were matched on data for 3 years prior to participation in the programme. Attainment data at the end of primary and secondary phases (GCSE) were used as the outcome measures.

Following the successful matching of selected good and outstanding schools to comparator schools, we used simple descriptive statistics (e.g. percentage level 4 and above) and then multilevel growth curve modelling to look at the relationship between support and pupil outcomes. Two-level multilevel models, with occasions nested in schools, were used to model growth in attainment over time.

5.3 Results

In all four projects we were thus looking to answer the question whether or not collaboration leads to improved pupil attainment. This is of course not the only question we could ask when it comes to collaboration. There are various other potential benefits to collaboration that may not directly impact on pupils, such as the opportunity to engage in more extensive professional development, better morale among teacher, and sharing knowledge, but ultimately even such factors should eventually lead to improved pupil outcomes. Below we will discuss the findings provided by the different projects.

5.3.1 Federations

Table 5.1 shows the multilevel modeling results for secondary schools from the Federations evaluation (full results can be found in Chapman & Muijs, 2014). The table shows the coefficients for the pupil level variables (FSM eligibility, IDACI

Table 5.1 Federations project – multilevel models for secondary schools – cohort 1

	Year of foundation	Year 1	Year 2	Year 3	Year 4	Year 5
	Coefficient (std error)					
Constant	**5.570 (0.351)**	**5.328 (0.267)**	**5.831 (0.269)**	**6.303 (0.294)**	**7.153 (0.295)**	**8.435 (0.294)**
Eligible for FSM	**−0.987 (0.097)**	**−1.022 (0.099)**	**−.993 (0.090)**	**−1.023 (0.095)**	**−0.951 (0.101)**	**−1.433 (0.108)**
IDACI score	**−3.851 (0.204)**	**−4.073 (0.235)**	**4.336 (0.218)**	**4.338 (0.232)**	**−4.165 (0.248)**	**−4.626 (0.267)**
Male	**−0.522 (0.199)**	**−0.781 (0.088)**	**−0.542 (0.064)**	**−0.572 (0.068)**	**−0.407 (0.074)**	**−0.576 (0.079)**
School action	**−3.374 (0.103)**	**−3.638 (0.098)**	**−3.237 (0.094)**	**−3.476 (0.094)**	**−4.022 (0.102)**	**−4.290 (0.105)**
School action plus	**−3.700 (0.142)**	**−3.786 (0.130)**	**−3.399 (0.128)**	**−3.627 (0.129)**	**−4.597 (0.131)**	**−4.898 (0.138)**
Statemented	**−3.991 (0.182)**	**−4.015 (0.203)**	**−3.700 (0.189)**	**−4.006 (0.202)**	**−4.688 (0.229)**	**−5.186 (0.245)**
Home language other than English	−0.1021 (0.108)	−0.011 (0.154)	−0.012 (0.154)	−0.023 (0.155)	0.039 (0.164)	−0.182 (0.169)
Black	−0.825 (0.775)	0.459 (0.509)	0.027 (0.159)	−0.154 (0.166)	**0.342 (0.177)**	−0.095 (0.185)
Pakistani/Bangladeshi	1.069 (0. 801)	0.715 (0. 665)	**0.432 (0. 214)**	**0.816 (0. 224)**	0.249 (0. 231)	0.307 (0. 238)
Other Asian	**1.236 (0.253)**	**0.714 (0.261)**	**0.688 (0.224)**	**1.108 (0.229)**	**1.413 (0.246)**	**0.892 (0.264)**
Mixed	−0.553 (0.319)	0.686 (0.457)	0.319 (0.186)	−0.405 (0.233)	**0.422 (0.194)**	0.276 (0.204)
Other	−0.793 (0.683)	−0.660 (0.512)	−0.270 (0.199)	−0.261 (0.187)	−0.065 (0.245)	**−0.737 (0.255)**
Prior attainment at KS2	**4.3295 (0.106)**	**4.108 (0.990)**	**4.155 (0.105)**	**4.084 (0.991)**	**3.924 (1.004)**	**3.964 (0.979)**
Performance federation	0.399 (0.365)	0.476 (0.441)	**0.512 (0.289)**	**0.825 (0.336)**	**1.178 (0.323)**	**1.733 (0.324)**
Academy federation	0.751 (0.600)	0.907 (0.634)	0.867 (0.695)	**1.546 (0.621)**	**1.436 (0.676)**	**1.902 (0.561)**
Cross-phase federation	0.592 (0.518)	0.719 (0.570)	0.895 (0.541)	0.206 (0.603)	0.469 (0.604)	0.234 (0.592)
Percentage variance level 2 (school)	13.8	12.1	11.8	11.9	10.2	8.9
Percentage variance level 1 (pupils)	86.2	87.9	88.2	88.1	89.8	91.1

(continued)

Table 5.1 (continued)

Variance explained compared to previous model						
Level 2	2.1%	8.7%	11.0%	15.6%	23.1%	37.4%
Level 1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Total	0.03%	0.9%	1.2%	1.6%	2.4%	3.8%

Note: Outcome measure: percentage 5A∗-C grades at GCSE. Statistically significant relationships at the .001 level are given in bold. Pupil level variables include FSM eligibility (binary: eligible or not), IDACI score (continuous), gender (binary: Male or Female), ethnic group (categorical: Black, Pakistani/Bangladeshi, Other Asian, Mixed, or Other), special educational needs status (categorical: School Action, School Action Plus, or Statemented, in order of increasing levels of Special Educational Needs) and language spoken at home (binary: English or not). School variables include whether or not a school was part of a Performance Federation, Academy Federation, or Cross-Phase Federation. "Variance explained" is based on comparison to previous models for each year with only pupil-level variables included as predictors

score, gender, ethnic group, special educational needs status and language spoken at home). In addition, variables are included for whether or not schools were part of a Federation, looking at three particular types. Performance federations are those federations in which a high performing school has federated with one or more low performing schools to support their improvement. Academy federations are federations formed of academies which form part of the same MAT (academy chain). Cross-phase federations are federations which combine primary and secondary schools to create a more seamless transition across phases.

These analyses and those for primary schools (Chapman & Muijs, 2014) suggest that federations can have a positive impact on student outcomes, and that the impact of Federations is strongest where the aim of the federation is to raise educational standards by federating higher- and lower-attaining schools. The other forms of Federation have less of a direct relationship with student attainment outcomes, and they may in many cases have been set up for somewhat different goals than raising attainment. Of course, other types of federations may be highly successful in achieving the goals for which they were set up, for example in improving arrangements for transition across school phases. However, they show limited or no impact on pupil outcomes. Notably, it appears to take some time before the impact of Federations appears (only from year 2 are significant differences evident), and effects appear to increase over time. This could reflect some of the legal and cultural complications in setting up Federations.

5.3.2 School-to-School Collaboration

In the evaluation of the school-to-school collaboration project the quantitative analyses showed an overall positive effect. There was clear evidence of an increase in pupil attainment associated with school-to-school collaboration, confirmation of the theoretical support for collaboration as a form of school improvement (Muijs, 2015a).

Table 5.2 shows the final multilevel model results, which regressed a pupil attainment outcome (an indictor of reaching level 4 in reading, writing and mathematics at the end of Key Stage 2; akey outcome variables in the English accountability system, as it is the level pupils are normally expected to achieve) on a range of background variables FSM Eligibility; IDACI score as an area-based measure of deprivation; gender; and special needs status – with statemented representing the highest level of need, followed by school action plus and school action; ethnicity; prior attainment; and whether or not the pupil spoke another language than English as first language at home).

The table above shows that pupils in partnership schools outperformed their peers in matched comparison schools, with the strength of the relationship growing over time. This suggests greater improvement in partnership schools than in matched comparators.

In the case studies, we were able to further probe some of the activities and factors associated with successful partnership. What was clear from these qualitative analyses was that successful partnership working is an intensive process, which requires active intervention in the key processes related to school outcomes: teaching, leadership and assessment. Many of these factors support what we already know from previous studies, such as the need for leadership support, mutual trust and goals, the need to focus on teaching and learning, and the development of capacity in the supporting and supported schools (e.g. Hadfield & Chapman, 2009; Rudd, Lines, Schagen, Smith, & Reakes, 2004; West, 2010). Others, however, reflect factors commonly reported in individual school improvement, but infrequently highlighted in the literature on collaboration, such as the need to build in quick wins (e.g. Hopkins et al., 2014). This study also pointed in particular to the potential power of very specific practices, such as coaching and mentoring, as part of collaborative work, and highlights the intensity required to make school-to-school support work, with all successful partnerships here reporting sustained hands-on engagement with supporting schools across both leaders and classroom teachers. In comparison to individual school improvement, this study highlighted a possible advantage of partnership in that it allowed schools to more quickly develop capacity by building on the existing capacity and expertise in the supporting school, without the need for contracting in external organisations that may have less direct understanding of schools, not least in their emotional and affective contexts. As such, collaboration may provide specific value added to school improvement efforts.

5.3.3 Impact Evaluation of Teaching School Alliances

Analyses of the impact of being part of a TSA on attainment were done for three cohorts of pupils in both primary and secondary. An example of the analyses is provided in Table 5.3, which shows the models for Cohort 1. This uses multilevel modelling to regress the percentage of pupils reaching level 4 in reading, writing and mathematics at the end of Key Stage 2, a range of background variables (FSM

Table 5.2 School-to-school support programme

	Year of foundation	Year 1	Year 2	Year 3
	Coefficient (std error)			
Constant	**5.570 (0.351)**	**5.328 (0.267)**	**5.831 (0.269)**	**6.303 (0.294)**
Eligible for FSM	**−0.987 (0.097)**	**−1.022 (0.099)**	**−.993 (0.090)**	**−1.023 (0.095)**
IDACI score	**−3.851 (0.204)**	**−4.073 (0.235)**	**4.336 (0.218)**	**4.338 (0.232)**
Male	**−0.522 (0.199)**	**−0.781 (0.088)**	**−0.542 (0.064)**	**−0.572 (0.068)**
School action	**−3.374 (0.103)**	**−3.638 (0.098)**	**−3.237 (0.094)**	**−3.476 (0.094)**
School action plus	**−3.700 (0.142)**	**−3.786 (0.130)**	**−3.399 (0.128)**	**−3.627 (0.129)**
Statemented	**−3.991 (0.182)**	**−4.015 (0.203)**	**−3.700 (0.189)**	**−4.006 (0.202)**
Home language other than English	−0.1021 (0.108)	−0.011 (0.154)	−0.012 (0.154)	−0.023 (0.155)
Black	−0.825 (0.775)	0.459 (0.509)	0.027 (0.159)	−0.154 (0.166)
Pakistani/Bangladeshi	1.069 (0. 801)	0.715 (0. 665)	**0.432 (0. 214)**	**0.816 (0. 224)**
Other Asian	**1.236 (0.253)**	**0.714 (0.261)**	**0.688 (0.224)**	**1.108 (0.229)**
Mixed	−0.553 (0.319)	0.686 (0.457)	0.319 (0.186)	−0.405 (0.233)
Other	−0.793 (0.683)	−0.660 (0.512)	−0.270 (0.199)	−0.261 (0.187)
Prior attainment at KS1	**4.3295 (0.106)**	**4.108 (0.990)**	**4.155 (0.105)**	**4.084 (0.991)**
Partnership	0.399 (0.365)	0.476 (0.441)	**0.512 (0.289)**	**0.825 (0.336)**
Percentage variance level 2 (school)	13.8	12.1	11.8	11.9
Percentage variance level 1 (pupils)	86.2	87.9	88.2	88.1
Variance explained compared to previous model				
Level 2	2.1%	8.7%	11.0%	15.6%
Level 1	0.0%	0.0%	0.0%	0.0%
Total	0.03%	0.9%	1.2%	1.6%

Note: Outcome measure is an indicator of whether a pupil reached level 4 or above in reading, writing and mathematics at the end of Key Stage 2. Statistically significant relationships at the .001 level in bold. Pupil level variables include FSM eligibility (binary: eligible or not), IDACI score (continuous), gender (binary: Male or Female), ethnic group (categorical: Black, Pakistani/Bangladeshi, Other Asian, Mixed, or Other), special educational needs status (categorical: School Action, School Action Plus, or Statemented, in order of increasing levels of Special Educational Needs) and language spoken at home (binary: English or not). School variable: Partnership (binary: either partnership school or comparison school). "Variance explained" is based on comparison to previous models for each year without the Partnership variable included

Table 5.3 Impact evaluation of Teaching Schools Alliances

	2012 level 4 and above in English and mathematics			2013 level 4 and above in reading, writing and mathematics			2014 level 4 and above in reading, writing and mathematics		
	Baseline model	Control model	Final model	Baseline model	Control model	Final model	Baseline model	Control model	Final model
Constant	**0.81**	**0.94**	**0.94**	**0.77**	**0.93**	**0.93**	**0.80**	**0.94**	**0.94**
FSM eligible		**−0.06**	**−0.06**		**−0.07**	**−0.07**		**−0.07**	**−0.07**
IDACI score		**−0.08**	**−0.08**		**−0.09**	**−0.09**		**−0.10**	**−0.10**
Gender = male		**0.03**	**0.03**		−0.00	−0.00		0.00	0.00
School action		**−0.39**	**−0.39**		**−0.45**	**−0.45**		**−0.43**	**−0.43**
School action plus		**−0.49**	**−0.49**		**−0.56**	**−0.56**		**−0.53**	**−0.53**
Statemented		**−0.49**	**−0.49**		**−0.68**	**−0.68**		**−0.67**	**−0.67**
Other language		**−0.03**	**−0.03**		**−0.07**	**−0.07**		**−0.05**	**−0.05**
Part of TSA			0.002			0.006			0.009
School level variance	**0.009**	**0.008**	**0.008**	**0.013**	**0.011**	**0.011**	**0.011**	**0.009**	**0.009**
Pupil level variance	**0.144**	**0.100**	**0.100**	**0.163**	**0.118**	**0.118**	**0.147**	**0.106**	**0.106**
% Variance explained		29.4	0.0		26.7	0.0		27.2	0.0

Note: Outcome measure is an indicator of whether a pupil reached level 4 or above in reading, writing and mathematics at the end of Key Stage 2. Significant coefficients at the .001 level in bold. Pupil level variables include FSM eligibility (binary: eligible or not), IDACI score (continuous), gender (binary: Male or Female), special educational needs status (categorical: School Action, School Action Plus, or Statemented, in order of increasing levels of Special Educational Needs) and language spoken at home (binary: English or not). School variable: Part of TSA (binary: either a school was part of a TSA or it was a comparison school)

Eligibility, the IDACI score (an area-based measure of deprivation), gender, special needs status (with statemented representing the highest level of need, followed by school action plus and school action), and whether or not the pupil spoke another language than English as first language at home. We also analysed the relationship of attainment to being part of a TSA. We can see that the pupil background variables were significantly related to attainment, explaining between 27% and 29% of variance. Being part of a TSA was not significantly related to attainment in these analyses. We conducted similar analyses for the other two cohorts in primary, and also looked at average point score in national assessments. These likewise showed no significant impact of being part of a TSA.

In the secondary sample we again conducted the same analyses. Here we found a small positive relationship between being part of a TSA and GCSE grades at the end of secondary schooling in Cohort 1, but no relationship in Cohorts 2 and 3. Being part of a TSA in Cohort 1 (where the relationship reached significance) explained less than 5% of the school-level variance.

The key finding is therefore that being part of a TSA was not generally significantly related to outcomes, except for a few analyses for Cohort 1 in 2013. In the latter cases, the effect size was always very small, however. It is therefore not possible to state a positive impact of being part of a TSA on the measures investigated. There are several possible reasons for this lack of relationship:

- The time the programme has been running is relatively short. Even pupils in Cohort 1 who took their Key Stage 2 tests in 2014 have not gone through a full Key Stage as part of a TSA.
- This problem is further exacerbated by the nature of the intervention. Previous research has shown that setting up successful school networks takes time, and it may take 1–2 years before actual network activities are fully under steam.
- As a school-level intervention, the impact on pupil outcomes is indirect, with improvements in teaching quality expected to follow from successful implementation of e.g. continuing professional development, research and initial teacher education. This improvement in teaching quality would then result in improved pupil outcomes. This makes it both less likely that we will find a direct effect on outcomes, and more likely that any effect would take time to manifest itself.
- There may be significant variability in the performance of TSAs, which could make overall programme effects hard to measure.

The latter hypothesis receives support from Gu et al.'s (2016) analysis, which strongly suggests both that there are large differences in the way TSA's function, and that in most cases the primary focus of attention was on developing the teacher training element of the alliance.

5.3.4 *Evaluation of a System Leadership Intervention*

Table 5.4 shows an example of the analyses conducted for the evaluation of the system leadership intervention. The table shows results for one of the cohorts of supporting Headteachers, and looks at both the sample as a whole and the sample of supported schools only. The analyses are based on Total Point Score on the GCSE (end of secondary schools) examinations. As in previous analyses, we regressed the outcome measure on a range of pupil background variables: FSM Eligibility, IDACI score (an area-based measure of deprivation affecting children), gender, special needs status, and whether or not the pupil spoke another language than English as first language at home.

We did not find a significant relationship in the sample as a whole, which included both supporting and supported schools. In the analyses for the supported schools alone, on the other hand, system leader support was significantly related to greater improved attainment over time than in the comparison schools. This improvement was temporally subsequent to initiation of system leader support, and occurred from 2013 onwards, allowing us to suggest that the relation could be causal. Being part of the intervention explained 11.4% of the total variance in pupil outcomes.

In the supported schools the interaction term between deployment and FSM was significant and the reduction occurred following the start of system leader support. This suggests that the intervention had a greater impact on the attainment of pupils eligible for FSM.

Looking at the analyses for the sample as a whole (which includes both the supported and the supporting schools) in primary, system leader support was not significantly related to attainment, but there was a significant relationship between system leader support and improvement in attainment in the supported schools. Furthermore, this improvement was temporally subsequent to initiation of support. The relationship was substantial in size, explaining 10% of variance in attainment changes over time. In the 2012/2013 cohort there was also some evidence that the gap in attainment between disadvantaged pupils and others was reduced in the supported schools.

In addition to the attainment analysis, we also analysed the impact of system leader support on Ofsted inspection grades. To do this, we compared inspection grades in the years following deployment with those of the last inspection before deployment, and conducted a statistical significance test. On average, Ofsted grades for overall effectiveness, pupil achievement, quality of teaching, and effectiveness of leadership and management improved significantly post- system leader support compared to pre-support, particularly in the supported schools.

Table 5.4 Evaluation of a system leadership intervention, 2012/2013 cohort

	Full sample (includes both the supported and the supporting schools) (N = 640)				Supported only (N = 425)			
	Baseline model	Control model	Final model	Closing the gap model	Baseline model	Control model	Final model	Closing the gap model
Constant	**262.1**	**290.1**	**293.1**	**294.8**	**297.5.0**	**318.2**	**320.5**	**322.1**
FSM eligible		**−73.1**	**−73.8**	**−73.3**		**−37.2**	**38.1**	**37.8**
IDACI score		**−85.6**	**−85.4**	**−85.4**		**−178.4**	**−180.4**	**−180.4**
Gender = male		**−23.2**	**−23.3**	**−23.3**		**−23.5**	**−23.4**	**−23.4**
SEN score		**−120.5**	**−122.5**	**−122.5**		**−155.6**	**−158.1**	**−158.1**
Other language		**8.4**	**8.3**	**8.3**		**11.4**	**−11.3**	**−11.3**
System leader support			9.4	9.4			**16.2**	**16.0**
System leader support /FSM interaction				3.1				**10.2**
School level variance	**5015.7**	**1852.4**	**1851.5**	**1851.0**	**3602.8**	**1528.4**	**1402.8**	**1391.2**
Occasion level variance	**3608.8**	**2321.6**	**2318.2**	**2317.6**	**1910.2**	**1302.9**	**1104.3**	**1003.6**
Occasion/ interaction random slopes coefficient				286.5				**627.8**
% variance explained		51.6	0.0	0.0		48.6	11.4	4.5

Note: Outcome measure is Total Points score at GCSE. Significant coefficients at the .001 level in bold. Pupil level variables include FSM eligibility (binary: eligible or not), IDACI score (continuous), gender (binary: Male or Female), special educational needs status (ordinal: No *SEN* School Action, School Action Plus and Statement from lowest to highest level of need) and language spoken at home (binary: English or not). School variable: System leader support (binary)

5.4 Discussion

We started this chapter with the intention of looking at whether or not collaboration and networking between schools can lead to school improvement. We specifically looked at pupil attainment as the outcome measure, as the ultimate aim of these educational interventions is to improve pupil outcomes, with learning outcomes being the primary factor that can be influenced by schooling.

To accomplish this aim, we looked at four interventions in England, all of which used forms of collaboration and networking as a lever for school improvement: the Federations of schools programme, a school-to-school collaborative improvement programme, a national programme of school networks with a somewhat disparate set of aims, and a system leadership intervention in which successful headteachers were asked to support schools in difficulty. So what do our findings suggest?

- In the Federations analyses, we found some evidence of a positive impact, though this was mainly the case in so-called performance federations, where strong schools supported weaker ones.
- In the school-to-school collaboration project, which took place in a single district, results showed a positive impact of collaboration.
- In the national network programme, no impact on attainment was found.
- In the system leadership project, we found a positive impact in supported schools.

These mixed findings suggest that we potentially still have little conclusive to say about the impact of networking and collaboration. However, when we consider the characteristics of the analysed projects, some interesting similarities and differences emerge. These may suggest a way forward in terms of looking at the impact of networking as a school improvement approach.

The most positive data in terms of impact come from the system leadership intervention and the local school-to-school collaboration project. Though one is a large-scale national initiative and the other a small-scale local district one, they share a number of similarities. Firstly, both have *clear aims* in terms of what the project is about. In both cases, strong schools/leaders support schools in difficulty, with the express aim of engendering improvement in the supported school. This is intended to happen relatively quickly, with the support being time-limited. Secondly, in both cases the role of the principal or Headteacher is key in leading on the support role and working with senior leaders in the supported school. Thirdly, there is a central role for external brokers, the local authority in the local intervention and a national agency in the system leadership intervention. These brokers match supporting to supported schools based on knowledge of school characteristics and capacity. Qualitative data from the local study is suggests ways in which this support is provided.

In the TSA initiative, in which we found no effect, the picture is rather different. The networks were expected to fulfil six goals, only one of which was school improvement through collaboration and mutual support. Gu et al.'s (2016) analysis suggests that the bulk of effort to date has gone into the goal of setting up structures

and systems for delivering initial teacher education, while other goals are delivered to a different extent. Support was arranged internally in a given network, with no external brokering. Networks were not time-limited, and were fixed in nature compared to the more transient arrangements in the system leadership intervention.

The Federations project saw differential impacts, but the fact that it was performance federations which showed an impact leads us tentatively to similar conclusions, as it is these Federation configurations that most clearly focussed on improvement and support.

Theoretically, these findings are not dissimilar to those from a range of school improvement research, which similarly points to the importance of factors in improving individual schools such as:

- Focus on a small number of clear, shared and inflexible goals;
- Early achievements to create momentum;
- The key role of leadership;
- The creation of an ecology of support; and
- Focus on those most in need of improvement (Potter, Reynolds, & Chapman, 2002).

It would appear that similar factors may need to be present in networks and collaboratives if their potential for improving schools is to be realised. This does bring into question some of the truisms often posited about collaboration and networks, such as the need for long-term maintenance of structures (Hopkins et al., 2014). In terms of the networking theories mentioned above, they would appear to most closely align with social capital theory, in that they specifically aim to address key gaps that may hinder high performance in supported schools.

Of course, there are clear caveats to these conclusions. This chapter draws on four studies of different networking approaches, which remains a small set of studies to look at. The studies were not systematically chosen to highlight different forms of collaboration and networking, so the interpretation of findings is necessarily tentative. Quasi-experimental research is imperfect in terms of providing causal evidence, the lack of true randomisation leaving open the possibility of alternative explanations for differences between intervention and comparison groups, such as prior differences in leadership capacity. The studies were all conducted in England, with its particular policy context, which by definition limits generalizability of the findings to other countries. We matched schools on those variables that were available in the national datasets used, which in general are decent proxies of intake, but we have not controlled for all possible differences between schools. Schools in collaborations may be more dynamic or have greater capacity for change than comparator schools. They may also have experienced unmeasured staff changes over the time period studied. Our measures of pupil intake are incomplete, not providing, for instance, a fully valid measure of parental socio-economic status (we used proxies such as IDACI scores and FSM eligibility rather than more direct measures such as parental education and occupation levels, as the latter are not collected in the national datasets used). We are also lacking process data, and a clear

requirement for future work in this area is to link quantitative analyses as done here to both survey and qualitative evidence on processes.

Another issue is what the goals of collaboration and networking are. In these analyses we have strongly focussed on pupil attainment as an outcome. However, networks can have different goals, which they may very well be achieving while not raising attainment. There is also a question of sustainability, as the analyses undertaken here were over a relatively short period of time, and we thus do not know whether improvements measured were sustained. It may be that sustainable improvement requires more of the elements mentioned by Rincon-Gallardo and Fullan (2016) to be present.

What these findings suggest is that though potentially valuable, networking is not a panacea for school improvement. As such, this chapter may help to explain some of the mixed findings from international research on middle tier structures. It is not a new finding in school improvement research that it is implementation and processes which ultimately determine outcomes, rather than structures, but it is something that education systems tend to lose sight of in the search for structural solutions to knotty educational problems.

References

Armstrong, P. (2015). *Effective school partnerships and collaboration for school improvement: A review of the evidence*. London, UK: Department for Education.

Azorin, C., & Muijs, D. (2017). Networks and collaboration in Spanish education policy. *Educational Research, 59*(3), 273–296.

Caldas, S. J., & Blankston, C. L. (1999). Multilevel examination of student, school, and district-level effects on academic achievement. *The Journal of Educational Research, 93*(2), 91–100.

Chapman, C., & Muijs, D. (2014). Does school-to-school collaboration promote school improvement? A study of the impact of school federations on student outcomes. *School Effectiveness and School Improvement., 25*(3), 351–393.

Croft, J. (2015). *Collaborative overreach: Why collaboration probably isn't key to the next phase of school reform*. London, UK: Centre for Education Economics.

De Witte, K., & Schiltz, F. (2018). Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional benefit-of-the-doubt approach. *European Journal of Operational Logistics, 267*(3), 1172–1181.

Finn, C. E., Manno, B. V., & Wright, B. L. (2016). *Charter schools at the crossroads*. Cambridge, MA: Harvard Education Press.

Finch, A., Dobson, B., Fischer, E., & Riggs, A. (2016). *Academy chains unlocked*. London, UK: Reform.

Fullan, M., & Hargreaves, A. (2016). *Bringing the profession back in*. Oxford, OH: Learning Forward.

Greany, T., & Higham, R. (2018). *Hierarchy, markets, and networks: Analysing the self-improving school-led system agenda in England and the implications for leadership*. London, UK: IOE Press.

Gu, Q., Smethem, L., Rea, S., Dunford, J., & Varley, M. (2016). *Teaching schools evaluation. Cases study reports*. London, UK: Department for Education.

Hadfield, M., & Chapman, C. (2009). *Leading school-based networks*. London, UK: Routledge.

Hadfield, M., Jopling, M., Noden, C., O'Leary, D., & Stott, A. (2006). *What does the existing knowledge base tell us about the impact of networking and collaboration? A review of network-based innovations in education in the UK*. Nottingham, England: National College for School Leadership.

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York, NY: Teachers College Press.

Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: a narrative state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 257–281.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal – Applied Economics, 10*(2), 1–26.

Lee, M., Seashore Louis, K., & Anderson, S. (2012). Local education authorities and student learning: The effects of policies and practices. *School Effectiveness and School Improvement, 23*(2), 133–158.

Muijs, D. (2015a). Collaboration and networking among rural schools: Can it work and when? Evidence from England. *Peabody Journal of Education, 90*(2), 294–305.

Muijs, D. (2015b). Improving schools through collaboration. A mixed methods study of school-to-school partnerships in the primary sector. *Oxford Review of Education, 41*(5), 563–586.

Muijs, D., West, M., & Ainscow, M. (2010). Why network? Theoretical perspectives on networking and collaboration between schools. *School Effectiveness and School Improvement, 21*(1), 5–27.

Potter, D., Reynolds, D., & Chapman, C. (2002). School improvement for schools facing challenging circumstances: A review of research and practice. *School Leadership and Management, 22*(3), 243–256.

Rincon-Gallardo, S., & Fullan, M. (2016). Essential features of effective networks in education. *Journal of Professional Capital and Community, 1*(1), 5–22.

Rudd, P., Lines, A., Schagen, S., Smith, S., & Reakes, A. (2004). *Partnership approaches to sharing best practise*. Slough, England: NFER.

Segre, S. (2004). A Durkheimian network theory. *Journal of Classical Sociology, 4*(2), 215–235.

Simkins, T., Coldron, J., Crawford, M., & Maxwell, B. (2018). Emerging schooling landscapes in England: How primary system leaders are responding to new school groupings. *Educational Management Administration and Leadership., 47*, 331–348. https://doi.org/10.1177/1741143217751079

Tymms, P., Merrell, C., Heron, T., Jones, P., Albone, S., & Henderson, B. (2008). The importance of districts. *School Effectiveness and School Improvement, 19*(3), 261–274.

Unnever, J. D., Kerckhoff, A. C., & Robinson, T. J. (2000). District variations in educational resources and student outcomes. *Economics of Education Review, 19*(3), 245–259.

Weick, K. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

West, M. (2010). School-to-school cooperation as a strategy for improving student outcomes in challenging contexts. *School Effectiveness and School Improvement, 21*(1), 93–112.

Chapter 6
Extending Educational Effectiveness: A Critical Review of Research Approaches in International Effectiveness Research, and Proposals to Improve Them

David Reynolds, Anthony Kelly, Alma Harris, Michelle Jones, Donnie Adams, Zhenzhen Miao, and Christian Bokhove

6.1 Introduction: The Rise of International Effectiveness Research

In the last 20 years considerable attention has been focussed upon the variation between countries in educational achievement and the apparent relative effectiveness of their educational systems. Partly this is a reflection of a world that is, in many respects, becoming 'smaller' all the time. The spread of mass communications and new information technologies is affording all countries a more international 'reach' in their world views. The revolution afforded by the pervasive spread of information means that ideas now travel 'virally' around the world with great rapidity, making it increasingly possible for educational ideas and processes to move freely too. In education, the process of 'internationalisation' has taken longer to embed than in

D. Reynolds (✉) · A. Harris · M. Jones
Swansea University School of Education, Swansea University, Swansea, UK
e-mail: David.Reynolds@swansea.ac.uk; alma.harris@swansea.ac.uk; michelle.s.jones@swansea.ac.uk

A. Kelly · C. Bokhove
Southampton Education School, University of Southampton, Southampton, UK
e-mail: a.kelly@soton.ac.uk; c.bokhove@soton.ac.uk

D. Adams
Faculty of Education, Institute of Educational Leadership, University of Malaya, Malaya, Malaysia
e-mail: donnieadams@um.edu.my

Z. Miao
College of Elementary Education, Jiangxi Normal University, Nanchang, China
e-mail: z.miao@jxnu.edu.cn

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_6

areas such as the economy, but education now seems an international commodity too, and this Chapter analyses the problems and potential of international studies of educational effectiveness, one of the areas of that commodity.

In the 1990s there were interesting explorations of individual educational policy areas from different societies that achieved some attention, such as the possible relevance to the United Kingdom of Swiss methods of mathematics teaching (Burghes & Blum, 1995), the popularity and usefulness of Japanese 'lesson study' methods of professional development for teachers and the espousal of 'whole class interactive' teaching methods in mathematics for use in England based upon observations and apparent success in Taiwan (Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002; Reynolds & Farrell, 1996).

There had also been attempts historically within certain societies to model their entire educational systems wholesale upon what were seen as successful practices elsewhere, as with the modelling of the Hungarian system of pre-school education upon those of Japan, the modelling on the English Literacy and Numeracy Strategies of the 2000s in some parts of the United States (Barber, 2007), and the modelling of Welsh language learning upon the methods pioneered in Israel in the 1960s (Reynolds, 2008).

There were however, many factors that hindered the internationalisation of education until relatively recently. Comparative education as a source discipline had eroded internationally in quality, influence and significance from the 1990s onwards, particularly in the United Kingdom where the number of University Departments in the area dwindled. Looking specifically at the field of educational effectiveness research (EER) that was probably the most rapidly developing of all educational specialties in and after the 1990s, it is true there was an international organisation (the International Congress for School Effectiveness and Improvement; ICSEI) and increasingly prevalent cross cultural analyses of the factors shown in the *national* research studies in different countries that were associated with effectiveness. However, this was not paralleled by a corresponding growth in the number of *internationally* based research studies – which involved common conceptualisation, operationalisation and measurement. Instead, what existed, simply, were international reviews of national studies in different countries (Reynolds et al., 2014).

In the last two decades this situation has been rapidly changing, largely because of the funding and sponsorship of large-scale international achievement studies by organisations such as the Organisation for Economic Cooperation and Development (OECD), particularly the Programme for International Student Achievement (PISA), but also shown in the World Bank's and IEA's commissioned literature reviews of educational effectiveness and those emanating from the European Economic Community too. Private sector companies, such as Pearson, have also been instrumental in generating influential publications (Barber, 2007; Mourshed, Chijioke, & Barber, 2010) that have become a reference point for much of the global debate about international 'best performance'. Interestingly, the historic concern within the Comparative Education community to understand national educational cultures has been replaced by the assumption that 'what works' are organisational arrangements largely independent of cultures.

The global economic and financial crisis and retrenchment of the late 2000s undoubtedly focussed renewed attention on issues of comparative educational performance. In times of greater financial scarcity, the ways in which societies can maximise their economic productivity by utilising their educational systems to generate more 'human capital' assumes greater importance, and the possible blueprints that may exist from more 'effective' educational systems took on greater salience than in times of rapid economic growth like the early 2000s. We now proceed to look at some of the studies that were conducted in the 1960–1990s, before looking in some detail at the well-known PISA studies that began later from 2001.

6.2 The First International Studies, 1960–2000

The international studies conducted during these four decades were numerous, as were commentaries, critiques and reviews. The International Association for the Evaluation of Educational Progress (IAEP) published some of these but the greatest number came out of the work of the International Association for the Evaluation of Achievement (IEA).

Briefly, the IEA conducted the First and Second Science Studies (Comber & Keeves, 1973; Keeves, 1992; Postlethwaite & Wiley, 1992; Rosier & Keeves, 1991); the First and Second Mathematics Studies (Husen, 1967; Travers & Westbury, 1989); the Study of Written Composition (Purves, 1992) and the Classroom Environment Study (Anderson, Ryan, & Shapiro, 1989). The IAEP conducted two studies of science and mathematics achievement cross-culturally (Keys & Foxman, 1989; Lapointe, Mead, & Phillips, 1989). The IEA also published the findings of the First Literacy Study (Elley, 1992; Postlethwaite & Ross, 1992). Reviews of the great majority of these studies are in Reynolds and Farrell (1996), Reynolds et al. (2002).

The results of the studies differed to an extent, according to the precise outcome measures being utilised whether it was reading/literacy, numeracy or science achievement. What was common though was a 'country against country' league table ranking perspective. However, it remained unclear which cultural, social, economic and educational factors were implicated in the country differences.

There were of course further 'core' problems that placed severe limitations upon the capacity of these international studies to generate valid knowledge about educational effectiveness. Some problems are present in all cross-national research, such as those of accurate translation of test material, of ensuring reliability in the 'meaning' of factors (such as social class or status indicators for example), of problems caused by Southern Hemisphere countries having their school years begin in January and of problems caused because of retrospective, potentially out of date material information being used. In certain curriculum areas, the cross-national validity of the tests utilised gives cause for grave concern and, as an example, the IEA study of written composition failed in its attempt to compare the performances of groups of students in different national systems that used different languages, since the latter study concluded that 'The construct that we call written composition must be seen in

in a cultural context and not considered a general cognitive capacity or activity' (Purves, 1992, p. 199).

Even the simple administration of an achievement test in the varying cultural contexts of different countries may pose problems, particularly in countries where the 'test mode' in which closed questions are asked in an examination-style format under time pressure may not approximate to students' general experience of school. By contrast, the use of this assessment mode within societies such as those of the Pacific Rim, where these methods are very frequently experienced, may facilitate country scores.

In addition to these core problems that affect all large-scale international effectiveness research, there were specific problems concerning the IEA and IAEP international effectiveness studies that represented virtually the totality of this international effectiveness research enterprise until the late 1990s.

6.2.1 Methodological Deficiencies

- The basic design of these studies, which were concerned with explaining country against country variation, may have itself been responsible for problems. Generally, a small number of schools each possessing a large number of students were selected, which made it difficult to make valid comparisons between schools once factors such as school type, socio-economic status of students and catchment areas were taken into account.
- Curriculum subjects were studied separately, making an integrated picture of schools and education in different countries difficult.
- There was considerable difficulty in designing tests which sampled the curricula in all countries acceptably, although the Trends in International Mathematics and Science Study (TIMSS) project expended considerable energy to ensure a geographical reach of items, and also published its results 'unadjusted' and 'adjusted' for the curriculum coverage or 'opportunity to learn' of individual countries.
- Cross sectional rather than longitudinal surveys precluded understandings of student progress, instead limiting findings to snapshots of student performance at specified ages.

6.2.2 Sampling Issues

- There were instances of large variations in the response rates that made interpretation of scores difficult in cross country comparisons.
- Sometimes samples of students used were not representative of the country as a whole (e.g. one area of Italy was used as a surrogate for the whole country in one of the IAEP studies).

- Variations between countries in the proportion of their children who could have taken part in the studies made assessment of country differences difficult. Mislevy notes that, whilst 98% of American children were in the sampling frame and eligible to take part in one study, the restriction of an Israeli sample to Hebrew-speaking public schools generated only 71% of total Israeli children being in eligible schools.

6.2.3 Limited Data and Limited Analyses

- In many studies there was a lack of information upon the non-school areas of children's lives (family and home environment) that might have explained achievement scores. Surrogates for social class utilised, such as 'number of books in the home', were not adequate.
- Outcomes data was collected mostly on the academic outcomes of schooling, yet social outcomes may have been equally interesting and important.
- The factors used to describe schools were overly resource-based (because of the greater perceived chance of obtaining reliability between observers in the former factors across countries, no doubt), in spite of the clearly limited explanatory power of the latter variables. At classroom level, only some studies (including TIMSS and Progress in International Reading Literacy Study; PIRLS) have used any measures of teaching and learning processes, with the use of videotapes of classrooms by the TIMSS project (Hiebert et al., 2003; Stigler, Gonzales, Kawanaka, Knoll, & Serrano, 1999) being particularly interesting, although of course rare.
- Only limited attempts were made to analyse the international samples differentially, for instance by achievement or by social class, with the exception of a limited amount of analysis by gender.

From all these points above, it is clear that the international studies of educational effectiveness of the IAEP and the IEA from the 1960s to the late 1990s necessitate some caution in interpretation. Not all studies possessed the same design, analysis and methodological problems, and no studies possessed all the design, analysis and methodological problems in total. But enough studies possessed sufficient problems to make firm and generalisable conclusions difficult and problematic.

The attention given to these international achievement surveys was by the 1990s considerable, although most accounts do not grant them the same importance globally as the more recent PISA studies (e.g. Waldow, Takayama, & Sung, 2014). There were, of course, also a limited number of critiques of the individual studies, and of the paradigm within which they were constructed (Alexander, 2010, 2012; Reynolds et al., 1994). Additionally, much of the discussion of the findings of these studies and their merits and de-merits remained within the academic community, rather than spreading widely into the news media as have the results of PISA. The globalisation phenomenon, facilitated by the spread of information technology

(IT), had not yet fully emerged in the 1990s to spread interest in what certain countries were doing educationally across all countries. And the financial crisis, as well as its effects in multiplying the pressures upon countries and politicians to pay enhanced attention to their economic and educational systems, did not influence policy until the late 2000s, perhaps with the exception of the global interest in the 1999 TIMSS video study.

6.3 The PISA International Achievement Studies, 2001 Onwards

PISA has spawned an extensive literature, including commentaries that emanate from a politically and socially critical perspective (e.g. Bulle, 2011; Dobbins & Martens, 2012; Eivers, 2010; Fischbach, Keller, Preckel, & Brunner, 2013; Gaber, Cankar, Umek, & Tašner, 2012; Grek, 2009; Hanberger, 2014; Kankaraš & Moors, 2013; Lewis, 2014; Morgan & Shahjahan, 2014; Sellar & Lingard, 2013a, 2013b; Waldow et al., 2014).

The main points made within the group of studies cited above are that:

- PISA studies have encouraged a convergence of policy intention and borrowing in educational policies and practices across the world;
- PISA has encouraged a positioning of international organisations – e.g. OECD, World Bank – in international educational matters that could be viewed as problematic because these organisations possess their own agendas;
- PISA has encouraged a 'one size fits all' response by policymakers that does not attach enough salience to the importance of local and national cultures;
- PISA is part of the phenomenon of the globalisation of the world's social and political structures, which has the potential to limit national, country-level influences.

In addition to these important perspectives, severe doubts have been expressed about PISA in terms of its methodological adequacy.

The core purpose of PISA has been stated by the OECD (2009) as follows:

> Are students well prepared for future challenges? Can they analyse, reason and communicate efficiently? Do they have the capacity to continue learning throughout life? The OECD Programme for International Student Achievement answers these questions and more, through its surveys of 15 year olds in the principal industrialised nations. (p. 1).

In practice, PISA seeks to achieve this through IT-based and 'paper and pencil' tests that are given to students aged 15 in different countries, in three achievement areas: mathematics, reading and science, with an emphasis not upon measuring student factual 'knowledge', but more upon measuring the capacity of students to apply that knowledge in real world situations, the so called 'skills based' approach. Using Item Response Theory (IRT) and based upon the assumption that a latent 'trait' determines all responses to test items, comparisons of students who have taken

different test items can be made based upon the statistical equivalence of the items, in terms of the data they produce. Additionally, surveys are conducted on students, Headteachers/Principals and parents, focussing on their attitudes and particularly on student and parental backgrounds/perceptions. The PISA surveys were published in 2001, 2004, 2007, 2010, 2013 and 2016. An additional outcome – problem solving – has been added to the existing three skill areas over time, and there has been a movement towards the increased measurement of 'metacognitive' skills shown first in the 2013 PISA use of specific items in this area in the reading test, and in the increased proportion of metacognitive items across the skill areas in the 2015 testing. Something called 'Global Competencies' is be measured from 2018 testing.

Unquestionably, PISA represents a major improvement upon the methodology and analyses that were characteristic of the previous group of studies from the late 1960s to the late 1990s. Other studies, like TIMSS and PIRLS, have also shown methodological advances over time. By focussing on the 'skills' to apply knowledge in real world settings rather than on the knowledge itself, PISA has theoretically made it easier to generate cross-cultural assessments, since there is likely to be smaller variation in the 'skills' aimed at in different societies than in the actual 'knowledge bases' that are taught to generate those skills, though this argument itself has been the focus of some criticism as the distinction between knowledge and skills is seen as artificial by some.

Also, PISA has paid significant attention to the rigour of the sampling process, especially to ensuring uniformly high response rates across countries, although securing such high response rates has frequently required the use of replacement schools, and concerns have been raised about PISA's method of calculating its response rate (Murphy, 2010). Attempts have also been made to ensure that various countries – like China – enter more representative portions of their national populations into the sampling frame, with the exclusion from the main sample of any countries having unsuitably low response rates (as in the case of England in the 2004 PISA survey). In addition, criticism has been levelled about gender imbalances in some country samples and the varying levels of exclusions due to intellectual impairment or special educational needs (Wuttke, 2007).

PISA has also been particularly committed to its attempts to measure the effects of the 'macro' level of national level policies related to the governance, funding, organisation and management of education within different jurisdictions, a marked contrast to the earlier IEA studies that collected very little data in these areas.

PISA has also opened up – in the publicity it has had – an important debate about educational excellence and equity that has been generated by participation of many different constituencies and groups around the globe.

Finally, the historic drift of the design of the PISA studies represents a sensible reflection of how international emphases and knowledge in areas associated with educational effectiveness have been changing over time. More socio-economic background data on students is being collected in the more recent PISA studies, reflecting enhanced interest in the interactions between schools *and* their communities, although there remain concerns that data on structural inequality in different countries is still not fully reflected in the PISA data collection process and in its

analysis (Eivers, 2010). The data that have been collected on the levels 'above' that of the school, such as the 'meso' level of the District/Local Authority and the 'macro' level of national educational policies on such areas as accountability, assessment, governance and funding, reflect the increasing attention that is being given to educational policy issues within educational research. The movement towards the measurement of metacognitive skills is also sensible, given their increased salience in the recent international assessment literature (Muijs et al., 2014).

In many important respects, however, PISA may not have succeeded completely in addressing all the concerns of the critics of the earlier studies in important areas of research design, research methodology and data analysis. There are also additional specific issues to do with PISA's specific use of its data and its data reporting.

A particular focus recently has been upon five 'core' issues where PISA appears vulnerable to criticism:

- It is crucial that tests used in work like PISA should be 'culturally fair', with students of the same level of achievement receiving the same assessment test scores independently of the country they come from. Early suggestions from small scale studies were that measurement equivalence may have been limited (Allerup, 2007; Goldstein, 2008; Grisay & Monseur, 2007; Yildirim, 2006). However, the recent study of Kankaraš and Moors (2013), based upon 2009 OECD data from 64 countries and 475,460 15-year-olds, found that "equivalence occurred in a majority of test questions in all three scales researched and, is on average, of moderate size" (p. 1). Given the fact that the PISA tests are translated and administered in a large number of countries and in cultures that are diverse linguistically, socially and economically, it could be expected that some differences in interpretation and understanding might occur. However, the extent and degree of the inequivalence across datasets, as stated by Gorur and Wu (2014), suggests that it actually "impairs the validity of country comparisons" (p. 17).
- There is evidence that the OECD's written reports on their data may not be totally in accordance with the findings of their own data - in other words, that there may be mis-reporting or instances of cherry-picking some findings over others. Whilst considerable attention has been historically drawn in PISA publications to positive 'key findings' for 'demand side' national policies involving school autonomy, competition and accountability, secondary analysis of the 2007 PISA data showed that other consequential and related accountability policies such as using student achievement data to evaluate teachers and to allocate resources were associated with *worse* student performance (Murphy, 2014). PISA applied more consistent country correlation methods to the 2010 study than in previous years, but negative correlations involving achievement and some accountability policies – for example, the use of student achievement data to allocate resources -- were left unremarked-upon in tables in an annex to the relevant report. Secondary analysis showed this policy to again be associated with *worse* student performance on all measures of performance across all countries in the study (Murphy, 2014), and on mathematics within the OECD countries.

- There have also been concerns about sampling issues, such as variation in the participation of special schools in some societies and not others, and the possible effects of differential country response rates being determined by Principal/Headteacher enthusiasm for the PISA testing process (Meyer & Schiller, 2013). Some other design and methodological limitations related to cultural differences include the extent to which students from different countries take 'non-PISA' tests and how far they take 'tests' seriously. Sjoberg (2012) argues that students in different countries and cultures understand tests differently and place different emphases upon their importance, such differences being a by-product of deep historical cultural influences and differences. He also argues that the desire to do well in 'the test' and persisting to complete all of it is uneven across cultures and could be a variable that explains differences in outcomes and performance.
- Most importantly, it still seems possible for governments to 'game' PISA, as with the early use of 'Shanghai' as a surrogate for all China when in fact it is a highly atypical region, both educationally and socio- economically. In Shanghai, it has been argued, that many migrant children have not been systematically involved with PISA testing, and that over half of all 15 year olds may have been excluded from testing because of the effect of the 'hukou' identity card or passport system (Loveless, 2013, 2014). The 'hukou' controls access to Municipal services and those Chinese with rural 'hukou' may be discriminated against in the urban environment, for example by being banned from public schools and put into migrant schools. Also, there are a proportion of migrant children who are left behind in villages as their parents migrate to Shanghai, and a further proportion of migrant children leave school before the age of 15 when children are meant to take the PISA tests. Both these factors may generate an unrepresentative sample of children in the Shanghai PISA sample. A larger sample of Provinces has been used in 2015 and 2018 testing, however.
- There are also methodological issues of concern, some of which relate directly to PISA's overarching goals. The OECD states that tests are designed to assess to what extent students at the end of compulsory education "can apply their knowledge to real-life situations and be equipped for full participation in society" (OECD, 2015). This is apparent in the nature of the content where 'real life' is demonstrated by the heavy use of contexts in, for example, maths and science assessment items. This might be problematic for two reasons. Firstly, contexts are often culturally rooted. Some contexts might be more appropriate for certain countries; for example, assessment items about bicycles and sailing ships might be less appropriate for some countries. Secondly, the context often adds additional reading demands to the assessment items. A valid question to ask, then, is whether test items might measure more than just maths and science constructs, which might weaken the validity of the measurements by creating additional variance irrelevant to the intended construct (Eivers, 2010; Messick, 1998).

The issue of reading load and difficulty has also been raised by Bodin (2007), who states that it is unclear whether difficulties in PISA's mathematics items are caused by the underlying text or the mathematical problem's degree of difficulty. A

'relationship' report from the IEA combined data from their maths and science test (TIMSS) and literacy test (PIRLS) and indicated that reading difficulty was associated with achievement (Mullis, Martin, & Foy, 2013). Results varied from country to country and even between mathematics and science within countries, yet there was overall support for the idea that higher reading demands can make the fourth grade TIMSS items more challenging for weaker readers. It is possible that this point applies to PISA, too, and indeed Ruddock, Clausen-May, Purple, and Ager (2006), noted that "the higher reading demand of questions in PISA is often accompanied by a relatively lower demand in the mathematics or science required" (p. 123).

Methodological challenges are not only apparent in PISA's assessment items, but also in several of the assumptions underpinning the statistical methods used. One criticism concerns the use of the Rasch model (Kreiner & Christensen, 2014). The OECD clearly explains its methods in the PISA technical manual, but the underlying assumption of it is that all the questions used in the study would have to function in exactly the same way (be equally difficult) in all participating countries. Given the cultural differences, but also differing curricula, this seems an unlikely assumption. If indeed the approach includes removing outlier items because of a 'unidimensional' requirement, this might create cultural bias.

6.4 PISA: A Perspective from Educational Effectiveness and Improvement Research (EEIR)

In addition to the range of recent methodological concerns and criticisms we have outlined above, our extensive historic bodies of knowledge, methodology and agreed approaches within EEIR suggest a further set of limitations that must be placed upon the utility, reliability and validity of the PISA studies. We outline these now, before concluding with a plea for, and some suggestions about, the generation of an enhanced focus upon differently conceptualised international effectiveness studies to be undertaken within the general field of EEIR.

6.4.1 The Absence of Teaching/Pedagogical Focus

All existing reviews of research conducted within EEIR have argued for the primacy of teacher effects and for these effects to be bigger than the 'levels' of the school and of the District/Local Authority. However, PISA collects no data upon the methods of teaching used in different countries, focussing upon the organisational arrangements of classrooms within schools more than on the actual behaviours of teachers and support personnel that have been shown to be highly important in the five decades of research on what can be called 'teacher effectiveness' (Brophy, 1979; Brophy & Good, 1986; Muijs & Reynolds, 2011).

Clearly, this absence of classroom data is because individual teachers would find it hard to rate and describe their own teaching in ways that could command cross cultural validity, and alternatively sending researchers to observe the classrooms of teachers in different countries would be hugely time intensive and expensive.

While these are sound reasons, it remains the case that the absence of any major focus upon pedagogy may well function both to severely limit the capacity of PISA to understand the causes and nature of country differences in educational effectiveness and also to imperil the prospects of success of any policies that may be tried out in different countries based upon PISA findings, given that it is likely to be 'teaching,' the 'alterable variable' with the largest likely effects, that countries may well wish to influence and improve. It may also be that the continued concern with the managerial arrangements of schools, Districts/Local Authorities and national policymaking/policies promoted by those who draw upon PISA to offer policy guidance, rather than with pedagogical practices, may not resonate with practitioners who are more interested in the classroom itself rather than the organisational layers that sit above it. However, the linkage between PISA and teacher behaviours being trialled in the Teaching and Learning International Survey (TALIS) suggests some progress in this area.

6.4.2 The Limited Use of a 'Value Added' Approach

EEIR has, over time, successfully established its bodies of knowledge about 'what worked' in generating its described 'effective schools' and 'effective teaching practices' (see reviews in Reynolds et al., 2014, for example), and generally enjoyed much practitioner and policymaker enthusiasm for the knowledge bases as they began to appear after the 1990s. Over those years, it became axiomatic that in order for knowledge to be reliable and accurate concerning the nature of the important school and classroom effectiveness factors, the 'raw' achievement results of schools should be made reflective of the variation in the nature of the intakes that went into the particular schools and classrooms, in order for better estimates of the 'value added' by particular educational settings to be available to correlate against their educational practices. This is what generated the effectiveness knowledge base. Using 'raw', non-value added measures of achievement that do not distinguish between the contribution of educational factors and the contribution of non-educational factors like socio economic status, parental attitudes, and cultural factors linked to differences between schools in their intake characteristics, risks invalid conclusions.

This is what PISA has done. In the earlier PISA studies there were limited measures of non-educational factors employed, so there was no systematic attempt to control out the major differences in social, cultural, environmental and economic factors in the wide range of countries utilised as the sample. Effectively, by default, the assumption was being made that it was educational factors solely that were involved in determining country differences.

More recently, by the time of the 2010 study, PISA was in fact also using 'national income per head of employed population' to provide a measure of the quality of what the educational systems of different countries were receiving as intakes, but only 6% of the differences in average student performance were due to GDP per capita (Reynolds et al., 2015), suggesting the need for other factors to be used. A wider range of socio-economic background and attitudinal data on students and parents were used in and after the 2013 studies, but the continued absence of any student achievement or student ability measure as controls means that even this wider range of social background factors were probably not functioning adequately to control out non-educational factors and influences. Although more socio-economic background data on students has been collected in recent years, the attempt has not been made to use these to generate 'value added' measures of relative country performance to supplement the much publicised and prevalent raw 'league tables'.

It is interesting that the PISA 2013 and 2015 studies do indeed show considerable recognition of the very important role played by non-educational factors in the determination of achievement outcomes, evidenced in the large number of analyses presented as to how all countries perform in the educational achievement of their lower socio-economic status groups of students. This attention given to 'equity' of performance within different countries, rather than merely 'excellence', is much to be welcomed and parallels the emphasis upon the differential effectiveness of schools shown within the EEIR community, particularly in research from Continental Europe in the last 15 years (Reynolds et al., 2014).

But if the effect of home social background and other non-school factors upon children's educational prospects *within* all different PISA societies deserves attention, these should surely also deserve to be used as factors in the analysis of differences *between* societies, by making allowance for background effects upon country achievement scores. Educational effectiveness research made its rapid progress in understanding how schools had their effects only *after* it had adopted 'value added' perspectives, particularly those involving multilevel analyses (Goldstein, 2003) and reflecting other methodological advances in the field (Creemers, Kyriakides, & Sammons, 2010). It is suggested that 'raw' achievement data is useful given that countries can see in *absolute* terms how they are doing with the development of their human capital, but that 'value added' or 'relative' data may be very useful too.

6.4.3 The Absence of a Longitudinal Research Design

PISA uses a cross sectional research design, whereas in contrast, EEIR now regards it as best practice and axiomatic to use longitudinal research designs. No doubt reasons of cost may be the explanation for use of the current PISA research design, given that it involves one testing point only. Also, to wait for differences in the gains of a cohort passing through schools in different countries to appear would certainly

be half a year, and even then these 'gain' differences *over* a short time would be nowhere near as large as differences between countries *at a point in time* that have evolved, making 'at a time point' the preferred option.

However, longitudinal studies following students for a year have been profitably used in exploratory work in this area (Reynolds et al., 2002, 2014), and a longitudinal research design would be helpful for PISA in two ways. Firstly, it would permit a more valid exploration of the effects of non-educational determinants of achievement – even if (as at present) only socio-economic background factors were measured rather than also using a prior achievement measure, the moderate correlations likely between all of these factors and achievement means that the levels of achievement at the first stage of the two stage testing involved in a cohort design ('pre' and 'post') would reflect the influence of these socio-economic factors. The gain over time in different countries that remained after 'stripping out' the start scores would be highly likely to reflect educational influences, thus generating more valid, 'true' educational effects.

Secondly, the study of the same children over a given time period would be likely to increase our understanding of the complex interactions between schools, educational systems and their children. Following the same children over time, with repeated visits made necessary by the need to do 'pre', and 'post' testing, does not necessarily improve our understanding of educational processes – as 'one off' events, how can they? But this approach does make more possible the collection of longitudinal data on student school experiences that are likely to give greater understanding of educational processes, and explain more variance. Interestingly, the effects of educational factors in the cohort studies following the same children over a longer period of time are typically much higher than those from cross-sectional work, or from those longitudinal studies undertaken over a short period of time (e.g. Guldemond & Bosker, 2009).

6.4.4 *The Use of Educational Policy/Educational Process Factors of Limited Explanatory Power*

It is important to note firstly that 'supply side' policies – concerned with teacher professional development, or national level programmes to build capacity for example – are utilised in PISA much less than those related to the 'demand side'. All things being equal, 'demand side' effects are likely to feature more strongly than the 'supply side' in the explanations of the success/failure of educational policies, then.

More importantly, this effort put into the data collection in the 'demand side' 'macro' policy areas may not be particularly useful in explaining variance. For example, the United States 'No Child Left Behind' Act (NCLB) of 2001 requires States to have accountability systems which typically involve State-wide testing for all children in grades 3–8, the disaggregated reporting of data on student

performance and the employment of sanctions when student performance is poor. Hanusheck and Raymond (2005), show an effect of only 0.2 of a standard deviation (using individual State data), on test scores. Dee and Jacob (2011), report a 0.5 student standard deviation impact of NCLB on student Maths scores, but no impact upon Reading scores. The Burges, Wilson, and Worth (2010) report on the effect of the national regime in Wales that abolished the publication and consequential use of the individual school national performance tables finds, after stripping out socio-economic factors by matching schools in 'experimental' Wales with 'control' England, that this is equivalent to a 0.23 of a (school level) standard deviation negative effect. All these studies suggest low effects for any of the 'demand side' policy levers.

The enthusiasm shown within PISA for collection of large amounts of data upon a limited range of 'demand side' policies may be understandable, given that the OECD wishes to influence the practices of policymakers, but it may be that some of the 'supply side' factors should interest them too to them rather than being consigned to the TALIS programme alone.

6.4.5 The Absence of an Efficiency Perspective

PISA has focussed primarily upon an 'effectiveness' perspective related to attempting to explain and understand national differences in the output of educational achievement. In doing this it has paralleled the EEIR field closely, which itself has focussed more on 'effectiveness' than 'efficiency'.

The 'efficiency' of countries, in the sense of the scale of the material 'inputs' that are necessary to generate the 'outputs' of the effectiveness levels shown in different societies, have so far received little attention in PISA data collection, with the exception of the expenditure levels of different societies being shown as unrelated to overall country variation in achievement test scores. Such a finding is not surprising, of course. About 80–90% of the variation in 'per student expenditure' is due to variation in the pay of teachers, which in PISA is then expressed on a linear scale. But since the overall individual country level of national income is closely related to national individual country teacher pay, this finding only tells us that national income per head is not related to student achievement. It does not tell us that other 'efficiency' measures may be unimportant. Indeed, more recently a study by Dolton, Gutierrez, and Still (2015), did relate PISA scores to a range of financial inputs, finding two – teacher salaries and class size – to be significant, creating an 'efficiency index' that is essentially a measure of how highly a country's pupils score on PISA given how much (or little) a country spends on its teachers. This method provides a useful starting point for looking at the question of efficiency. However, it also further illustrates some of the limitations of PISA data (i.e. the difficulty of causal attribution from cross-sectional datasets) and the misuse of the data to which PISA has been prone, through (again) creating a league table notwithstanding the measurement error and the overlapping confidence intervals involved, and through

simplistic policy advice. In general, of course, efficiency is important, and therefore both the use of production functions taken from economics to study efficiency, and the use of cost-benefit analysis when looking at particular interventions or policy changes, would be beneficial to the study of education as a whole.

One 'efficiency' measure that was also strangely neglected in PISA is 'time', in terms of the 'inputs' of time that students in different countries are exposed to in their instructional activities. Time is an international measure that has a common metric and means the same in all countries. An hour of time is exactly the *same* in Oman as it is in Shanghai (whereas an hour of teacher's *pay* is highly variable depending on the country setting). Exposure to this thing called 'time' is highly variable cross culturally – there is a range in the days of schooling per student per year in different societies of from 230+ at the top to perhaps 160 days at the bottom, with some Pacific Rim societies scoring particularly highly on this. Is it possible that this efficiency measure may be related to country scores?

If one were to add an 'hours in school' to a 'days in school' measure, then country differences may widen. In addition to 'time in school', there also is some scale and impact of private supplementary tutoring, often referred to as 'shadow education'. Shadow education (Bray & Kobakhidze, 2014) has expanded significantly worldwide and is now recognised to potentially have considerable significance. Thus measures of the duration and intensity of schooling and also additional learning outside school via tutors/private tutoring schools need to be recognised as of potential influence.

Whether one stays with financial factors as 'inputs', or adopts additional non-financial ones like time, an 'efficiency' perspective may be useful for PISA, with its policymaking evidence, to employ.

6.4.6 *The Absence of National Cultures and Context in the Analysis of Effectiveness*

The central assumption underlying PISA is that national educational structures and policies – and by default only these things – explain global variation in students' academic performance (Feniger & Lefstein, 2014). In other words, there is an underlying assumption that cultural factors or features play little or no part in explaining differences in relative country performance.

Explanations of success in comparative work – as in any educational research work – are dependent on a number of inter-related factors, but studies such as PISA may produce analyses that are not sensitive to the *relationality* of the phenomena being studied. Consequently, many of the factors that might affect educational performance, particularly those that are culturally defined or contextually shaped, are not included or captured in existing PISA analyses. This leads to what Gorur and Wu (2014) have nicely termed the 'problem of the unmeasured' and the fact that

drawing any meaningful parallels or conclusions about cross-national performance from the existing PISA studies will be difficult, if not impossible. Pereyra et al. (2011, p. 261) aptly argue that 'PISA is a brilliant big-social science mapping of outcomes but in no anthropological, historical or cultural sense is it comparative work'.

There are a number of specific issues that result from the absence of cultural and social contexts as factors within PISA:

- Unmeasured cultural/social factors may have effects on country achievements that are at the moment explained by default as due to educational factors. Examples of this, from research outside PISA, involve the stress on ambition related to the simultaneous 'internal/external' loci of control on Asian children (Reynolds et al., 2002), the value given to education through positive parental perceptions of literacy as a goal in Finland (Sahlberg, 2011), or the historical enthusiasm – verging on idolatry – for education in the Welsh society of the mid-twentieth century, formed by religion, socialist policies and the effects of a self-educated working class (Reynolds, 2008).
- Without an understanding of context we cannot know 'how' any possible educational policy factors have their effects, and causal inferences and attributions cannot be made. At present, romanticised accounts of how policies have their apparently dramatic effects dominate discussions (e.g. Mourshed et al., 2010), as Zhao (2014) notes, and there is no explanation of the possibility that sets of educational policies and educational factors are *differentially* effective across societies in accordance with country cultures and social structures, in other words that there may be an interaction between societies and their educational systems.
- PISA reflects the view that the *same* educational/school/policy factors are effective everywhere independent of context, but one pilot study (Reynolds et al., 2002) found that the same effectiveness factors only 'travelled' across their sample of eight countries at the level of the classroom, where the effects of detailed factors such as structured teaching, high expectations and the other factors from the teacher effectiveness literature all appeared important in all the different societies studied. By contrast, details and operationalised characteristics of the school level factors and educational policy factors associated with effectiveness were very *different* in different societies, raising the possibility that high levels of country achievement may be generated by factors that are *different* at the higher levels of educational systems. Do different countries need different systems – in accordance with their cultures – to generate teaching and classrooms which, because of the nature of children's physiology internationally, should be the same if they are to be effective? We do not know.

The analyses of PISA pay little attention to cultural and contextual factors that may partly explain the differences in the performance of education systems (Harris & Jones, 2015). As Feniger and Lefstein (2014) underline, "we need a much better

understanding of comparative cultural contexts" to explain relative educational performance and outcomes. Without such understanding, the educational policies, strategies and interventions associated with, and indeed endorsed by PISA and the OECD, will continue to be founded on scientific sand.

6.5 Conclusions: The Potential Value of Improved International Effectiveness Research

We have seen so far that international surveys of educational achievement have been receiving more and more attention in the last decade. One early analysis found 322 publications from 2002 to 2010 on PISA in the three main scientific databases for the social sciences: ERIC, EBSCOhost and the ISI Web of Science. This number would miss those that have appeared since, and the considerable number of articles in the 'grey' area of magazines, web sites, think tank publications and the like. Additionally, the volume of material about PISA in newspapers and in other media is extensive. Studies by the IEA, that have continued to be published, also receive extensive coverage. A more recent review moves discussion into a more considered and comprehensive direction (Hopfenbeck et al., 2016).

What *has* been pursued however, as comparative work by EEIR, are mostly studies that are done in multiple countries, collecting data on a defined issue, or problem, or factor, which *purport* to be cross cultural or comparative work but which lack across nation methodological communality in conceptualisation, operationalisation and measurement (that PISA does show considerably, of course). We have in EEIR also just lots of national reviews, not a comparative knowledge base.

It would be wrong to omit to mention here the efforts of those who have attempted to do cross cultural work from a 'school improvement' perspective, particularly those concerned to generate strategies that may be useful for system reform internationally (e.g. Barber, 2009; Fullan, 2009; Hargreaves & Shirley, 2009; Harris & Chrispeels, 2008; Hopkins, 2007). A number of interesting country system-level studies have been published, such as those of Whelan (2009). Further, improvement researchers have generated some analysis of context specificity whereby 'what works' for improvement is seen as variable at different growth states for schools (Harris & Jones, 2015; Hopkins, 2013; Hopkins, Harris, & Jackson, 1997; Hopkins & Reynolds, 2001).

Much of this work is also sensitive to local country context, and attempts to avoid a 'list of ingredients' approach, preferring a 'recipe' based upon a different 'mix' of approaches to system level reform in different countries (Hopkins, Stringfield, Harris, Stoll, & Mackay, 2014). However, it is sometimes unclear in many school improvement studies what data exist to support their formulations, how they have been analysed, how intensive has been the immersion of the individual 'improvers'

in their different societies and whether the understandings of the highly complex interactions between systems, countries and cultures are quite advanced enough.

Nowadays, therefore, the pressures on us in the EEIR field to create better research and better explanations in the field of cross cultural/comparative work are immense, given PISA's partial contribution as outlined above. Educational knowledge is travelling now with same rapidity as other bodies of knowledge – in medicine, industry and commerce – because of the effects of the information society created by the pervasiveness of the internet. There is much evidence that the world benefits by the spread of good practice in non-educational fields – by spreading best practice in automobile engineering, by medical best practice helping to deal with infections that stalk the planet, and by management systems such as 'just in time' or 'big data' or 'simultaneous top down and bottom up'. However, we now need to ensure that 'what travels' in the area of education is also valid and reliable because of the considerable practical and intellectual benefits of comparative work in our field:

- There is a real danger of educational damage when the interpretations of big data sets like PISA are associated with the attempted transplant of factors from country to country independent of the cultural contexts of these countries.
- Culturally sensitive explanations of 'what works' and 'why' in international studies can improve the quality of educational discourse more generally, generating more nuanced, complex sets of understandings that have greater explanatory power by being generated in multiple different national settings.
- Internationally based studies tap the full range of variation in school and classroom effects. *Within* any country, the range of school and teaching factors in their 'quality' and their 'quantity' is likely to be much smaller than *between* countries. Put simply, therefore, international studies are likely to show *greater* educational effects than within nation ones which have settled into a range of perhaps 10–20% in variance explained (Chapman, Muijs, Reynolds, Sammons, & Teddlie, 2015). The true power of school and classroom is, if this reasoning is correct, only likely to be shown by authentic international comparative work.
- Internationally comparative work helps in the generation of theory, where we still have an absence of theorising of even a middle-range variety, although some recent attempts may suggest this to be changing (Creemers & Kyriakides, 2008). Why is it that 'assertive Principal leadership' does not predict school effectiveness status in the Netherlands but does in Anglo Saxon societies (Bosker & Scheerens, 1994)? Why do some of our factors travel across socio economic contexts within countries better than others (Teddlie & Stringfield, 1993)? Answering these questions forces us to develop a deeper analysis, more complex explanations and more multi layered interactions between 'levels' of education which have theoretical potential;
- If comparative work were to show that 'what works' varies within different contexts, it would compel us to generate more sensitive, contextually specific explanation in our field. In the early days of EEIR, we largely researched in low socio-economic status communities, rendering ourselves unable to see if there was within-nation contextual variation in effective educational process factors.

Later work, particularly in the United States, found these differences, with the particularly interesting finding of effective schools in poorer Californian communities actively pursuing policies to dis-involve their parents (Hallinger & Murphy, 1986; Teddlie & Stringfield, 1993). The District/Local Authority context-specific policies necessary to improve in socially challenged communities were a focus also in the UK (Harris, Chapman, Muijs, Russ, & Stoll, 2006; Muijs, Harris, Chapman, Stoll, & Russ, 2004). This tradition was largely eroded by the 2000s, perhaps due to the across-context 'lists' (e.g. Reynolds et al., 1996), that were a feature of the 1990s and which were an attempt to extract 'what works' from the context-bound early studies, and which also sometimes arose out of government sponsored and funded projects where researchers might have seen any context specific formulations as 'inconvenient' to conventional 'one size fits all' educational policies.

For the policymaker/political sponsors of EEIR, it may also be that the absence of context-specific effectiveness research also reflected the desire to ape medical research by generating universal findings that could be applied in all contexts, as in Slavin's (1996) use of the phrase 'wherever and whenever we choose'. This is of course to fundamentally misunderstand medical research and practice, since the latter involves 'universal' treatments (e.g. statins) but applied in highly context-specific fashions to individual patients (with variation in the type of drug used, the dosage, the length of use, the possible combination with other drugs, and the sequence of combination if it was used). This context specificity in medicine has been ill understood. Another factor limiting context-specific formulations may have been the popularity of meta-analyses which avoided splitting up samples by context in the interests of maintaining high sample size (e.g. Hattie, 2009).

To conclude on a positive note, there is of course much that the EEIR community can do to generate high quality research into International Educational Effectiveness, and the following things seem to be particularly useful:

- The OECD has made available individual student level PISA data to facilitate research;
- EEIR researchers have been involved in conducting PISA studies, providing advice on issues such as questionnaire design and use, and providing advice on methodological issues;
- EEIR shows interest in interacting with the existing international comparative work on the effectiveness of international systems (Kyriakides, Giorgiou, Creemers, Panayiotou, & Reynolds, 2018). Existing studies like PISA and TIMSS use a repeated series design approach, so it is possible to add educational factors like national policies in educational areas and changes in these policy areas to the student achievement data, to test the extent to which changes in these factors are associated with the effectiveness status of different country systems;
- EEIR shows recent interest (Chapman et al., 2015) in conceptualising and measuring social, in addition to academic, outcomes and there are increasing hints that these may be highly relevant to the apparent economic success of "high performing" societies;

- The OECD is gathering 'teacher' and 'teacher behaviour' data in the current TALIS, an area in which EEIR researchers have become more interested in the last decade;
- Small scale studies that use routinely available existing international social and economic data to consider the contextual effects upon educational achievement of different countries are powerful and profitable (Kelly, 2018). In this work, contextual factors such as targeted expenditure on education relate to overall PISA country scores, more than does overall expenditure. Likewise, mean PISA scores and 'resilience' in children (being in the bottom quarter socio economically but in the top quarter on achievement) are closely related;
- The picture of high achieving countries that has begun to emerge suggests more complex explanations than early formulations, focusing on cultural factors and pedagogical practice (Deng & Gopinathan, 2016). Disentangling these influences would seem entirely appropriate for researchers and practitioners in the EEIR paradigm.
- Methodological advances in the EER field (Creemers, Kyriakides & Sammons, 2010) including interest in mixed methods research may provide a fruitful way forward to explore in more detail within and between country school and classroom differences in student outcomes of various kinds (not just academic) and exploring the role of differences in culture and educational systems and their associations with variations in outcomes in new and creative ways (Sammons, 2010). It could also provide a richer picture of *within as well as between* system variation including country, regional and contextual perspectives, and within school variation too.

We would suggest with others (Rutkowski & Rutkowski, 2016) that it is time for EEIR and those responsible for international surveys and studies to suspend any possible turf wars and work collaboratively to deliver the promise of International Effectiveness Research.

References

Alexander, R. (2010). *Culture and pedagogy: International comparisons in primary education*. Oxford, UK: Basil Blackwell.

Alexander, R. (2012). Moral panic, miracle cures and educational policy: What can we really learn from international comparison? *Scottish Educational Review, 44*(1), 4–21.

Allerup, P. (2007). Identification of group differences using PISA scales – Considering effects of inhomogeneous items. In S. Hopmann & G. Brinek (Eds.), *PISA according to PISA*. Wien, Austria: Lit-Verlag, University of Vienna.

Anderson, L. W., Ryan, D. W., & Shapiro, B. J. (1989). *The IEA classroom environment study*. Oxford, UK: Pergamon Press.

Barber, M. (2007). *Instruction to deliver: Tony Blair, the public services and the challenge of delivery*. London, UK: Politico's.

Barber, M. (2009). From system effectiveness to system improvement. In A. Hargreaves & M. Fullan (Eds.), *Change wars*. Bloomington, IN: Solution Tree.

Bodin, A. (2007). What does PISA really assess? What does it not? A French view. In S. Hopman, G. Brinek, & M. Retzl (Eds.), *PISA according to PISA: Does PISA keep what it promises?* Wien, Austria: LIT.

Bosker, R., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research, 21*, 159–180.

Bray, M., & Kobakhidze, M. M. (2014). Measurement issues in research on shadow education: Challenges and pitfalls encountered in TIMSS and PISA. *Comparative Education Review, 58*(4), 590–620.

Brophy, J. (1979). Teacher behaviour and its effects. *Journal of Educational Psychology, 71*(6), 733–750.

Brophy, J., & Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Bulle, N. (2011). Comparing OECD educational models through the prism of PISA. *Comparative Education, 47*(4), 503–521.

Burges, S., Wilson, D. & Worth, J. (2010). *A natural experiment in school accountability: The impact of school performance information on pupil progress and sorting*. CMPO Working paper series no. 10/246. Bristol: CMPO.

Burghes, D. & Blum, W. (1995). *The Exeter Kassel comparative project: A review of year 1 and year 2 results in Gatsby Foundation*. In Proceedings of a seminar on mathematics education. London, UK: Gatsby.

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., & Teddlie, C. (2015). *The international handbook of educational effectiveness and improvement*. London, UK: Routledge.

Comber, L. C., & Keeves, P. (1973). *Science education in nineteen counties*. London, UK: Wiley.

Creemers, B., Kyriakides, L., & Sammons, P. (Eds.). (2010). *Methodological advances in educational effectiveness research*. London, UK: Routledge Taylor Francis.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK: Routledge.

Dee, T., & Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418–446.

Deng, Z., & Gopinathan, S. (2016). PISA and high performing education systems: Explaining Singapore's educational success. *Comparative Education, 52*(4), 449–472.

Dobbins, M., & Martens, K. (2012). Towards an education approach à la finlandaise? French education policy after PISA. *Journal of Education Policy, 27*(1), 23–43.

Dolton, P., Gutierrez, O. M., & Still, A. (2015). *Educational efficiency: Value for money in public spending on schools* (Paper No. CEPCP 441).

Eivers, E. (2010). PISA: Issues in implementation and interpretation. *The Irish Journal of Education/Iris Eireannach an Oideachais, 38*, 94–118.

Elley, W. B. (1992). *How in the world do students read?* Newark, DE: IEA.

Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Education Policy, 29*(6), 845–855.

Fischbach, A., Keller, U., Preckel, F., & Brunner, M. (2013). PISA proficiency scores predict educational outcomes. *Learning and Individual Differences, 24*, 63–72.

Fullan, M. (2009). Large scale reform comes of age. *Journal of Educational Change, 10*, 101–113.

Gaber, S., Cankar, G., Umek, L. M., & Tašner, V. (2012). The danger of inadequate conceptualisation in PISA for education policy. *Compare: A Journal of Comparative and International Education, 42*(4), 647–663.

Goldstein, H. (2003). *Multilevel models in educational and social research* (3rd ed.). London, UK: Edward Arnold.

Goldstein, H. (2008). Comment peut-on utiliser les etudes comparatives international pour doter les politiques educatives d'information fiables? How can international comparative studies be used to provide reliable educational information policies? *Revue Francaise de Pedagogie/French Review of Pedagogy., 164*, 69–76.

Gorur, R., & Wu, M. (2014). Leaning too far? PISA, policy and Australia's 'top five' ambitions. *Discourse: Studies in the Cultural Politics of Education, 36*(5), 647–664.
Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy, 24*(1), 23–37.
Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation, 33*, 69–86.
Guldemond, H., & Bosker, R. (2009). School effects on student progress – A dynamic perspective. *School Effectiveness and School Improvement, 20*(2), 255–268.
Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education, 94*, 328–355.
Hanberger, A. (2014). What PISA intends to and can possibly achieve: A critical programme theory analysis. *European Educational Research Journal, 13*(2), 167–180.
Hanusheck, E., & Raymond, M. (2005). Does school accountability lead to improved school performance? *Journal of Policy Analysis and Management, 24*(2), 297–327.
Hargreaves, A., & Shirley, D. (2009). *The fourth way*. Thousand Oaks, CA: Corwin.
Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: Exploring the possible. *School Effectiveness and School Improvement, 17*, 409–424.
Harris, A., & Chrispeels, J. (Eds.). (2008). *International perspectives on school improvement*. London, UK: Routledge Falmer.
Harris, A., & Jones, M. (Eds.). (2015). *Leading futures*. Singapore: Sage.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.
Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., & Jacobs, J. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: National Centre for Educational Statistics.
Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2016). Lessons learnt from PISA: A systematic review of peer reviewed articles on the programme for international student assessment. *The Scandinavian Journal of Educational Research, 62*(3), 333–353.
Hopkins, D. (2007). *Every school a great school*. Maidenhead, UK: Open University Press.
Hopkins, D. (2013). *Exploding the myths of school reform*. Maidenhead, UK: Open University Press.
Hopkins, D., Harris, A., & Jackson, D. (1997). Understanding the school's capacity for development: Growth states and strategies. *School Leadership and Management, 17*, 401–412.
Hopkins, D., & Reynolds, D. (2001). The past, present and future of school improvement: Towards the third age. *British Educational Research Journal, 27*, 459–475.
Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: A narrative state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 257–281.
Husen, T. (Ed.). (1967). *International study of achievements in mathematics, volumes one and two*. Stockholm, Sweden: Almquist and Wiksell.
Kankaraš, M., & Moors, G. (2013). Analysis of cross-cultural comparability of PISA 2009 scores. *Journal of Cross-Cultural Psychology, 45*(3), 381–399.
Keeves, J. P. (1992). *The IEA study of science 111: Changes in science education and achievement, 1970 to 1984*. Oxford, UK: Pergamon Press.
Kelly, A. (2018). *PISA – Explaining new contextual factors at the level of the nation state*. Unpublished paper. University of Southampton, Southampton, UK.
Keys, W., & Foxman, D. (1989). *A world of differences (a United Kingdom perspective on an international assessment of mathematics and science)*. Slough, UK: National Foundation for Educational Research.
Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika, 79*(2), 210–231.

Kyriakides, L., Giorgiou, M. P., Creemers, B. P. M., Panayiotou, A., & Reynolds, D. (2018). The impact of National Policies on student achievement: A European study. *School Effectiveness and School Improvement, 29*(2), 171–203.

Lapointe, A. E., Mead, N., & Phillips, G. (1989). *A world of difference: An international assessment of mathematics and science*. Princeton, NJ: Educational Testing Services.

Lewis, S. (2014). The OECD, PISA and educational governance: A call to critical engagement. *Discourse: Studies in the Cultural Politics of Education, 35*(2), 317–327.

Loveless, T. (2013). *Attention OECD-PISA: Your silence on China is wrong*. Washington, DC: Brown Center on Education Policy at Brookings.

Loveless, T. (2014). *PISA's China problem continues: A response to Schleicher, Zhang & Tucker*. Washington, DC: Brown Center on Education Policy at Brookings.

Messick, S. (1998). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). New York, NY: Macmillan.

Meyer, H. D., & Schiller, K. (2013). Non-educational influences on PISA outcomes. In H. D. Meyer & A. Benavot (Eds.), *PISA, power and policy: The emergence of global educational governance*. Oxford, UK: Symposium Books.

Morgan, C., & Shahjahan, R. A. (2014). The legitimation of OECD's global educational governance: Examining PISA and AHELO test production. *Comparative Education, 50*(2), 193–205.

Mourshed, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. London, UK: McKinsey.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socio-economically disadvantaged areas: A review of research evidence. *School Effectiveness and School Improvement, 15*, 149–175.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – Teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256.

Muijs, D., & Reynolds, D. (2011). *Effective teaching. Evidence and practice*. London, UK: Sage.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2013). The impact of Reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item Reading demands. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade – Implications for early learning* (pp. 67–108). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murphy, D. (2014). Issues with PISA's use of its data in the context of international education policy convergence. *Policy Futures in Education, 12*(7), 893–916.

Murphy, S. (2010). The pull of PISA: Uncertainty, influence, and ignorance. *Inter-American Journal of Education for Democracy, 3*(1), 27–44.

OECD. (2009). *PISA 2009 assessment framework: Key competencies in reading, maths and science*. Paris, France: OECD.

OECD. (2015). *PISA frequently asked questions.* Retrieved from http://www.oecd.org/pisa/aboutpisa/pisafaq.htm

Pereyra, M. A., Kotthoff, H. G., & Cowen, R. (2011). *PISA under examination*. Rotterdam, The Netherlands: Sense Publishers.

Postlethwaite, T. N., & Ross, K. (1992). *Effective schools in reading: Implications for educational planners*. Newark, DE: IEA.

Postlethwaite, T. N., & Wiley, D. E. (1992). *The IEA study of science 11, science achievement in twenty three countries*. Oxford, UK: Pergamon Press.

Purves, A. C. (1992). *The IEA study of written composition 11: Education and performance in fourteen countries*. Oxford, UK: Pergamon Press.

Reynolds, D. (2008). New labour, education and Wales: The devolution decade. *Oxford Review of Education, 34*, 753–765.

Reynolds, D., Bollen, R., Creemers, B. P. M., Hopkins, D., Stoll, L., & Lagerweij, N. (1996). *Making good schools: Linking school effectiveness and school improvement*. London, UK: Routledge.

Reynolds, D., Caldwell, B., Cruz, R. M., Miao, Z., Murillo, J., Mugendawata, H., … Ramon, M. R. R. (2015). Comparative educational research. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The Routledge international handbook of educational effectiveness and improvement*. London, UK: Routledge.

Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., & Schaffer, E. (2002). *World class schools: International perspectives in school effectiveness*. London, UK: Routledge Falmer.

Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., Schaffer, E., & Nesselrodt, P. S. (1994). *Advances in school effectiveness research and practice*. Oxford, UK: Pergamon Press.

Reynolds, D., & Farrell, S. (1996). *Worlds apart? A review of international surveys of educational achievement involving England*. London, UK: HMSO for OFSTED.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state of the art review. *School Effectiveness and School Improvement, 25*(2), 197–230.

Rosier, M. J., & Keeves, J. P. (1991). *The IEA study of science 1, science education and curricula in twenty three countries*. Oxford, UK: Pergamon Press.

Ruddock, G., Clausen-May, T., Purple, C., & Ager, R. (2006). *Validation study of the PISA 2000, PISA 2003 and TIMSS 2003 international studies of pupil attainment*. Nottingham, UK: Department for Education and Science.

Rutkowski, L., & Rutkowski, D. (2016). A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher, 45*(4), 252–257.

Sahlberg, P. (2011). *Finnish lessons*. New York, NY: Teachers College Press.

Sammons, P. (2010). The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioral research* (pp. 697–724). Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781506335193

Sellar, S., & Lingard, B. (2013a). Looking east: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. *Comparative Education, 49*(4), 464–485.

Sellar, S., & Lingard, B. (2013b). The OECD and global governance in education. *Journal of Education Policy, 28*(5), 710–725.

Sjøberg, S. (2012). PISA: Politics, fundamental problems and intriguing results. *Recherches en Education, 14*, 1–21.

Slavin, R. E. (1996). *Education for all*. Lisse, The Netherlands: Swets & Zeitlinger.

Stigler, J. W., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS videotape classroom study: Methods and findings from an exploratory research project on eighth grade mathematics instruction in Germany, Japan and the United States*. Washington, DC: National Center for Education Statistics.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a 10-year study of school effects*. New York, NY: Teachers College Press.

Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics 1: Analysis of mathematics curricula*. Oxford, UK: Pergamon Press.

Waldow, F., Takayama, K., & Sung, Y. K. (2014). Rethinking the pattern of external policy referencing: Media discourses over the 'Asian Tigers' PISA success in Australia, Germany and South Korea. *Comparative Education, 50*(3), 302–321.

Whelan, F. (2009). *Lessons learned: How good policies produce better schools*. London, UK: Author.

Wuttke, J. (2007). Uncertainty and Bias in PISA. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *PISA according to PISA: Does PISA keep what it promises?* (pp. 241–263). Berlin, Germany: Lit Verlag.

Yildirim, H. H. (2006). *The differential item functioning (DIF) analysis of mathematical items in the international assessment programs.* Unpublished doctoral thesis. Middle East Technical University, Ankara, Turkey.

Zhao, Y. (2014). *Who's afraid of the big bad dragon: China the best and worst education system in the world*. San Francisco, CA: Jossey Bass.

Chapter 7
Policies and Practices of Assessment: A Showcase for the Use (and Misuse) of International Large Scale Assessments in Educational Effectiveness Research

Eckhard Klieme

7.1 International Large Scale Assessment (ILSA) and Educational Effectiveness Research (EER)

International Large Scale Assessments (ILSAs) are international assessments of educational topics that target large and representative samples of students and/or teachers, as well as other stakeholders in education such as school principals or parents. They started around 1960 with the first studies of the International Association for The Evaluation of Student Achievement (IEA). Today, the most cited studies include the Progress in International Reading Literacy Study (PIRLS, run by IEA), the Programme for International Student Assessment (PISA, run by OECD) and Trends in International Mathematics and Science (TIMSS, run by IEA).

The overarching and initial goal of such ILSAs is to provide indicators on the effectiveness, equity, and efficiency of educational systems (Bottani & Tuijnman, 1994), to set benchmarks for international comparison, to monitor trends over time and thus inform educational policy on an international, national, regional and even local (school) level, e.g. with regard to innovations in educational governance and curriculum (Klieme & Kuger, 2015). Consequently, ILSAs have attracted much media attention in many countries and have exerted sometimes far-reaching influence on education policy. Bogdandy and Goldmann (2009), scholars in international public law, claim that ILSAs even allow international organizations like OECD to establish a new legal mechanism they call "governance by information". In addition to educational politics, administration and the public, researchers increasingly draw on the results of these assessments to study, on the one hand, the universality and generalizability of certain findings in educational effectiveness and, on the other

E. Klieme (✉)
DIPF| Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany
e-mail: klieme@dipf.de

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_7

hand, the respective national, regional, cultural, and other group-specific features that may moderate universal mechanisms.

Scholars repeatedly claimed that studies must take into account theoretical considerations, modeling approaches and research results from Educational Effectiveness Research (EER) to develop a meaningful system of reporting indicators (e.g., Bryk & Hermanson, 1994). However, when the U.S. National Research Council attempted to summarize what had been learned from more than three decades of ILSA up to the year 2000, in the concluding chapter Rowan (2002, p. 338) argued for a deeper analysis of "relationships among school characteristics and student achievement" (p. 338) and "effects of educational practices" (p. 339). Thus, ILSAs had failed to align with EER in the twentieth century. One reason was that ILSA designs had almost exclusively focused on student tests, and most of the administration time had been spent with students working on test items, while only a small number of student, classroom, school or even system-level characteristics were measured in so-called "background questionnaires". Fortunately, this situation has changed over the last decade, especially as PISA introduced "analytical frameworks" (most recent: Klieme & Kuger, 2015) which systematically linked questionnaire development to policy issues, research questions, and scientific constructs informed by the growing knowledge base of EER. In a recent review of ILSA questionnaire design, Jude and Kuger (2018, p. 5) mention three advantages of basing ILSA frameworks on EER:

> (1) EER acknowledges the complexity of educational systems, (2) EER frameworks ultimately aim at explaining student outcomes, and (3) overarching EER theories offer a number of different anchors to relate to other, interdisciplinary theories or frameworks.

Vice versa, ILSAs can as well contribute to the further development of EER (Klieme, 2012) by providing data, triggering new research studies, and providing instruments which work across multiple cultures. First, ILSA data are based on large representative samples assessed in multiple countries, usually with high quality; yet they are easily accessible. Researchers may use these data for fruitful secondary analysis, both within and across countries. Opportunities and limitations of using such data, including the implementation of enhanced designs, will be discussed in the present chapter; for a recent overview see Singer, Braun, and Chudowsky (2018). Second, the rich description of education in different cultures, school systems, and *s*chool contexts that is provided by ILSA studies – although limited by its descriptive nature – can inspire EER to discover new fields of research. National and international patterns or regional peculiarities are easily accessible through screening publicly available ILSA data, and can trigger new research questions that lead to the careful development of smaller, targeted EER studies. For example, PISA findings on disadvantages for migrant students and students from low SES families motivate EER to have a closer look at differential school effectiveness. Third, ILSAs carefully develop research instruments and methodology that may be used in further studies both within and across countries. The high quality standards typically involved in the preparation and implementation of ILSAs, including sophisticated procedures for translation, adaptation, administration, data cleaning, and scaling,

provide EER with high quality, culturally adapted, policy-relevant material in a large number of languages to support innovative EER studies.[1] Therefore, ILSAs offer an unmatched source of ready-to-use instruments for EER that has been developed and refined under strict quality guidelines and discussed by education, policy, questionnaire, and survey method experts.

When using ILSA data in the context of EER, researchers need to cope with typical limitations of ILSA designs, above all with the cross-sectional nature of data and the lack of cross-cultural comparability. On one hand, state-of-the-art Item Response Theory (IRT) methods are applied, missing data are treated in very sophisticated ways, and complex, multi-level models are used for analysis. On the other hand, the cross-sectional nature of the data severely limits any interpretation in terms of the direction of relationships in educational effectiveness or ability to infer causality. Unobserved confounding variables such as prior achievement might explain existing patterns and relationships, and the direction of causality might not be clear. At least, this holds for interpretations of findings made at the student, class, or school level. To overcome these limitations, researchers, especially in Germany, have added follow-up measures to study conditions of student learning over one school year (see Baumert et al., 2010, for a study of teacher effectiveness based on a longitudinal extension of PISA 2003; Kuger, Klieme, Lüdtke, Schiepe-Tiska, & Reiss, 2017, for a study of teaching quality enhancing PISA 2012) and effects of school policies such as internal evaluation and all-day-programmes on long-term change in school climate and school outcomes (Bischof, Hochweber, Hartig, & Klieme, 2013, based on a sample of schools participating both in PISA 2000 and in PISA 2009). Other scholars have been focusing on the country or "system" level, using repeated measures from trend studies such as PISA and TIMSS, analyzed through difference-in-difference estimation (Gustafsson, 2007), or fixed effects techniques (Bergbauer, Hanushek, & Wößmann, 2018).

Whenever analyses are run across countries, they require some level of measurement invariance (van de Vijver & He, 2016). In particular, "metric invariance" typically is required in order to warrant the claim that the construct of interest has the same meaning across countries. Running linear models, e.g. regression analyses, in parallel for a number of countries and comparing effect sizes are legitimate only if metric invariance can be established. Furthermore, a higher level of invariance, namely "scalar invariance", is required to allow mean differences on the construct to be compared across countries. Researchers who have checked invariance most often found that metric invariance can be established for many questionnaire scales, while scalar invariance seems to be a rare exception (He, Buchholz, & Klieme, 2017; He & Kubacka, 2015).

[1]Usually, questionnaires are published in the source language (mostly English) only. For PISA 2015, translated versions from 75 countries, item- and scale-level statistics are available at https://daqs.fachportal-paedagogik.de/search/show/survey/177?language=en. This online depository includes Field Trial material not yet used in the Main Study. For an introduction and conceptual overview, see Kuger, Klieme, Jude and Kaplan (2016).

Because of these limitations, policy making has often been misinformed and misled by shortcut interpretations and too-far-reaching conclusions (Baker, 2009). For example, based on PISA 2006 the OECD reported that "Students in schools posting their results publicly performed 14.7 score points better than students in schools that did not, and this association remained positive even after the demographic and socio-economic background of students and schools was accounted for" (OECD, 2007, p. 243). They concluded "that the impetus provided by external monitoring of standards, rather than relying principally on schools and individual teachers to uphold them, can make a real difference to results" (p. 276), although the reverse causality interpretation seems to be more realistic: schools might tend to publish their results if they have been successful in an assessment. More recently, OECD claimed that extracurricular activities and school climate would increase the proportion of "resilient" students in Germany (OECD & Vodafone Stiftung, 2018), while this finding was due to a neglect of the tracked structure of the German system (Klieme, 2018). There have also been examples of policy makers cherry picking results and making inappropriate claims about dramatic falls in student attainment to justify new and controversial 'education reforms'. For example in England a so called plummeting of student performance in PISA tests in country league tables was used to justify the introduction of the free school and academisation programme in 2010 although the actual performance of England was not statistically different in terms of the country reference group at the time (Jerrim, 2011). In Germany, a massive investment into after school programmes was publicly claimed to be a consequence from PISA 2012, which in fact it was not (Klieme, Jude, Baumert, & Prenzel, 2010). Within the research community, overstatements have been made as well. This includes the present author, who interpreted cross-sectional relationships between perceived teaching quality and student outcomes as indicative of teaching effects in PISA 2000 (Klieme & Rakoczy, 2003). More recently Schmidt, Burroughs, Zoido, and Houang (2015) drew far-reaching conclusions on the effects of opportunity to learn on mathematics achievement using cross-sectional PISA 2012 data and a mis-specified indicator for "applied mathematics" (as can be seen by comparing with OECD, 2014, pp. 56 and 324).

Rather than supporting claims on educational effectiveness proper, i.e. estimating the *effects* of specific policies and practices on student outcomes, ILSA data may be used to inform about the *distribution* of educational opportunities among students, families, schools, and regions. Policies and practices would be treated as dependent variables, while student achievement as well as student and family background would be treated as independent variables. E.g., do migrant students and students from socially disadvantaged families have an equal share of well-trained teachers, engaged school principals, of well-ordered, supportive and challenging classroom environments and out-of-class learning opportunities? Who receives differentiated instruction, supportive feedback and direct guidance from his or her teachers? Which schools report policies for assessment and evaluation, and which don't? Similarly, differential uptake of activities, use of opportunities and engagement in learning may be studied to understand inequity. E.g., does student truancy and attention in classroom differ between subpopulations? While studies on effectiveness typically

require experimental, or at least quasi-experimental designs, cross-sectional ILSAs are well prepared to answer questions about the provision of, differential access to, and differential use of learning opportunities. These kinds of questions may not be considered crucial in EER, but they are highly relevant for policy making and for understanding (in)equity in education.

In an attempt to illustrate the methodological issues raised above, to further explore the opportunities and limitations of ILSA data and to discuss their relevance for EER, the present chapter will use a specific showcase: policies and practices of educational assessment. This topic will be introduced in Sect. 7.2, while Sect. 7.3 will present and critically discuss related findings from PISA 2015. Thus, the chapter integrates three layers of academic discourse: (a) The meta-theoretical issue of how ILSAs relate to EER. (b) The substantive EER research question on how assessment is implemented in education, how it is shaped by systems and schools, and how it relates to student learning. (c) Specific methodological issues in analyzing ILSA data, which are discussed in several excursus spread across the chapter.[2]

7.2 Policies and Practices of Assessment as a Topic in Educational Effectiveness Research[3]

For at least three decades, assessment and evaluation have been major strands of educational policy and practice internationally. In recent years, there has been growing interest in the use of assessment and evaluation results through feedback to students, parents, teachers, and schools as one of the most powerful tools for quality management and improvement. Reporting and sharing data from assessments and evaluations with different stakeholders provides multiple opportunities for monitoring both individual learning and institutional development, for certification and accountability (Elacqua, 2016). The volume *Schools and Quality*, published by OECD in 1989, marked the initiation of a global trend that is still ongoing: "educational assessment, evaluation, and accountability are still evident in educational practice and policy making in virtually every country" (Huber & Skedsmo, 2016, p. 1). This trend is part of an overarching change in concepts and measures of educational governance (Altrichter & Maag Merki, 2016). New forms of educational governance, such as school performance feedback systems (Visscher & Coe, 2003), systemic approaches to educational evaluation and monitoring (Scheerens, Glas, & Thomas, 2003) and concepts of data-driven school improvement (Coburn & Turner, 2011; Spillane, 2012) have become popular among policy makers. Over the years,

[2]The author wants to thank Anindito Aditomo, Sonja Bayer, Janine Buchholz, Jessica Fischer, Jia He, Nina Jude and Susanne Kuger for collaboration on this topic at the DIPF Department for Research on Educational Quality and Evaluation.

[3]This section is in part based on Bayer, S., Klieme, E. & Jude, N. (2016). Assessment and evaluation in educational contexts. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective* (pp. 469–488). Cham: Springer.

the assessment/evaluation paradigm has shifted from a focus on measurement towards a focus on efforts to improve learning (Wyatt-Smith, 2014). Formative Assessment and Feedback to students have been shown to be among the most powerful tools teachers can use to boost their students' understanding and achievement (e.g., Bennett, 2011; Hattie, 2009; Kingston & Nash, 2011).

In the following, we broadly discriminate two areas of assessment and evaluation that seem to become increasingly popular around the globe: assessing and evaluating schools on the one hand, assessing and measuring student learning in the classroom on the other hand. In both cases, data may be used either for formative purposes, informing school improvement activities and classroom teaching, respectively, or for summative and accountability purposes, such as ranking schools with regard to national standards and issuing certificates for individual students. It should be noted, however, that test measures may be used across areas. For instance, student outcomes, aggregated to the appropriate level, may be used to judge educational systems, individual schools, and teachers alike.

7.2.1 School Evaluation

The evaluation of schools is an important instrument of educational governance used in decisions and judgments about processes, programmes, reforms, and educational resources (Faubert, 2009). Moreover, the evaluation of schools can help school leaders to make better decisions about processes, build knowledge and skills, or to facilitate continuous improvement and organizational learning. The improvement of schools participating in evaluation programmes can be explained by feedback theory, (Visscher & Coe, 2003), or as an effect of stakeholders within school being held accountable for evaluation results (Donaldson, 2004):

- Feedback is a core element of data-driven school development (Scheerens et al., 2003), at best pushed by a combination of internal and external evaluation. Feedback may also be provided by national test programs allowing schools to compare their own performance with national standards. Scheerens et al. assume evaluation to be the fundamental process through which a school becomes a learning organization, and they believe evaluation- and feedback-based school improvement to be more effective than any forward-planning strategy.
- From an accountability perspective, rewards and penalties are assumed to change the behaviours of stakeholders in ways that improve student achievement (Wößmann, Lüdemann, Schütz, & West, 2009). Strong accountability practices include the public availability of assessment and evaluation results (Scheerens et al., 2003). Such information could be used by parents for school choice, or by local communities for resource allocation. Bergbauer et al. (2018) provide an econometric model based on principal-agent-theory, summarized as follows (p. 6): "By creating outcome information, student assessments provide a mechanism for developing better incentives to elicit increased effort by teachers and students, thereby ultimately raising student achievement levels to better approximate the desires of the parents".

School evaluation and improvement can indeed affect students' outcomes. For instance, Scheerens (2002) and also Creemers and Kyriakides (2008) report evidence that systematic school evaluation can positively impact students' outcomes. On the basis of a school panel added to the PISA 2000 and 2009 samples in Germany, Bischof et al. (2013) report that schools that had done some internal evaluation improved in terms of both student achievement and school climate.

Different evaluation practices generally coexist and benefit from each other (Ryan, Chandler, & Samuels, 2007). External evaluation can expand the scope of internal evaluation, and also validate results and implement standards or goals. Internal evaluation can improve the interpretation of external evaluation results (Nevo, 2002). In a review of 41 empirical studies on evaluation use, Johnson et al. (2009) found the involvement of stakeholders to be most important condition of effective school evaluations. Engagement, interaction, and communication between evaluation clients and evaluators are critical to the meaningful use of evaluations for improvement purposes.

Common steps of effective evaluation can be identified (e.g., Sanders & Davidson, 2003), yet school evaluation approaches are multifold and vary across educational systems (OECD, 2013). Therefore, it is difficult to report on and compare the effects of evaluation across different evaluation systems and education systems.

7.2.2 Assessment Embedded in Classroom Teaching and Learning

In its summarizing function, assessment takes place in order to grade, certify or record progress. A summative assessment therefore indicates and monitors standards, but it may also raise standards by stimulating students, as well as teachers and schools, to invest more effort in their work (Harlen & Deakin Crick, 2002). On the other hand, summative assessment might lead to lower self-esteem and diminished effort in students at risk, which could increase the gap between lower- and higher-achieving students (Black & Wiliam, 2004). Another side effect can emerge if teachers neglect skills development and knowledge in opting rather to train their students in test-taking strategies (Harlen & Deakin Crick, 2002).

Apart from summative assessments, formative assessment plays a key role in classroom learning (e.g., Shepard, 2006; Black & Wiliam, 2004; McMillan, 2007; OECD 2005). Several meta-analyses indicate that formative assessment is a significant source of improvement in student learning processes. In particular, low achievers benefit from formative assessment, which can lead to sizable gains in student achievement (Abrams, 2007). However, there is large variation in the implementation and in the impacts of formative assessment (e.g., Bennett, 2011; Black & Wiliam, 1998; Hattie & Timperley, 2007; Kingston & Nash, 2011; Shute, 2008). Effects of formative assessment have been shown to be moderated by generic

teaching quality (Decristan et al., 2015) and by students' perception of usefulness (Rakoczy, Klieme, Leiss, & Blum, 2017).

Feedback plays a key role in formative assessment. Hattie and Timperley (2007) have identified four types of feedback provided to students that have differential effects on learning: Feedback may refer to (1) the student, evaluating him or her on a personal level, (2) task performance (3) task processing and (4) self-regulation (see also Kingston & Nash, 2011). Most commonly, feedback is given about task performance (2; also called corrective feedback). This feedback can be useful if the recipient uses it to reconsider and, if necessary, adapt their strategies or to enhance self-regulation. Otherwise, feedback can explicitly refer to processes to solve a specific kind of task (3) or to non-task-specific strategies (4): for example, how to learn, or how to structure a learning process. The latter two types of feedback have been shown to be the most effective, but learners need to know how to incorporate the feedback into their thinking. Feedback on a personal level (1; e.g., "you are a nice student") is less effective. In general, feedback to students needs to be simply coded, and suggestions need to be realistic (Sadler, 1989). Feedback that meets these conditions will allow students to understand the gap between the intended learning goal and what they have achieved so far, and guide them to take appropriate steps.

7.2.3 *Using ILSAs to Inform Research on Assessment and Evaluation*

Assessments (ILSAs) like TIMSS, PIRLS and PISA are major driving factors for system-level monitoring. They provide complex techniques to be used *for* assessment, evaluation, and accountability at all levels of the educational system. At the same time, these international surveys can be used as sources of information *about* assessment, evaluation and accountability practices in cross-national comparison. This is illustrated by the findings which will be presented in the remaining part of this chapter.

ILSA data may inform critical debates on assessment, evaluation, and accountability systems in the public sphere, in policy and pedagogy, and overcome the purely ideological debates that oftentimes dominate this discourse. Another advantage is the broad coverage of geographical areas and cultural contexts, which helps widening the scope of insights on this topic beyond the "Western", mostly English speaking world which dominated both policy and research on assessment and evaluation for a long time.

A recent, ground-breaking study using ILSA data from 59 countries is the paper entitled "Testing" by Bergbauer et al. (2018). Combining indicators from PISA 2000 to 2015 (mainly the ones discussed later in the present chapter) as well as from international comparative reviews of assessment policy, they claim to measure how strongly each of four different types of assessments has been implemented in a

country: (1) Standardized External Comparisons such as national tests or central exams. (2) Standardized Monitoring, i.e. using standardized tests for internal purposes without necessarily comparing to external standards, (3) Internal Testing, informing local stakeholders based on any measure of student achievement, and (4) Internal Teacher Monitoring, using any kind of evaluation mechanism to judge teachers. The authors find that only the first type of assessment is associated with improvements in student achievement. This finding is supported by a sophisticated set of models and robustness checks. However, the following interpretation may raise some skepticism: "Internal testing that simply informs or monitors progress without external comparability and internal teacher monitoring including inspectorates have little discernible effect on overall performance" (Bergbauer et al., 2018, p. 2). The problem is that assessment policies and practices are measured on the country level only. Thus, for each of the 59 countries, there are 4 indicators measured up to 6 times (in 2000, 2003, 2006, 2009, 2012, 2015). These indicators describe how strongly different kinds of assessments are implemented on average across a country, but there is no attempt to measure policies and practices at the school- or classroom level. This kind of analysis is appropriate when comparing national evaluation and accountability policies such as centralized exams between countries and studying their impact over time (as Sects 7.3.4, 7.3.5 and 7.3.7 below will do), but it is inappropriate for studying the impact of internal evaluation and classroom assessment on student learning in local contexts. To answer the latter research question, longitudinal and (quasi-)experimental enhancements of PISA would be needed. Once again, the limitations of ILSA designs need to be taken seriously (as discussed in Sect. 7.1).

7.3 A Comparative Analysis of Assessment Policies and Practices, Implemented in PISA 2015

Starting with the first wave in 2000, school questionnaires in all cycles of OECD's Programme for International Student Assessment have addressed policies of evaluation and assessment, and how results were used within countries. Thus, existing PISA trend data helps us understand how the use of student assessments has widened over the past 15 years in almost all OECD countries and how this has impacted system-level change in student achievement (see Sect. 7.3.7 below). In PISA 2015, the author, in close collaboration with staff at the German Institute for International Educational Research (DIPF) and the International Questionnaire Expert Group, developed a broader set of questions covering details of school evaluation and classroom assessment (Bayer, Klieme & Jude, 2016). These new measures will be presented in Sect. 7.3.1. Using PISA 2015 Main Study data from 55 participating countries (Sect. 7.3.2), empirical analyses will present findings on formative assessment practices in classrooms (Sect. 7.3.3), assessment-related school policies (Sect. 7.3.4), national contexts for assessment and accountability (Sect. 7.3.5) and the

relationships between them (Sect. 7.3.6). In the course of the section, methodological issues related to comparability across countries and interpreting cross-sectional findings will be discussed.

7.3.1 Developing Measures for PISA 2015

The PISA 2015 Context Questionnaires (Kuger et al., 2017; OECD, 2013) allow for studying how often students are assessed through highly standardized tests, teacher-made tests or teachers' judgmental rating (Question SC034 on "General assessment practice", see Table 7.1) and whether certain measures for school improvement, including internal and external evaluations of schools are common practice (SC037). Moreover, the impetus for action is also relevant in order to analyze system policies. Thus, the PISA 2015 questions referring to school improvement policies (SC037) and standardized testing (SC034) distinguish action that is mandatory, i.e. required by educational policies, from action that is based on the school's initiative.

To support the description and analysis of data use by schools, a set of items from previous PISA cycles (2000–2012) was taken up addressing various kinds of usage for student test scores, such as informing parents, deciding upon student promotion, or comparing the school with other schools (Purpose of assessment results; SC035). Some items on formative use (e.g., guiding student learning and adapting teaching) were newly added, and the response format was changed with the intention to discriminate the use of standardized tests from the use of teacher-developed tests. In addition, three items asking whether schools publish test results, track them over time and/or provide scores to parents have also been taken up from previous cycles of PISA (Use of achievement data for accountability; SC036).

In order to understand the link between assessment, evaluation, and school development, fine-grained information on processes of external evaluation (e.g., Have the data been used to inform specific action for improvement of teaching? Were such measures put into practice promptly?; SC041) and consequences of internal evaluation (Which areas have been affected by change in school policies:

Table 7.1 Assessment-related questions in the PISA 2015 school (SC) and student (ST) questionnaires

Policies for assessment, evaluation and accountability (individual indicators and indices)	
General assessment practice	SC034
Purpose of assessment results	SC035
Use of achievement data for accountability	SC036
Measures for school improvement, including internal and external evaluation	SC037
Consequences of internal evaluation	SC040
Processes of external evaluation	SC041
Formative assessment and feedback (latent scale for classroom practice)	
Perceived feedback	ST104

curriculum, professional development, parental engagement, etc.?, SC040) has been added.

It should be noted that two questions (namely, SC034 and SC035) are referring to student assessment at the "national modal grade for 15 year old students", i.e. the grade which enrolls most of the PISA target population nationwide, while all other questions refer to school assessment policies in general.

All topics mentioned so far have been addressed within the PISA 2015 School Questionnaire, which was meant to be answered by the school's principal. Each item in this part of the School Questionnaire asked whether (all questions), why (SC037 and SC040 only) and how frequently (SC034 only) specific activities had been implemented by the school. Answers were treated as individual indicators, and sometimes indices were computed summing up across several activities. For example, three indices based on SC035 are summing up across different kinds of testing purposes, each of which is presented with a forced-choice ("yes"–"no") response format:

- "Formative use": This index counts how many of the following had been identified as purposes of using standardized tests in school: guiding student learning, adapting teaching to students 'needs, identifying aspects of instruction or curriculum that should be improved, informing parents.
- "Summative use on the student level" counts how many of the following purposes of standardized testing have been identified by the school principal: decision about retention, certification, or grouping students.
- "Use for school evaluation" counts positive answers to the following items: compare test results with other schools, compare to national or district performance, monitor school progress.

We do not expect these three indices to represent any "latent" construct. Rather than measuring dispositional concepts of school evaluation, the indices pragmatically summarize school policies which analytically fall within the three categories. In methodological terms, these indices are treated as "formative" measures, i.e. indices are defined by the items covered, rather than items "reflecting" a latent construct (for the distinction, see Ellwart & Konradt, 2011).

Arguably the most prominent form of assessment studied so far in educational research is formative assessment (see Sect. 6.2.2 above). Since feedback is essential in formative assessment, we assessed this concept in the PISA 2015 student questionnaire (Perceived Feedback; ST 104; see Table 7.2). This scale was developed to assess the frequency of (formative) feedback activities as perceived by each individual student, asking how often the teacher would inform the student about his or her performance (Item 1), identify strengths (Item2), tell the students where (Item 3) and how (Item 4) he or she can improve, and provide advice on how to reach the learning goals (Item 5). Students were asked to respond in reference to one science course they had chosen before. Contrary to individual indicators included in the School Questionnaire, these five items are supposed to reflect an underlying latent dimension of classroom practice, i.e. a view on teaching which is understood in a similar way by students within a given class, school or even system. OECD (2017a,

Table 7.2 Items on "perceived feedback" in the PISA 2015 student questionnaire. (OECD 2017a, p. 315)

Item number	Item	Corrected Item-Total-Correlation (Median across 55 countries)
ST104Q01NA	The teacher tells me how I am performing in this course.	.70
ST104Q02NA	The teacher gives me feedback on my strengths in this <school science> subject.	.78
ST104Q03NA	The teacher tells me in which areas I can still improve.	.81
ST104Q04NA	The teacher tells me how I can improve my performance.	.81
ST104Q05NA	The teacher advises me on how to reach my learning goals.	.79

p. 315) provides scale scores estimated within an Item Response Theory (IRT)-based approach; the respective variable is named "PERFEED" in the PISA 2015 data file.

7.3.2 Data and Methods

As described in the Technical Report (OECD, 2017a), PISA 2015 sampled 546,299 students from 18,817 schools representing 26.9 Mio 15 year old students in 74 countries (35 OECD members plus 39 "partner" countries). The study implemented a two-stage stratified strategy, sampling about 150 schools per country and about 30 students per school. Within schools, students of the target age (15 years) were selected across grade levels and classrooms. Thus, contrary to PIRLS and TIMSS, PISA does not allow for an analysis of classroom-level variation in teaching and learning.

Students worked on cognitive tests for about 2 h, followed by a Student Questionnaire which took about 35 min. Both tests and questionnaires were administered on computer in the vast majority of countries. In addition to students, school principals and – in 19 countries – also teachers were asked to fill in web-based questionnaires. The technology-based administration allows for routing procedures to guide the individual respondent through the questionnaire. E.g., school principals are asked about the purpose of using standardized tests (SC035) if and only if they have said they implement such tests in their school (SC034). This procedure should help avoid invalid responses, but as a side effect, a significant part of the data matrix will be "missing by design". E.g., across 55 countries, data on testing purposes are available from two thirds of the schools only.

The full international data file may be downloaded from the OECD website. The SPSS student-level data file contains about 1000 variables, including 130 measures of student achievement (13 separate domains or sub-domains, each represented by 10 "Plausible Values" to cope with the matrix-design used for testing) and about

160 technical variables (e.g., information on sampling and weights). More than 700 variables are based on the student questionnaire, including several optional add-ons. The School file consists of about 300 variables based on the School Questionnaire.

OECD reports routinely cover all participating countries and "systems". Nevertheless, experts are well aware that data quality varies between countries. For example, the feasibility and appropriateness of the study design for developing countries may be questioned. As a consequence, some researchers only use data from OECD countries. However, for cross-cultural studies, this results in a severe loss of cultural and systemic variation. As a compromise, the analyses reported in this chapter have been run for 55 selected countries, which may (a priori) be grouped into ten categories based on geographical and/or linguistic proximity:

- English speaking: AUS, CAN, IRL, NZL, UK, USA
- German/Dutch speaking: AUT, BEL, CHE, DEU, LUX, NLD
- Roman Europe: ESP, FRA, ITA, PRT
- Nordic States: DNK, FIN, ISL, NOR, SWE
- Baltic States: EST, LTU, LVA
- Central Europe: BGR, CZE, HUN, HRV, POL, ROU, SVK, SVN
- Eastern Europe: GEO, KAZ, MDA, RUS
- Eastern Mediterranean: CYP, GRC, ISR, MLT, TUR
- East Asia: HKG, JPN, KOR, MAC, SGP, TAP and China (representing several industrialized regions from the Eastern part of China, including Beijing and Shanghai),
- South America: ARG, BRA, CHL, COL, MEX, PER, URY

Thus, we include 14,111 schools, i.e. 75% of all schools participating in PISA 2015. Cases were weighted using the so-called "Senate weight" which standardizes all national or system samples to an equal size. Descriptive analyses, including exploration of relationships with student outcomes and other kinds of linear relationships, were run in parallel for all 55 countries using SPSS 22. As we do not use any significance testing, findings in Sects. 7.3.3 through 7.3.5 are not distorted by the clustered sampling. In Sect. 7.3.3.1, we report results from Multi-group Confirmatory Factors Analyses including multiple countries, executed in MPLUS. In Sect. 7.3.6.2 we do report significance testing, as we study policies, practices, and mean achievement on the school level only.

7.3.3 Formative Assessment and Feedback: Studying Teaching Practice from a Comparative Point of View

For our measure of "perceived feedback", students responded on a four-point Likert scale with the categories "never or almost never" (1), "some lessons" (2), "many lessons" (3), and "every lesson or almost every lesson" (4). Table 7.2 shows the item

wording and provides information on item-total correlation. Across all items and all countries, the minimum value for any item-total-correlation was .36. In all 55 countries, these items form a highly coherent scale, with Cronbach's alpha mostly above .80 (Median: .91); the single outlier is Romania with alpha = .74.

The five items cover different, yet related facets of a classroom practice which is co-constructed by students and teachers, based on shared norms and expectations and coherent chains of teacher and student activities. As any kind of social practice (Reckwitz, 2002), and classroom practice specifically, we assume the practice of formative assessment and feedback to be socially constructed within a culturally shaped social, physical, and intellectual space. Integrating this theoretical view (which is rooted in sociological theory and usually studied by qualitative methods) into the quantitative measurement approach, "Perceived Feedback" may be modeled as a latent variable, assuming the latent structure (i.e. dimensionality, factor loadings/discrimination and intercepts/difficulty) to be at least partially culture-specific, and expecting some agreement among students within the same institutional context (classroom, and to some extent school).

The assumption of perceptions being shared within the same school is in fact supported by the decomposition of variance: Within countries, between 6 and 19% of the variance in Perceived Feedback is between-school variance (Median across all 55 countries = 9.8%). This supports the claim that Formative Assessment and Feedback is a social practice shaped by the learning environment, and perceived in a somewhat similar way by students sharing the same environment. We would expect even higher levels of agreement on the classroom level, but unfortunately the PISA data set does not allow to identify that level. The structural assumptions are tested in the following excursus, summarized in Table 7.3.

7.3.3.1 Excursus on Cross-Cultural Measurement

Applying common conventions for acceptable model fit in structural equation modeling (CFI > .90; RMSEA <.08) to the first row in Table 7.3, we conclude that configural invariance holds across countries. I.e., within each of the 55 countries all five items can be assumed to represent a unidimensional latent construct. Based on criteria suggested by Rutkowski and Svetina (2014) for large international data sets, the loss in model fit when assuming metric invariance on top is negligible (ΔCFI < .02, ΔRMSEA < .03). Thus, factor loadings can be assumed to be equal

Table 7.3 Multiple group confirmatory factor analyses for perceived feedback

	Model assuming configural invariance		Model assuming metric invariance		Model assuming scalar invariance	
Countries	RMSEA	CFI	RMSEA	CFI	RMSEA	CFI
55 (10 regions)	.061	.981	.059	.969	.075	.928
6 English speaking	.038	.993	.038	.988	.045	.976
UK, IRL, AUS, NZL	.038	.993	.034	.991	.034	.988

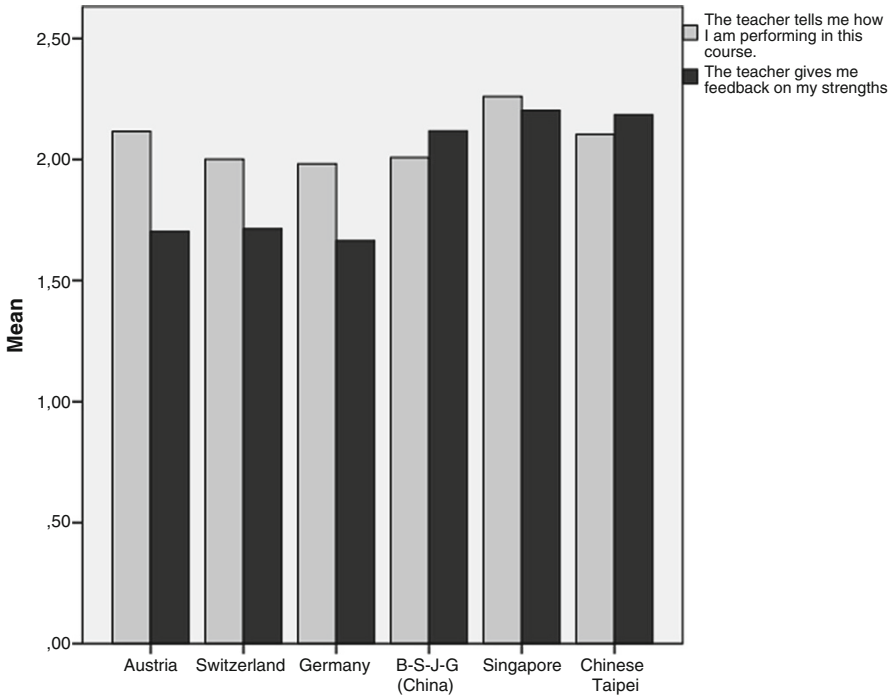

Fig. 7.1 Mean response regarding two feedback items for German-speaking and Chinese-speaking countries

across all 55 countries, meaning that items have similar discriminative power. However, when comparing the model of metric invariance with a third model assuming scalar invariance, the loss in model fit is substantial (neither ΔCFI nor ΔRMSEA are <.01, as requested by Rutkowski & Svetina, 2014). This means that item difficulties vary across countries, an important finding from a substantive point of view.

Figure 7.1 provides an empirical illustration of scalar non-invariance, based on data for items 1 and 2 from three German speaking and three Chinese-speaking countries. First, it should be noted that students report these activities to happen rather seldom; the average score is close to 2 ("some lessons") in all countries and for both items. In the three German-speaking countries participating in PISA, giving feedback on students' strengths is slightly more "difficult" (i.e. mean perceived frequency is lower) than providing feedback on student performance with respect to the course. In the three Chinese-speaking countries shown in this graph, however, the relative "difficulty" of these two items is nearly reversed. What does that mean in terms of "feedback cultures"? Within a Chinese context, teachers relatively often provide feedback based on an individual frame of reference (i.e. interpreting student achievement in relation to individual strengths), while within German culture, teachers seem to prefer a social frame of reference (i.e. interpreting student

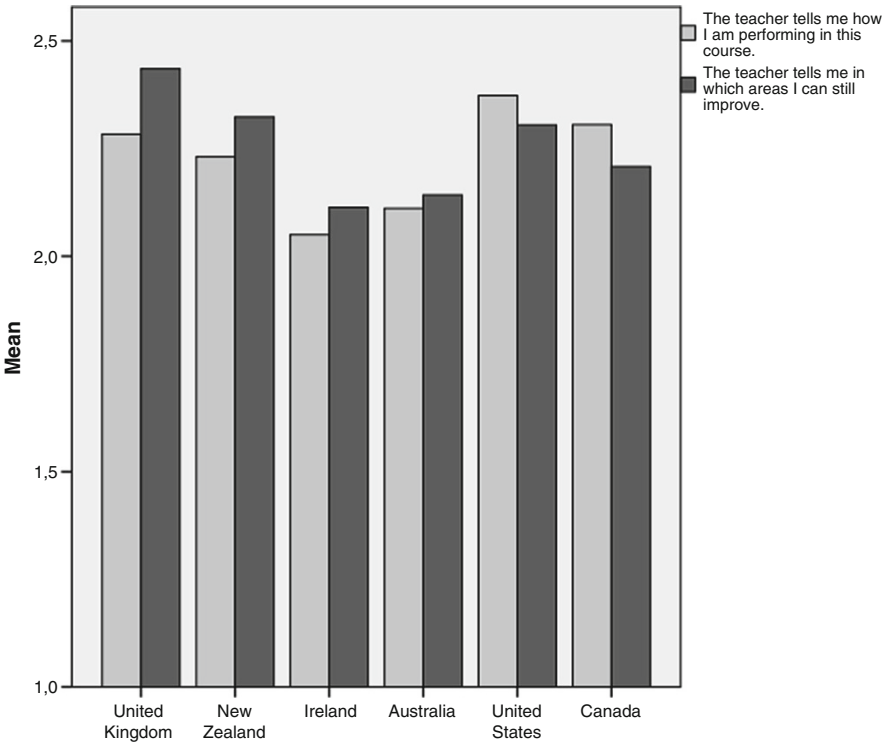

Fig. 7.2 Mean response regarding two feedback items for English-speaking countries

achievement in relation to the class). In statistical terms, the difference in relative "difficulty" implies that these items cannot be assumed to have equal intercepts on a common scale of "Perceived Feedback". Thus, scalar invariance has to be rejected.

The example shown in Fig. 7.1 suggests that linguistic and/or regional proximity might allow for higher levels of comparability. (See Fischer, Klieme & Praetorius, submitted, for similar findings regarding student-reported teaching quality.) This hypothesis is tested in the second row of Table 7.3, checking invariance across all six English-speaking countries included in our analysis. Again, metric invariance holds, but scalar invariance doesn't. Figure 7.2 illustrates, why. Once again, countries differ with respect to the prevalence of applying social comparison within the class vs. an individual frame of reference. Teachers in North America (Canada as well as the US) relatively often provide performance feedback in comparison to the class or course, while in other English-speaking countries, especially in the UK; teachers seem to prefer informing students about areas for individual improvement. As a result, scalar invariance does not hold across all six countries. If however, North America is excluded, in the remaining four countries even scalar invariance holds, as shown in row 3 of Table 7.3.

What does this analysis of measurement invariance imply for ILSA reporting? Still, common practice in policy reports published both by IEA and OECD includes ranking countries on scale means for all kinds of constructs, including measures of classroom practice such as "Perceived Feedback". In volume II of its policy report on PISA 2015, OECD (2017b, p. 67) provides a ranking of mean perceived feedback across countries, with the lowest value for Iceland and high values for the Dominican Republic and (among OECD countries) Mexico. All English-speaking countries are positioned above the OECD-mean, while all German-speaking countries are positioned below the OECD-mean. Based on our analyses of (non-) invariance, this kind of "finding" is not defendable. Rather than providing meaningful and useful information, such OECD ranking produces misleading myths on country-differences in teaching practice, in this case: on the use of formative assessment and feedback.

Strictly speaking, the IRT-based scaling for that set of items provided by OECD is inappropriate. Until PISA 2012, the questionnaire scaling procedures assumed discrimination and difficulty to be the same across all countries, and this assumption had not been tested at all – at least in routine analysis. Following a suggestion from Glas and Jehangir (2014), PISA 2015, for the first time ever, introduced a more sophisticated model, the Generalized Partial Credit Model, checked country- and item-specific misfit as documented in the Technical Report (OECD, 2017a, p. 290), and allowed for some variation in country-specific item parameters.[4] However, the analysis was based on a proprietary software owned by Educational Testing Service, with no prior application to cross-national questionnaire data. Thus, conventions for judging item (mis)fit had to be borrowed from cognitive tests or developed from the scratch. This led to all five items from the "Perceived Feedback" (PERFEED) scale being unflagged; therefore, the IRT scaling was done with common parameters across countries. Using the traditional approach of Multi-Group CFA, the present analysis calls for a revision of those conventions. In general, the methodology of establishing and testing measurement invariance for questionnaire scales in ILSAs is still evolving (van de Vijver, 2018). Organizations such as the OECD and IEA need to be much more hesitant when reporting and statistically comparing country means based on such data. Even within a country, questionnaire scales may lack invariance across different types or tracks of schools (Bayer, 2019), so the caveat applies to EER as well.

7.3.3.2 Restricting Comparison of Scale Means to a Smaller Sample of Countries

Applying rigorous standards of cross-cultural research to the construct of "Perceived Feedback", the discussion summarized in Table 7.3 leads to the decision to restrict

[4]This chapter of the Technical Report was authored by Janine Buchholz from DIPF, who kindly shared findings on the PERFEED scale with the present author. For a review of the scaling method, see Buchholz & Hartig, 2017.

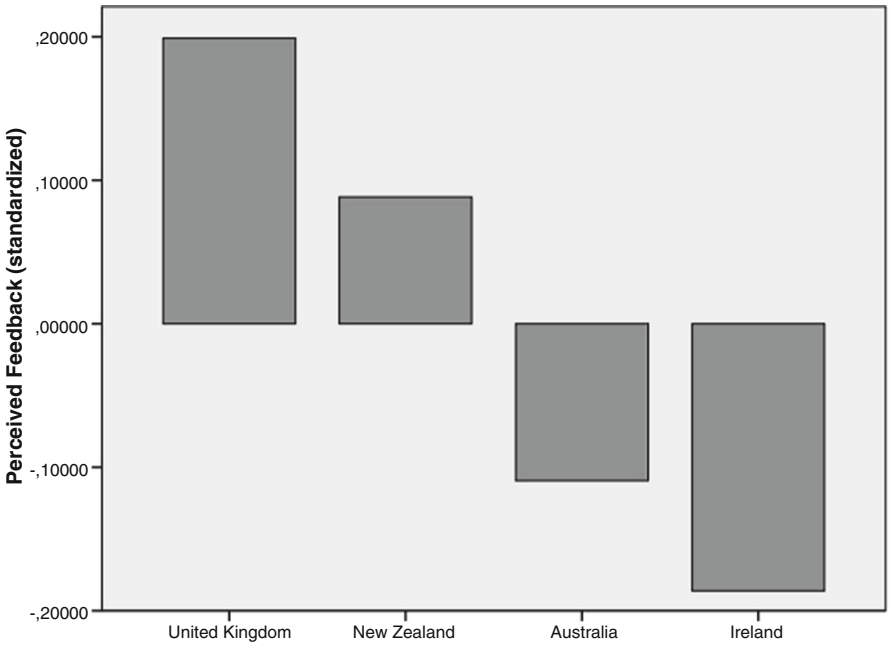

Fig. 7.3 Mean scale score for perceived feedback, standardized across four countries

comparative reports on scale means to four English-speaking countries with proven measurement invariance. Within this group of countries, students from the UK most often report receiving formative feedback (see Fig. 7.3). The difference with New Zealand, Australia, and Ireland roughly amounts to .1, .3, and .4 standard deviations, respectively. This information is trustworthy, and it can be relevant for teacher trainers and policy makers when discussing policies and practices related to formative assessment and feedback. E.g., professionals in the UK may conclude that in their system, promoting *more* use of assessment and feedback in classrooms is less of an issue than addressing the *quality* of assessment. Australian professionals may learn that the strong focus on assessment and feedback established in neighbouring New Zealand has not been implemented in their country (yet?).

7.3.4 The Purpose of Student Testing: Assessment as a School Policy

PISA 2015 provided school principals with an extended list of potential purposes of student testing, asking them to identify those that were relevant for their individual school when using standardized tests. As described in Sect. 7.3.1, three indices were created:

- formative use (e.g., guiding student learning and improving the curriculum),
- summative use on the student level (e.g., certification),
- use for school evaluation (e.g., comparing to national standards).

These indices represent different kinds of school policy towards student testing. Nevertheless, for all three of them, 23% of the overall variance can be identified as between-country-variance, which is quite a large share of the variation in assessment purposes. This proves that preference for any of these purposes is to a large extent driven by national (or system-level) policies, norms, and/or practices. Even the geographical/linguistic grouping of country does have some explanatory power: 10, 11, and 7% of variation in index 1, 2 and 3, respectively can be explained by differentiating between the ten "regions" described above. All in all, English speaking countries, especially the United Kingdom, New Zealand and the US, tend to rank at the top of this "hierarchy", while German-speaking countries are positioned at the bottom (see Fig. 7.4). This shows that traditional differences in how to use (or not to use) standardized student testing still prevail. On the other hand, it is instructive to see variation between countries belonging to the same geographical/linguistic group. E.g., Australia does not fit into the common pattern for English-speaking countries. Within the Nordic group of countries, strong differences can be found with regard to school evaluation purposes; these are rather weak in Finland and Norway, a little stronger in Denmark, and most important in Sweden and Iceland. At the same time, even in Sweden and Iceland, summative use of tests for judging individual students is quite low from an international point of view.

This is just descriptive data, but it may help a lot with public debates on testing. For example, claims that schools suffer from "Testeritis", meaning a move towards extensive testing practices, are quite popular among teacher unions in Germany nowadays,[5] but they can be challenged from a comparative point of view using these empirical data, as Germany has the lowest reported use.

7.3.5 *National Contexts for School Evaluation and Accountability*

While the previous section informed on school-based policies, which nevertheless were shown to be partly driven by national patterns, the following section looks into the wider context, mainly driven by accountability rules established by national, state, or district administration. Two questions in the PISA School Questionnaire which have been used in several cycles may help identify that context:

[5] https://www.focus.de/politik/deutschland/bildung-lehrer-machen-gegen-testeritis-an-schulen-front_id_3819831.html

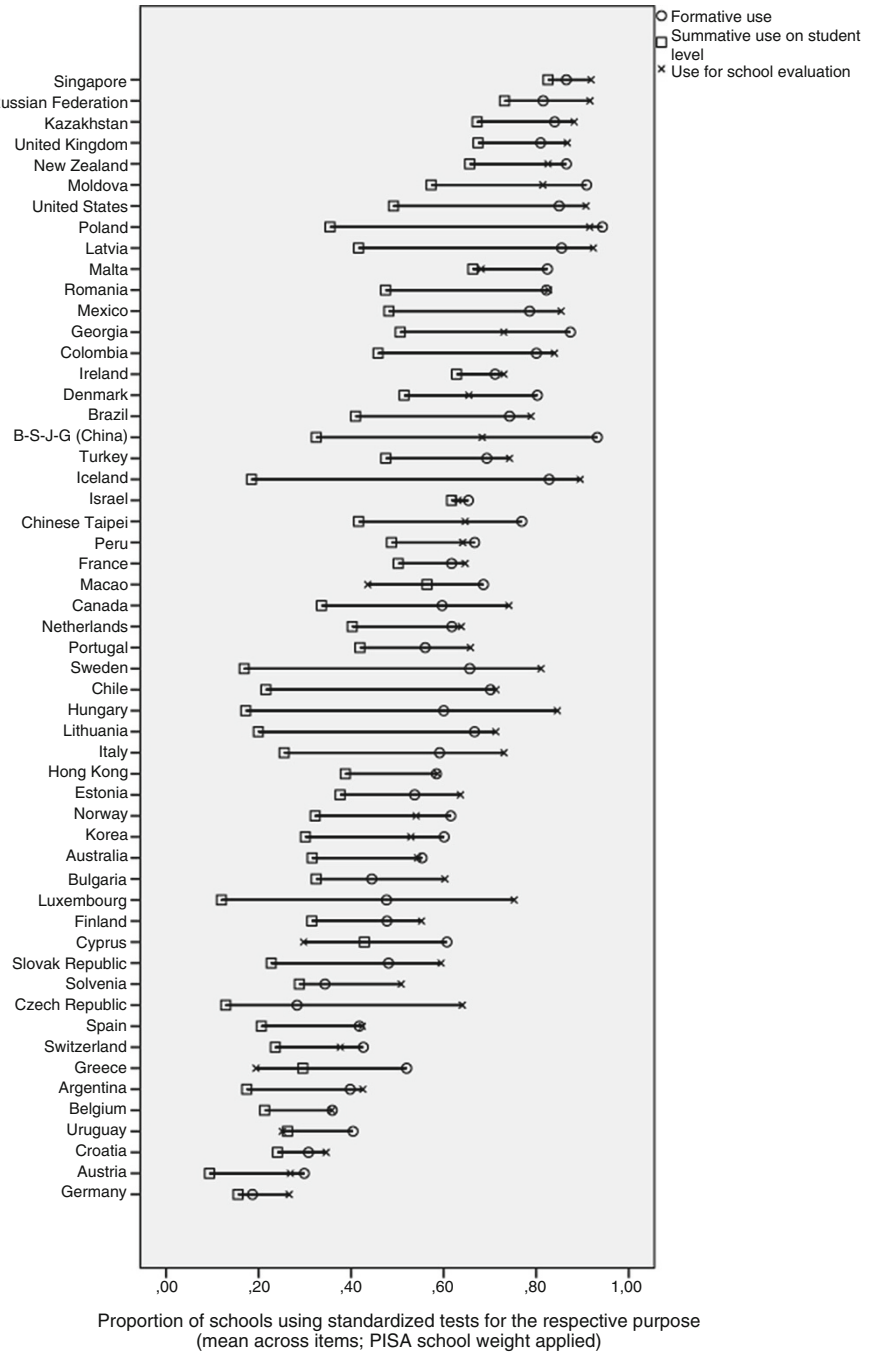

Fig. 7.4 Mean indices of using standardized tests for different purposes across countries. (Data on this question are not available for Japan)

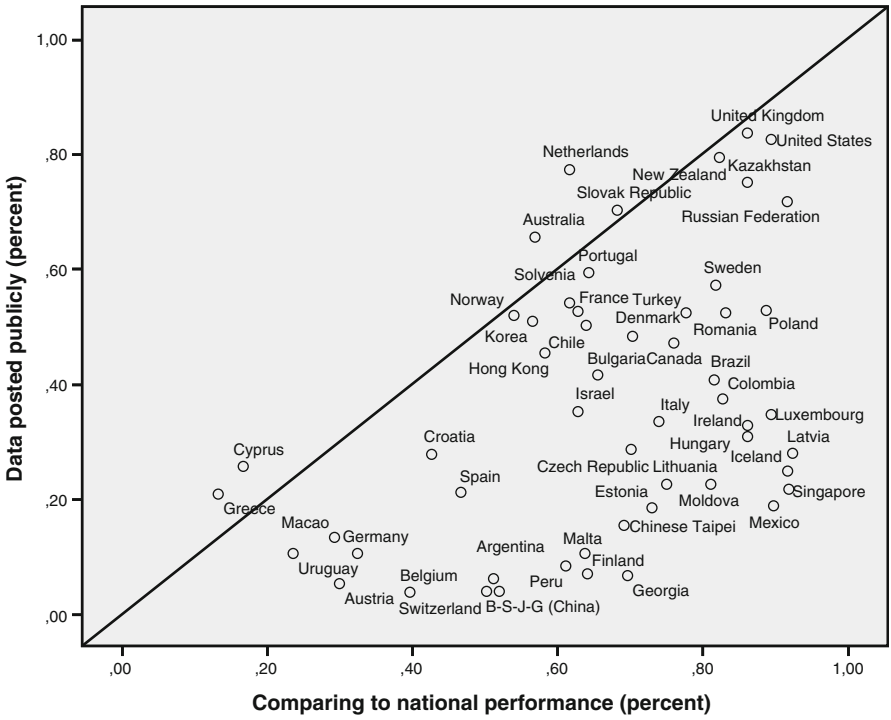

Fig. 7.5 Percentage of school principals reporting evaluation and accountability activities by country

- The question on Use of achievement data for accountability (SC036) asks, among other aspects of accountability, whether achievement data are posted publicly (e.g., in the media) by the school. The response format is "Yes/No".
- One item from the question on Purpose of standardized testing (SC035) asks whether standardized tests are used to compare the school to national, state or district performance. Again, the response format is "Yes/No".

With regard to the mechanisms explaining the impact of school evaluation discussed in Sect. 7.2.1, the first item is indicative of strong accountability practices, while the second item indicates a feedback-based approach to school improvement. Country-level means, as shown in Fig. 7.5, are in line with the patterns reported in the previous section.

- All in all, English speaking countries, especially the United Kingdom, New Zealand and the US, are international leaders in accountability, while German-speaking countries are positioned at the bottom.
- Once again, there are notable differences within country groups. E.g., Ireland and, to some extent (meaning: in some provinces), Canada tend to publish school results less often than other English-speaking countries. Comparing Norway,

Korea and Chile with Iceland, Singapore, and Mexico, respectively, we find complementary patterns within the Nordic, East Asian, and South American region.

- However, this analysis conveys two "global" messages that speak to the theories of school evaluation and accountability (see Sect. 7.2.1). These messages are easily visible by noting that the space below the diagonal is filled with all kinds of combinations of the two variables, while the space above the diagonal is practically void, with the exception of the Netherlands. (1) In practically all countries and systems, comparing test results to national or regional performance is reported more often than making results public. (2) High percentage of comparing is a prerequisite for high percentage of public reporting, but countries may show high prevalence for comparing while differing strongly in their prevalence for publication (see, e.g., Mexico vs. USA). Thus, there seems to be a hierarchy among the two mechanisms of school evaluation mentioned in Sect. 7.2.1. At a lower level, which might be called "soft accountability", schools are expected to compare their own test results to some external standards, probably as a kind of feedback mechanism.[6] At a higher level ("strong accountability"), data are made public to external users, such as parents, establishing a kind of quasi-market for schooling if it is used for the 'choice' of schools. These two kinds, or levels of accountability should not be mixed up.

7.3.6 Integrating the Picture: How Formative Assessment Practice Relates to Student Composition, Evaluation and Accountability Policies on the School Level

In this section, we take a closer look at formative assessment and feedback as perceived by students (see Sect. 7.3.3), using the IRT-based estimate of the latent construct named PERFEED in the PISA 2015 database. As PISA does not identify classrooms, analyses will be run on the school level, aggregating data from all students sampled within that school. Thus, we are dealing with formative assessment and feedback as a social practice of science education, as established in a given school, measured through the shared (mean) perception of 15 year-old students. Our research goal is to understand the relationship between this practice on one hand, the school's achievement level in science (represented by the school average of the first plausible value for science literacy), student composition (mean socio-economic status, as measured by the HISEI index of occupational status in PISA) and policies related to assessment, evaluation, or accountability (as measured by the indices discussed in Sects. 7.3.4 and 7.3.5) on the other hand.

[6]Bergbauer, Hanushek & Wößmann (Bergbauer et al., 2018, p. 17) classified this item as an instance of "school-based external comparisons".

7.3.6.1 Excursus on the Methodology of ILSA

Since the core variable used in this study, Perceived Feedback, does have metric, but not scalar invariance across countries, pooling data from all countries is not an option. The approach mostly used by econometricians (e.g. Bergbauer, Hanushek & Wößmann, 2018), introducing a fixed effect for every single country, helps to control for country-specific impact in regression-type models, but it does not take the lack of measurement invariance into account; it is bluntly ignorant with regard to measurement issues. Policy reports from IEA or OECD usually follow a third approach, running analyses in parallel for every country and summarizing findings in a qualitative way, as we did in Sects. 7.3.4 and 7.3.5 when interpreting descriptive findings across countries. However, this approach has severe drawbacks when applied to explanatory rather than descriptive findings. First, sample sizes may be quite small within individual countries. E.g., for studying purposes of student testing (SC035) on the school level, only 149 units would be available in Ireland, and 131 in New Zealand. Thus, the very asset of ILSAs to build a large international database would be lost. Second, summarizing and reporting results from complex models run in more than 70 countries is more an art than a science. Quite often, countries are grouped based on the results of within-country statistical tests, without any conceptual foundation.

When pooling student- or school level data for any analysis across countries, the appropriate approach seems to be the following: checking measurement invariance through rigorous methods (such as Multi-Group CFA, as applied in Sect. 7.3.3) and combining data sets from countries if and only if they meet the criteria of scalar invariance. Thus, in our case, we will work with the combined data from Australia, Ireland, New Zealand and the UK in the following.

We assume measurement invariance for the PISA science literacy test as documented in the Technical Report (OECD, 2017a). The standard measure of socio-economic status in PISA, the ESCS, cannot be assumed to be fully invariant across countries because item parameters from one of its components, the IRT measure of home possessions, have been shown to vary in meaning (Watermann, Maaz, Bayer, & Roczen, 2016). Therefore, we are using the international index of parental occupational status (HISEI) instead which is based on a transnational measurement approach in sociology (OECD, 2017a, p. 298).

For other predictive variables and control variables used in the following, we assume they do not represent latent variables. Rather, as discussed in Sect. 7.3.1, we treat them as "formative" indices, pragmatically summarizing reported activities of a certain kind (teaching and learning activities, professional activities at school, activities related to assessment, evaluation and accountability). Thus, there is no need to check measurement invariance for those variables.

7.3.6.2 Relating Perceived Feedback to Other School-Level Variables: In Search of the Proper *Explanandum*

In the light of conceptual debates on the relationship between ILSA and EER (see Sect. 7.1), it is interesting to note that OECD (2017b) treats Perceived Feedback both as a predictor for student achievement, *and* as an aspect of school practice that can be predicted from other variables.

The latter approach is visible in the following citation from Volume II of the PISA 2015 Policy Report (OECD, 2017b, p. 66): "Students in disadvantaged and rural schools were more likely to report that their teachers provide them with feedback . . . More perceived feedback is also associated with poorer performance in science, probably because low-performing students need and receive more feedback than better-performing students."

The former approach is at least implicitly used just one paragraph after the first citation: "Across OECD countries and after accounting for socio-economic status, students score between 5 and 17 points lower in science when they reported that their teachers use these strategies 'in many lessons' or 'every or almost every lesson' than when they reported that they use them in 'some lessons' or 'never or almost never'" (OECD 2017b, p. 66). Some pages further down (p. 73), the authors even talk about "impact on student performance" when, after controlling for reading and math achievement, they find a very low, but positive correlation between Perceived Feedback and student outcomes in Science Literacy.

Within our school-level analysis of these data, we illustrate both approaches in Table 7.4. Mean school-level achievement (PV1SCI) and mean school-level

Table 7.4 Prediction of mean student achievement vs. prediction of mean Perceived Feedback on the school level. (Data from AUS, IRL, NZL and UK; n = 1276 schools)

	Dependent variable	
Predictor[a]	Mean science achievement	Mean perceived feedback
Perceived feedback (PERFEED)	−.112 ***	
Science achievement (PV1SCIE)		−.181 ***
Socio-economic status (HISEI)	.654 ***	−.099 **
Goal-oriented curricular development	−.048 *	.061 *
Direct instruction (TDTEACH)	.131 ***	.307 ***
Inquiry-based teaching (IBTEACH)	−.037	.170 ***
Purpose of testing: Formative	−.001	.020
Purpose of testing: Summative/student level	.002	.127 ***
Purpose of testing: School level	.049[b]	.084 *
Data posted publicly	.022	.061 *
R^2	.516	.213

[a]Names in capital letters refer to aggregated student variables; all other predictors are based on the School Questionnaire. Parameters are standardized regression coefficients; [b]p < .10, * p < .05, ** p < .01, *** p < .001

Perceived Feedback (PERFEED) are used as *explanandum* (dependent variable) and predictor, respectively, in the left column, and vice versa in the right column. Additional predictors include the school's social composition (mean HISEI) and four indices of school policies in assessment, evaluation and accountability (see Sects. 7.3.4. and 7.3.5). Further, we include three control variables that cover major professional activities at the school level.

- One variable based on the school questionnaire (named "leadcom" in the PISA data base) summarizes the frequency of self-reported professional activities lead by the principal aimed at strengthening goal orientation in the curriculum. Items include "I ensure that teachers work according to the school's educational goals" and "I use student performance results to develop the school's educational goals".
- Two indices cover the frequency of different kinds of teacher activities in science classrooms as perceived by the students (school level average). (1) "Direct instruction" includes four core components of classroom teaching, namely "explaining ideas", "demonstrating an idea", "discussing student questions" and "conducting a whole class discussion". (2) "Inquiry-Based teaching", e.g., includes the following items: "Students are given opportunities to explain their ideas", "Students are allowed to design their own experiments".

Compared to the prediction of mean perceived feedback (right column), mean Achievement (left column) is obviously much easier to predict using this set of predictors; more than 50% of school-level variation can be explained. However, the prediction is mainly due to the relationship between achievement and socio-economic composition. Otherwise, Direct Instruction has a significant "effect" – showing that the achievement level is higher if schools succeed in implementing core activities of teaching across their science courses. However, as stated above, the direction of this relationship remains unclear: Probably, the more knowledgeable students are, the easier it is for teachers to enact core teaching activities. Perceived Feedback and goal-oriented curriculum development are associated with lower achievement – which may be interpreted as a case of reversed causality: both kinds of activities are probably implemented in response to low student outcomes.

Once again, we run into undecidable questions of directionality and causality when using cross-sectional ILSA data in attempts to "explain" the variation in achievement between schools. Regarding the topic of this chapter, assessment-related policies, however, the message is clear: Controlling for student composition and some basic kinds of professional activity, there is no relationship between any of the school policy indices and student achievement. The single index which is close to significance ($p < .10$) is *not*, as put forward by OECD in 2006, public posting of data. Rather, it is the use of data for school-level evaluation purposes such as comparing the school's test results with national standards. Referring to Sect. 7.3.5, we conclude that "soft accountability" tends to be related to a school's achievement level, while "strong accountability" is not.

The complementary research question, "Which schools implement formative feedback, under which conditions?" can be answered more clearly (see right column

in Table 7.4). Only 21% of the school-level variance can be explained, but there are more significant relationships with predictors:

- Schools that provide feedback with high frequency seem to be working 'against the odds', i.e. serving students with relatively low SES and low achievement.
- Also, students report higher intensity of formative assessment and feedback if they report higher levels of other teaching activities as well. Interestingly, the relationship with "Direct instruction" is much stronger than the relationship with "Inquiry-based teaching". In fact, it has been claimed in conceptualizations of teaching quality (e.g., Rosenshine & Stevens, 1986) that "Direct Instruction"--kind of approaches are successful, especially for groups of low achievers, because they provide lots of opportunities for feedback "on the fly" – contrary to inquiry-based learning where students spent more time working on their own.
- Most important, and not trivial at all, is the finding that several school-level policies and activities are related to students' shared perception of receiving feedback. There are relations with goal-oriented professional planning (remember one item referred to the use of performance results!) as well as with evaluation/accountability-related school policies. Both soft accountability policies (purpose of testing: school level) and strong accountability policies (data posted public) show small, but significant regression parameters. Even more relevant is the school's use of standardized testing for decisions about retention, grouping, and certification ("summative" use at the student level). The school's formative use of standardized tests in adaptive teaching and learning, curriculum development and parental involvement, however, seems not be specifically related to feedback approaches as perceived by students.

When choosing the use of formative assessment and feedback to be the *explanandum* (right column in Table 7.4), directionality is easier to decide. There is little sense in assuming that student perceptions of classroom practices in science courses determine school policies reported by principals. Hence, we may interpret any relationship *between* school policies and student-perceived practices as an effect of school evaluation policy *on* assessment and feedback practices embedded in science teaching. Our findings show that school policies do make a difference for everyday classroom practice experienced by students: The more tests are used for summative decision making, school evaluation and accountability, the more students report receiving formative feedback in classrooms. At least in the four English speaking countries covered here, formative and summative assessment seem to be connected rather than being mutually exclusive.

The findings reported in Table 7.4 can inform the development of new research questions. Here are some relatively general hypotheses to be tested in future studies:

1. Formative assessment and feedback is a classroom practice which is closely related to "traditional" (direct) teaching activities. It is still less integrated in "constructivist" teaching activities such as inquiry-based teaching. (Fischer, He and Klieme, submitted, seek to test this hypothesis across cultures.)

2. School policies on assessment, evaluation, and accountability can promote and foster classroom-based assessment if these policies are touching medium to high stakes, e.g. the use of test data in school evaluation or (even more important) the use of tests in decisions on student careers. (Above, we interpreted our data from four English speaking countries in line with this hypothesis. However, we do not know if the statement holds for other systems, especially when the overall level of assessment and evaluation is much lower than in English-speaking countries. Also, intervention studies should be implemented to test causal claims.)
3. (a) Policies regarding assessment, evaluation and accountability are mostly unrelated to student achievement at the school level, if student background and teaching practices are controlled for. (b) Effects on student outcomes may, if any, be expected from assessment data being used more often for (self-) evaluation on the school level. (Again, our cross-sectional findings from English-speaking countries need to be tested under different conditions. Statement (b) will be checked based on country-level trend data in the next section, and it is in line with findings reported by Bergbauer et al., 2018, using different methodology.)

7.3.7 Long-Term Changes in Assessment Strategies

In order to test Hypothesis 3(b) generated in the previous section, school policies on test use and school evaluation would have to be changed under controlled experimental or quasi-experimental conditions. Such studies are very hard to implement. However, as we have seen in Sect. 7.3.4, those school policies seem to be shaped by national (perhaps also state or district) contexts. Thus, instead of implementing (quasi-)experimental treatments, "natural" change in national policies may be studied in relation to changes in student achievement on the national level across different ILSA testing occasions.

ILSAs, in this case PISA, provide trend data that can allow this question to be studied. Bergbauer, Hanushek and Wößmann (Bergbauer et al., 2018, p. 16), mainly using PISA data, observed "a tendency for increased prevalence of the measures of standardized external comparison over time". Teltemann and Klieme (2016) focused on a single indicator which was available for almost all PISA cycles so far[7]: the item "In my school, assessments of 15-year old students[8] are used ... to compare the school to district, state or national[9] performance." In Sects. 7.3.4. and 7.3.5, we classified this item as indicating "use of assessment for school evaluation" and "soft accountability", in line with Bergbauer, Hanushek and Wößmann (Bergbauer et al., 2018, p. 17) who describe it as an instance of "school-based external comparisons". Unfortunately, in PISA 2015 this item was split up, one version addressing

[7]With the exception of PISA 2006.

[8]Later replaced by 'students in national modal grade for 15-year-olds'.

[9]The international school questionnaire allows for national adaptations regarding the level on which comparisons are made.

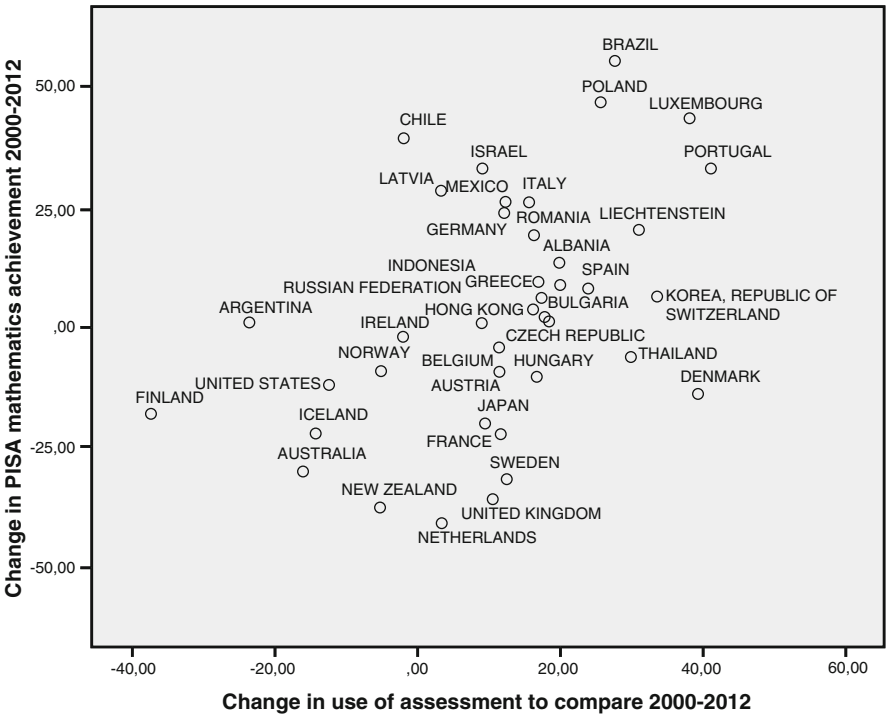

Fig. 7.6 Changes in use of assessment results to compare school performance with national, state, or district data (difference in percentage, as reported by school principals) and changes in mean mathematics achievement, PISA 2000–2012

"standardised tests", one version addressing "teacher-developed tests". Therefore, we restrict the discussion of system level change to the time interval from 2000 to 2012.[10]

As Teltemann and Klieme (2016) documented, from 2000 to 2012 the vast majority of OECD countries increased the use of assessment to compare with national or state/district performance. This global trend is also visible in Fig. 7.6. Furthermore, this figure shows that, on the country level, the change in use of "soft accountability" (horizontal axis) is related to the change in mean student achievement in mathematics (vertical axis). The correlation is $r = .449$ ($p < .01$), computed across all 35 out of the 55 countries (as selected in Sect. 7.3.2) for which both indices are available. Thus, "soft" accountability policy, as indicated by an increased proportion of schools comparing their assessment data to system-wide norms, seems to trigger gains in student achievement on the country level. PISA provides

[10]Changes in background questionnaire wording across cycles of measurement are yet another obstacle against analyzing trend data from ILSA's; cf. Singer et al., 2018, p. 64.

support for the hypothesis put forward at the end of Sect. 7.3.6: using assessment data for self-evaluation of schools has an impact on student outcomes.

How can this directional, even causal claim be justified? Gustafsson (2007) as well as Strietholt, Bos, Gustafsson, and Rosén (2014) provide several examples of such a "longitudinal cross-cohort design". By analyzing change on the country level, this method controls for fixed characteristics of the countries, thus reducing the effects of unobserved (omitted) variables. The method has been criticized by Singer et al. (2018, p. 59, pp. 65–66) mainly because it neglects the multi-level structure of the data, interpreting country level effects (e.g., the impact of mean use of computers at home on mean achievement) on a lower level (e.g., the impact of individual computer use on students' achievement). Similar criticism has been raised above, in Sect. 7.2.3, against the analysis of internal (classroom) testing by Bergbauer et al. (2018). However, accountability policies such as encouraging schools to compare their own performance to national norms are largely decided and implemented at the system level (see Sect. 7.3.4). Thus, when interpreting the finding from Fig. 7.6, we may indeed draw conclusions on the country (system) level.

So far, surprisingly few studies report quantitative analyses of system-level change. (See, however, Aloisi & Tymms, 2017; Lenkeit & Caro, 2014; Strietholt et al., 2014). The more waves of data collection PISA, PIRLS and TIMSS have implemented, the better researchers are prepared to model longitudinal data on the country (system) level. In order to strengthen such research, the meaning of aggregated measures has to be better understood (for examples, see Klieme, 2016; Rozman & Klieme, 2017), new research methods such as Bayesian modeling (Kaplan & Lee, 2018) should be adapted, and, most importantly, testable theories of educational change on the system level (e.g., Sun, Creemers, & de Jong, 2007) need to be further developed.

7.4 Summary and Discussion: Connecting ILSA and EER

The use and misuse of data from International Large Scale Assessments (ILSAs) has been repeatedly discussed both in public and by scholars in educational research (e.g., Rowan, 2002; Singer et al., 2018). Much of this discussion is dealing with how ILSAs may respond to the needs and practical questions of policy makers, professionals, and other stakeholders in education while avoiding overstatements, overgeneralizations or simplifications. While such pitfalls are common both in policy and in research (as shown in the introductory section of this chapter), experts agree that ILSAs should be conceptually based on Educational Effectiveness Research (EER) and adhere to rigorous methodological principles. The methodological foundations and challenges of ILSAs may be grouped into three major areas:

- Design: How to deal with the cross-sectional nature of individual ILSAs; how to use the trend design on the country level; how to specify the proper explanandum (explaining learning opportunities or school process quality rather than searching

for an explanation of student outcomes which is done almost "automatically" by most analysts); being cautious with regard to causality and the direction of effects;
- Sampling and data structure: how to deal with the multi-level nature of data, with clustered sampling, and with missing values (including missing by design): how to avoid ecological fallacies when interpreting country level relationships;
- Measurement: how to discriminate formative vs. reflective constructs, or manifest indices vs. latent (dispositional) measures; how to assess item and scale (mis)fit and establish measurement invariance across cultures.

Some of these issues were discussed and empirically illustrated across the chapter, but most of them could only be touched briefly. For example, issues of multi-level model specification were beyond the limits of the present chapter.

Both the fragile relationship with policy and practice as well as most of the methodological problems are shared features of ILSAs and EER. Therefore, each of the two paradigms of school research can mutually benefit from solutions developed by the other. Even more important seems to be the connection with regard to theoretical foundations and empirical findings. The present chapter explored such connections within one particular area of research and one particular study: the study of policies and practices of assessment and evaluation in PISA 2015.

First (in Sects. 6.2, 6.3.1 and 6.3.2), we showed how related constructs from EER have been taken up and implemented in PISA. Through PISA, national patterns of classroom assessment practices (Sect. 6.3.3), use of student assessment (Sect. 6.3.4), school evaluation and accountability policies (Sect. 6.3.5) have been identified. E.g., it turned out that English-speaking countries are similar in many respects, while full (scalar) invariance could be established for student reports from four countries only: UK, New Zealand, Australia, and Ireland can legitimately be ordered according to their respective prevalence of formative assessment and feedback. On the school level across those countries, formative assessment practices reported by students are related to summative use of assessment reported by principals. Thus, at least in some English-speaking countries, formative and summative assessment are positively connected rather than complementary. Regarding school accountability, the distinction between "soft accountability" (comparing performance with a national standard) and "strong accountability" (making test results public) proved to be informative. On the country level, soft accountability seems to be a necessary, but not sufficient prerequisite for strong accountability. For soft accountability only, a slightly positive relationship with student achievement on the school level (Sect. 6.3.6) and a positive impact on country-level math achievement (Sect. 6.3.7) were found.

From an EER perspective, we conclude that ILSA data help understand the effects of assessment, evaluation, and accountability on student outcomes. Experimental research has proven formative assessment and feedback to be an effective classroom practice (e.g., Kingston & Nash, 2011). Cross-sectional PISA data are not suitable for testing this claim. Nevertheless, they provide additional information on the variation of formative assessment practices within and between countries. In addition, PISA trend data show that summative assessments, such as national

surveys of student achievement, may trigger country-level growth in student outcomes as they provide feedback to schools. Overall, the pattern of our findings is consistent with a theory of school improvement based on "soft accountability", feedback and professional learning as the main mechanisms.

Policies and practices of assessment thus provide a showcase for how EER constructs can inform ILSA design, and how ILSA data in turn can inform the EER knowledge base.

References

Abrams, L. M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 79–98). New York, NY/London, UK: Teacher College/Columbia University.

Aloisi, C., & Tymms, P. (2017). PISA trends, social changes, and education reforms. *Educational Research and Evaluation, 23*(5–6), 180–220.

Altrichter, H., & Maag Merki, K. (2016). *Handbuch Neue Steuerung im Schulsystem* (2nd ed.). Wiesbaden, Germany: Springer.

Baker, D. P. (2009). The invisible hand of world education culture. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 958–968). New York, NY: Routledge.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Bayer, S. (2019). *Alle alles ganz lehren – Aber wie? Mathematikunterricht vergleichend zwischen den Schularten* [Omnes omnia omnino doceantur – But how? Comparing mathematics teaching between school tracks]. Phil. Dissertation. Goethe University, Frankfurt am Main.

Bayer, S., Klieme, E., & Jude, N. (2016). Assessment and evaluation in educational contexts. In S. In Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning. An international perspective* (pp. 469–488). New York, NY: Springer.

Bennett, R. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Bergbauer, A. B., Hanushek, E. A., & Wößmann, L. (2018, July). *Testing* (CESifo working paper no. 7168 7168 2018).

Bischof, L. M., Hochweber, J., Hartig, J., & Klieme, E. (2013). Schulentwicklung im Verlauf eines Jahrzehnts: Erste Ergebnisse des PISA-Schulpanels [School improvement throughout one decade: First results of the PISA school panel study]. *Zeitschrift für Pädagogik, special issue, 59*, 172–199.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Black, P., & Wiliam, D. (2004). The formative purpose. Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability: 103rd yearbook of the national society for the study of education, Part II* (pp. 20–50). Chicago, IL: University of Chicago Press.

Bogdandy, A. V., & Goldmann, M. (2009). The exercise of international public authority through National Policy Assessment. The PISA study of the OECD as a template for a new international standard legal instrument. *Zeitschrift für ausländisches öffentliches Recht und Völkerrecht, 69*, 51–102.

Bottani, N., & Tuijnman, A. C. (1994). The design of indicator systems. In A. C. Tuijnman & T. N. Postlethwaithe (Eds.), *Monitoring the standards of education* (pp. 47–78). Oxford, UK: Pergamon.

Bryk, A., & Hermanson, K. (1994). Observations on the structure, interpretation and use of education indicator systems. In OECD (Ed.), *Making education count: Developing and using international indicators* (pp. 37–53). Paris, France: OECD.

Buchholz, J. & Hartig, J. (2017). Comparing attitudes across groups: An IRT-based item-fit statistic for the analysis of measurement invariance. *Applied Psychological Measurement*. Advance online publication. https://doi.org/10.1177/0146621617748323.

Coburn, C., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Practice, 9*(4), 173–206.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness. A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.

Decristan, J., Klieme, E., Kunter, M., Hochweber, J., Büttner, G., Fauth, B., et al. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting students' science understanding? *American Educational Research Journal, 52*(6), 1133–1159.

Donaldson, S. I. (2004). Using professional evaluation to improve the effectiveness of nonprofit organizations. In R. E. Riggo & S. S. Orr (Eds.), *Improving leadership in nonprofit organizations* (pp. 234–251). San Francisco, CA: Wiley.

Elacqua, G. (2016). Building more effective education systems. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective*. Dordrecht, The Netherlands: Springer.

Ellwart, T., & Konradt, U. (2011). Formative versus reflective measurement: An illustration using work-family balance. *Journal of Psychology, 145*(5), 391–417.

Faubert, V. (2009). *School evaluation: Current practices in OECD countries and a literature review* (OECD Education working papers, no. 42). Paris, France: OECD.

Fischer, J., He, J., & Klieme, E.. (Submitted). *The structure of teaching practices across countries: A combination of factor analysis and network analysis*.

Fischer J., Klieme E., & Praetorius A-K.. (Submitted). *The impact of linguistic similarity on cross-cultural comparability of students' perceptions of teaching quality*.

Glas, C. A. W., & Jehangir, K. (2014). Modeling country specific differential item functioning. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large scale assessment* (pp. 97–116). Boca Raton, FL: CRC Press.

Gustafsson, J.-E. (2007). Understanding casual influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 37–63). Washington, DC: The Brookings Institution.

Harlen, W., & Deakin Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning* (EPPI-Centre Review, version 1.1∗). London: EPPI-Centre. https://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/ass_rv1.pdf?ver=2006-02-24-112939-763. Accessed 17 June 2016.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

He, J., Buchholz, J., & Klieme, E. (2017). Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology, 48*(3), 319–334.

He, J. & Kubacka, K. (2015). *Data comparability in the teaching and learning international survey (TALIS) 2008 and 2013* (OECD education working papers vol. 124). Paris, France: OECD.

Huber, S. G., & Skedsmo, G. (2016). Editorial: Data use – A key to improve teaching and learning. *Educational Assessment, Evaluation and Accountability, 28*(1), 1–3.

Jerrim, J. (2011). *"England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline"* (DoQSS working papers 11–09), Department of Quantitative Social Science – UCL Institute of Education, University College London.

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410.

Jude, N. (2016). The assessment of learning contexts in PISA. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective*. Dordrecht, The Netherlands: Springer.

Jude, N., & Kuger, S. (2018). *Questionnaire development and design for international large-scale assessments (ILSAs)*. Washington, DC: National Academy of Education.

Kaplan, D. & Lee, C. (2018). Optimizing prediction using Bayesian model averaging: Examples using large-scale educational assessments. *Evaluation Review*. Advance online publication. https://doi.org/10.1177/0193841X18761421

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.

Klieme, E. (2012). The role of large-scale assessments in research on educational effectiveness and school development. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 115–147). Heidelberg, Germany: Springer.

Klieme, E. (2016, December). *TIMSS 2015 and PISA 2015 -How are they related on the country level?* (DIPF working paper). https://pisa.dipf.de/de/pdf-ordner/Klieme_TIMSS2015andPISA2015.pdf

Klieme, E. (2018, February). *Alles schräg* (Biased findings). https://www.zeit.de/2018/07/pisa-studie-oecd-politik-eckhard-klieme.

Klieme, E., Jude, N., Baumert, J., & Prenzel, M. (2010). PISA 2000–2009: Bilanz der Veränderungen im Schulsystem (Making up the balance of changes in the school system). In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Koeller, M. Prenzel, W. Schneider, & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt (Making up the balance a decade after).* Münster, Germany: Waxmann.

Klieme, E., & Kuger, S. (2015). PISA 2015 context questionnaires framework. In *PISA 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy* (pp. 101–127). Paris, France: OECD.

Klieme, E., & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht (Teaching quality from a student perspective: Culture-specific profiles, regional differences, and relationships with teaching effects). In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, (Hrsg.). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland (Hrsg.)* (S. 334–359). Opladen, Germany: Leske + Budrich.

Kuger, S., Klieme, E., Jude, N. & Kaplan, D. (Eds.) (2016). *Assessing contexts of learning: An international perspective*. Dordrecht, The Netherlands: Springer.

Kuger, S., Klieme, E., Lüdtke, O., Schiepe-Tiska, A., & Reiss, K. (2017). Mathematikunterricht von Schülerleistungen in der Sekundarstufe: Zur Validität von Schülerbefragungen in Schulleistungsstudien (Mathematics teaching and student achievement in secondary education: The validity of student surveys in school achievement studies). *Zeitschrift fuer Erziehungswissenschaft, 20*(2), 612. https://doi.org/10.1007/s11618-017-0750-6

Lenkeit, J., & Caro, D. H. (2014). Performance status and change – Measuring education system effectiveness with data from PISA 2000–2009. *Educational Research and Evaluation, 20*(2), 146–174.

McMillan, J. H. (2007). Formative classroom assessment: The key to improving student achievement. In J. H. McMillan (Ed.), *Formative classroom assessment. Theory into practice* (pp. 1–7). New York/London: Teacher College, Columbia University.

Nevo, D. (2002). Dialogue evaluation: Combining internal and external evaluation. In D. Nevo (Ed.), *School-based evaluation: An international perspective* (pp. 3–16). Amsterdam, The Netherlands/Oxford, UK: Elsevier Science.

OECD. (2005). *Formative assessment: Improving learning in secondary classrooms*. Paris, France: OECD.

OECD. (2007). *PISA 2006. Science competencies for tomorrow's world*. Paris, France: OECD.

OECD. (2013). *Synergies for better learning. An international perspective on evaluation and assessment. OECD reviews of evaluation and assessment in education*. Paris, France: OECD.

OECD. (2014). *PISA 2012 technical report*. Paris, France: OECD.

OECD. (2017a). *PISA 2015 technical report*. Paris, France: OECD.

OECD. (2017b). *PISA 2015 Results, Volume II. Policies and practices for successful schools*. Paris, France: OECD.

OECD & Vodafone Stiftung. (2018, January). *Erfolgsfaktor Resilienz* (Success factor resilience). https://www.vodafone-stiftung.de/uploads/tx_newsjson/Vodafone_Stiftung_Erfolgsfaktor_Resilienz_01_02.pdf

Rakoczy, K., Klieme, E., Leiss, D., & Blum, W. (2017). Formative assessment in mathematics instruction: Theoretical considerations and empirical results of the Co^{2}CA project. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Eds.), *Competence assessment in education: Research, models and instruments* (pp. 447–467). Cham, Switzerland: Springer.

Reckwitz, A. (2002). Toward a theory of social practices: A development in culturalist theorizing. *European Journal of Social Theory, 5*(2), 243–263.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Rowan, B. (2002). Large-scale, cross-National Surveys of educational achievement: Promises, pitfalls, and possibilities. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-National Surveys of educational achievement* (pp. 319–350). Washington, DC: National Academic Press.

Rozman, M., & Klieme, E. (2017). *Exploring cross-national changes in instructional practices: Evidence from four cycles of TIMSS* (Policy brief vol. 13). Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*(1), 31–57.

Ryan, K. E., Chandler, M., & Samuels, M. (2007). What should school-based evaluation look like? *Studies in Educational Evaluation, 33*(3–4), 197–212.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Sanders, J. R., & Davidson, E. J. (2003). A model for school evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation. Part one: Perspectives/ part two: Practice* (pp. 807–826). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Scheerens, J. (2002). School self-evaluation: Origins, definitions, approaches, methods and implementation. In D. Nevo (Ed.), *School-based evaluation: An international perspective* (pp. 35–69). Amsterdam, The Netherlands/Oxford, UK: Elsevier Science.

Scheerens, J., Glas, C. A., & Thomas, S. M. (2003). *Educational evaluation, assessment, and monitoring. A systemic approach*. Lisse, Ther Netherlands/Exton, PA: Swets & Zeitlinger.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher, 44*(7), 371–386.

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (pp. 623–646). Westport, CT: Rowman and Littlefield Publishers.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Singer, J., Braun, H., & Chudowsky, N. (Eds.). (2018). *International education assessments – Cautions, conundrums, and common sense*. Washington, DC: National Academy of Education.

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education, 118*(2), 113–141.

Strietholt, R., Bos, W., Gustafsson, J.-E., & Rosén, M. (Eds.). (2014). *Educational policy evaluation through international comparative assessments*. Münster, Germany: Waxmann.

Sun, H., Creemers, B. P. M., & de Jong, R. (2007). Contextual factors and effective school improvement. *School Effectiveness and School Improvement, 18*(1), 93–122.

Teltemann, J., & Klieme, E. (2016). The impact of international testing projects on policy and practice. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 369–386). New York, NY: Routledge.

van de Vijver, F. & He, J., (2016), Bias assessment and prevention in non-cognitive outcome measures in PISA questionnaires. In Kuger, S., Klieme, E., Jude, N. & Kaplan, D. (eds.). Assessing contexts of learning world-wide: An international perspective. New York, NY: Springer Science, p. 229–253. 24 p.

van de Vijver, F. J. R. (2018). Towards an integrated framework of Bias in noncognitive assessment in international large-scale studies: Challenges and prospects. *Educational Measurement: Issues and Practices, 37*(4), 49–56. 8p.

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School Effectiveness and School Improvement, 14*(3), 321–349.

Watermann, R., Maaz, K., Bayer, S., & Roczen, N. (2016). Social background. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning: An international perspective* (Methodology of educational measurement and assessment) (pp. 117–145). Springer. https://doi.org/10.1007/978-3-319-45357-6

Wößmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Cheltenham, UK: Edward Elgar.

Wyatt-Smith, C. (2014). *Designing assessment for quality learning: The enabling power of assessment*. Heidelberg, Germany: Springer.

Part II
Examples of Educational Effective Research from Around the Globe

Introduction

Following on from the historical context, theoretical framing, and extensions of EER presented in Part 1, Part 2 presents snapshots of EER from around the globe. These examples represent the state-of-the-art in EER in different parts of the world at the time of publication. The chapters were chosen to include research studies conducted by leading researchers in the EER tradition and to offer recent exemplars that illustrate the range of designs and approaches that are being used in current international EER in diverse contexts.

In some cases (e.g. Chaps. 8, 9, and 13), the examples presented in this section represent an expansion of EER research in previously under-represented regions. In other chapters, the reader will encounter examples of research from regions in in which EER has developed and matured (e.g. Chaps. 10, 11, and 12). The research methodologies used in these six chapters are broadly representative of those employed in EER and educational improvement, and while there is an emphasis on quantitative approaches (e.g. longitudinal surveys), qualitative approaches are also represented (e.g. case studies in Chap. 12, and less formally as a way to provide context in Chap. 8). A common feature across many of these chapters is the need for research that is grounded in partnerships.

In total, these chapters provide a methodologically-diverse and geographically comprehensive view of the range of topics, approaches and problems that current-day EER seeks to address.

Chapter 8
Educational Effectiveness Research in Africa: The Case of the Democratic Republic of the Congo (DRC)

Gratien Bambanota Mokonzi, Jan Van Damme, Bieke De Fraine, Paul Masimango Vitamara, Gaston Mabela Kimbuani, Augustin Tshite Mukiekie, Stanislas Lukula Maroyi, Oscar Asobee Gboisso, and Jean Paul Legono Bela

8.1 Introduction

Ever since Scheerens' (1992) call to increase the number of educational effectiveness studies, a lot of research has been carried out in the developed countries (Sammons et al., 2008; Van Damme, De Fraine, Van Landeghem, Opdenakker, & Onghena, 2002 ...). But in African countries in general, and in Sub-Saharan countries specifically, little has been done in this regard. Apart from some comparative studies conducted in Southern and Eastern African sub regions (ACER, 2015; Hungi, 2011; Hungi et al., 2010) as well as in French-speaking Africa (PASEC, 2015), the more systematic South-African research studies using multilevel models (Carrim & Shalem, 1999; Christie, 1998; Harber & Muthukrishna, 2000; Howie, 2006) must be considered as exceptions that prove the rule. In Africa, research focusing on educational effectiveness has typically been initiated by occidental researchers and supported by external financial aid. This type of research has known a less rapid development here than in other parts of the world, partly due to the fact that it had to face the competition of other urgent issues related to the expansion of the educational system (Fleisch, 2007). In addition, only very few countries in Sub-Saharan Africa have participated in international assessment

G. B. Mokonzi · P. M. Vitamara · G. M. Kimbuani · A. T. Mukiekie · S. L. Maroyi · O. A. Gboisso · J. P. L. Bela
Service de Planification et d'Évaluation en Éducation, Université de Kisangani, Kisangani, The Democratic Republic of the Congo
e-mail: gratien.mokonzi@unikis.ac.cd; paulvitamara@yahoo.fr; gaston.kimbuani@unikis.ac.cd; augustin.mukiekie@unikis.ac.cd; stanislas.maroy@unikis.ac.cd; oscar.gboisso@unikis.ac.cd; jean-paul.bela@unikis.ac.cd

J. Van Damme (✉) · B. De Fraine
Center for Educational Effectiveness and Evaluation/KU Leuven, Leuven, Belgium
e-mail: jan.vandamme@kuleuven.be; bieke.defraine@kuleuven.be

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_8

programs such as Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). Further, studies using experimental and quasi-experimental approaches have not been undertaken to the same extent in the different regions of Sub-Saharan Africa. Out of the 56 studies of this type that have been included in a meta-analysis by Conn (2014), only three were carried out in Central Africa and none in the Democratic Republic of the Congo (DRC).

In response to this lack of research on educational effectiveness in the DRC, the 'Service de Planification et d'Évaluation en Éducation' of the University of Kisangani and the Center for Educational Effectiveness and Evaluation of KU Leuven set up a partnership in 2010 to conduct a longitudinal study on educational effectiveness in primary and secondary schools in the Oriental Province of the DRC. This chapter addresses results from this research project, restricting the focus to findings related to primary schools.

8.2 Research Context

Since the convention of 1977 concluded between the Congolese state and the Churches,[1] there are three distinct categories of schools in the DRC: 'Ecoles Publiques Non Conventionnées'[2] (EPUNC), 'Ecoles Publiques Conventionnées'[3] (EPUC) and 'Ecoles Privées'[4] (EPR). The EPUNC are both funded and run by the government; the EPUC are generally funded by the government and run by the Churches, while the EPR are funded and run by their sponsors who are individuals or associations, more in specific Churches and NGOs. According to the 'Cellule Technique pour les Statistiques de l'Éducation' (2017), the DRC counts 51,578 primary schools of which 18% are EPUNC, 64% EPUC and 14% EPR (see Table 8.1). These schools provide education to 14,301,438 pupils. In the Orientale Province 77% of the pupils attend EPUC (see Table 8.2).

Since the armed conflicts that the country has experienced in the 1990s have come to an end, the educational system of the DRC -- especially at primary school level -- has made significant progress, especially with regard to access, attendance, equity and retention. Indeed, not only does 95% of a given cohort have access to

[1]In the history of the education system of the DRC, schools were initially created and run by protestant and catholic missionaries from the end of the nineteenth century until the beginning of the 1950 decade. The official schools were created only at the end of the 1940s. In order to extend its power on the whole education system, the dictatorial regime of Mobutu processed, in 1973, in the nationalization of church schools. However, four years after (1977), these schools were given again to churches through the convention signed between the Congolese Government and the protestant, catholic and Kimbangu churches. The convention obviously claims that only the Government has the organization power of schools but churches are simple managers.

[2]Public schools not under convention.

[3]Public schools under convention.

[4]Private schools.

Table 8.1 Number of primary schools of the DRC in 2014–2015

	EPUNC		EPUC		EPR		Total
	Number	%	Number	%	Number	%	Number
DRC	9205	18%	35,294	68%	7079	14%	51,578
Orientale province	931	16%	4253	76%	422	8%	5606
% Orientale Province	10%		12%		6%		11%

Source: 'Cellule Technique pour les Statistiques de l'Éducation'(2017)

Table 8.2 Number of pupils in the primary schools of the DRC in 2014–2015

	EPUNC		EPUC		EPR		Total
	Number	%	Number	%	Number	%	
DRC	2,430,756	17%	9,899,902	69%	1,970,780	14%	14,301,438
Orientale province	246,847	15%	1,245,737	77%	119,065	7%	1,611,649
% Orientale Province	10%		13%		6%		11%

Source: 'Cellule Technique pour les Statistiques de l'Éducation' (2017)

primary schools as of 2014–2015, but the gross enrolment rate also increased from 90% in 2006–2007 to 110.3% in 2014–2015 ('Cellule Technique pour les Statistiques de l'Éducation', 2008, 2017). The latter statistic might initially appear odd; this is an indication of the high number of children being older than 12 years but still in primary school. The increase in enrolment was accompanied by a higher gender parity of which the index increased from 0.80 in 2006–2007 to 0.90 in 2014–2015. In other words the rate of girls' school attendance was at 80% and 90% compared to that of the boys respectively in 2006–2007 and in 2014–2015.

In addition to improved access and attendance, the system also underwent a positive change in terms of pupil retention and progression. Indeed, even though the drop-out rate remains high, primary schools have recorded a decrease in the number of repeaters, declining from 25% in 2007 to about 10–12% in 2012–2013 ('Cellule Technique pour les Statistiques de l'Éducation', 2008, 2014). In the same way, 7 out of 10 children accessing school reach the 6th grade of primary school in 2011–2012 compared to 3 out of 10 in 2000–2001 and 5 out of 10 in 2006–2007 (UNESCO & IIEP 'Pôle de Dakar', 2014).

Despite these positive developments, the educational system in the DRC still faces important challenges, particularly with regard to equity and quality. According to the 'Ministère de l'Enseignement primaire, secondaire et professionnel' (2012), 28.9% (or 7,375,875 children and adolescents between the ages of 5 and 17 years) do not attend school. One of the key factors that affects school attendance is the unsystematic application of the principle of free primary education, despite the affirmation of this principle by the Constitution of the Republic and General Law nr. 14/004 of 11 February 2014 ('Présidence de la République', 2014).

In most of the schools, the conditions under which the pupils receive education do not meet the standards warranted to provide a high-quality education. It is not uncommon to encounter overcrowded classes led by unqualified and untrained teachers, or schools lacking basic equipment such as benches, blackboards, teaching material and text books (Mokonzi, 2009).

In terms of quality, pupil achievement remains low, especially with regard to reading, writing and mathematics (Banque Mondiale, 2005; De Herdt, Marivoet, & Muhigirwa, 2015; Mokonzi & Issoy, 2002). According to a 'Programme d'Analyse des Systèmes Educatifs de Confemen' (PASEC, 2011) study, 26% of 2nd-grade pupils and 51% of 5th-grade pupils in the DRC have learning difficulties. With regard to mathematics, the Early Grade Reading Assessment (EGRA) tests administered in 2012 showed that a significant proportion of the 2nd- and 4th-grade primary school pupils do not meet the level of achievement required according to the official curriculum (UNESCO & IIEP 'Pôle de Dakar', 2014).

8.3 Research Motivation

To face the challenges of its educational system, the government of the DRC adopted in March 2010 a 'Stratégie de développement de l'enseignement primaire, secondaire et professionnel 2010/11-2015/16'. This strategy has three major objectives: (a) to improve access, accessibility, equity and retention at all educational levels, especially at the basic educational level; (b) to improve educational quality; and (c) to strengthen the governance ('Ministère de l'Enseignement Primaire, Secondaire et Initiation à la Nouvelle Citoyenneté' et al., 2015).

Underlying assumptions are that the quality of education can be improved by (a) restricting class size to 40 pupils per class, (b) recruiting new teachers, (c) creating canteens, (d) distributing basic school text books to pupils, (e) providing continued training to teachers, and (f) offering school furniture and teaching material ('Ministère du Plan', 2011).

Will these strategies improve the effectiveness of Congolese schools? What factors should be taken into account to improve this effectiveness? Considering the need to carry out educational effectiveness research in the DRC as stated in the introduction of this chapter, we have initiated this educational effectiveness research on primary and secondary schools of the Oriental Province of the DRC to address these questions.

8.4 Analytical Modelling Approaches

Educational effectiveness research (EER) has a long tradition. It intensified in the 1990s and benefitted from the contributions of several streams of research, more specifically "education production functions" (Coleman et al., 1966; Hanushek, 2003; Krueger, 1999), the "process-product" approach (Gage, 1963), and the "Aptitude-Treatment-Interaction" approach (Cronbach, 1957). EER aims to explain different educational products through (multilevel) models which consider not only teachers and classes, but also the characteristics of schools and even educational systems (Van Damme et al., 2009).

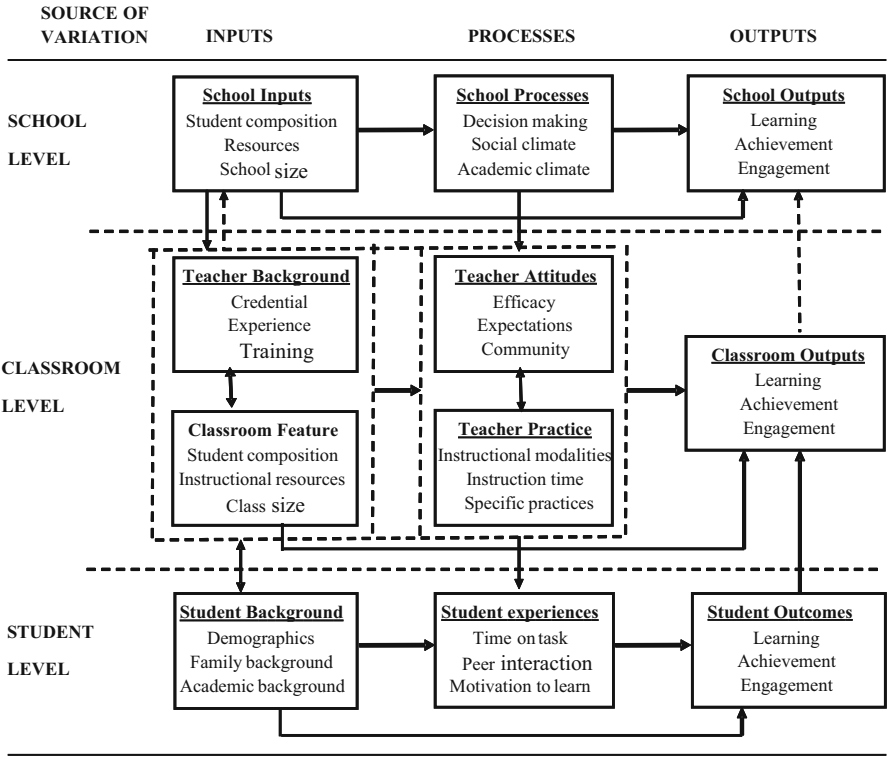

Fig. 8.1 Theoretical multilevel model for the study of the class and school effects

Since the mid-90s, syntheses of educational effectiveness research have been produced and conceptual models including the results of studies derived from production functions, the process-product models and considerations of educational characteristics have been developed (e.g. Creemers, 1994; Creemers & Kyriakides, 2008; Palardy & Rumberger, 2008; Scheerens, 1990; Scheerens & Bosker, 1997; Teddlie & Reynolds, 2000). Above and beyond their specific characteristics, these models take into account the hierarchical nature of the educational systems in which pupils are affected by the classes which are, in turn, affected by the schools (the same observation applies to the teachers). They admit the influence of inputs, processes and outputs at each one of these levels. One of these models is the model of Palardy and Rumberger (2008, p. 118) to which the research in this chapter is referring (see Fig. 8.1).

8.5 Methodological Choices

Four methodological choices were predetermined to carry out this research. The first consisted of conducting a longitudinal study, as the school effect can only be effectively established over the long term (Creemers, Kyriakides, & Sammons,

2010). To this end, we followed over three consecutive years a sample that initially consisted of 4787 pupils from the 4th grade of primary school. Data were collected at four time points: At the beginning of 4th grade (2011), and at the ends of 4th grade (2012), 5th grade (2013) and 6th grade (2014).

The second methodological choice defining this research regards the necessity to collect data related to pupils' initial levels of achievement. This option justifies the data collection at the beginning of 4th grade.

The third methodological choice concerns the necessity to measure both the cognitive and non-cognitive aspects of the learning process. To do so, this research measured pupil achievement in reading, writing and mathematics (to measure the cognitive aspects) as well as their self-concept and their feeling of well-being at school (to account for some non-cognitive aspects).

Finally, the fourth methodological choice consisted of using a multilevel model. In line with this purpose, data were collected at pupil, class and school level.

8.6 Research Questions

In addition to providing a general overview of the effectiveness study on schools and classes, we intend to address two specific questions in this chapter:

(a) What is the size of school and class effects on writing, reading and mathematics achievement in grades 4, 5 and 6 in the DRC education system?
(b) Which variables explain the effects on mathematics achievement in grade 4?

8.7 The Sample of Schools and Classes

The sample of schools and classes was selected in the two major cities of the Oriental Province, i.e. the province where the University of Kisangani is located. To this end, we used a stratified, weighted sampling of schools of the cities of Kisangani and Bunia. Out of the 280 schools established in both these cities as of 2010, 50 schools—i.e. about 18% of the primary schools—were selected. As a result of the structure of the educational system in the DRC, three strata were identified for the sampling of schools, corresponding to the three school types: (a) non-conventioned public schools, (b) conventioned public schools and (c) private schools. The sampling procedure consisted of a random selection of a proportional number of schools within each category. The next stage consisted of selecting classes within schools: if a school had only one or two classes of the 4th grade, all of these classes were systematically retained in the sample; if a school had more than 2 classes of the 4th grade, only 2 classes were selected at random to be part of the sample. Finally, all pupils in the selected classes were invited to participate in the research project. This sampling procedure resulted in a selection of 50 schools and 82 4th grade classes. But as one school in Bunia informed us during the first stage of data collection that it

Table 8.3 The sample of schools and classes

City	Schools	2011–2012		2013		2014	
		4th grade classes	pupils of 4th grade	5th grade classes	pupils of 5th grade	6th grade classes	pupils of 6th grade
Kisangani	36	59	3271	60	3000	70	2886
Bunia	13	21	1516	21	1132	19	910
Total	49	80	4787	81	4132	89	3796

didn't want to take part in the research, the analytic sample consisted of 49 schools, 80 classes and 4787 pupils in the 4th grade (see Table 8.3).

The number of schools remained constant throughout the study, while the number of classes and pupils changed over the years (see Table 8.3). So, in 2013, we selected in every school that had been retained no more than two classes of the 5th grade, grouping the largest number of pupils who were involved in the survey conducted in 2011–2012. In 2014, however, data collection was realized in all of the classes and all pupils in 6th grade, within the 49 primary schools selected in 2011, were included in the research.

The fluctuation of the number of pupils during the first 3 years of this research is mainly explained by pupil absenteeism, which is partly due to the fact that people are often unable to pay school fees for their children. In addition, some pupils frequently changed schools or had to repeat classes. So, out of the 4787 subjects selected in the 4th grade of primary schooling during 2011–12, only 1868 are included in the 2013 sample in 5th grade and only 1,202 in the 2014 sample in 6th grade. Similarly, out of the 4,132 pupils in 5th grade who took part in the 2013 survey, only 2008 are included among the 3796 pupils involved in the 2014 data collection. But it also has to be pointed out that each year, new pupils have been added to the sample. These pupils were either repeaters or came from other schools, schools which often were not included in the sample.

8.8 Variables and Research Instruments

The realization of this research required the use of various data collection instruments; some of them were used with pupils, others with teachers and school principals. The variables explored via these instruments are derived from three main sources: (a) international research on educational effectiveness, (b) observation of the instructional practice in the Congolese schools and (c) discussions held with teachers, principals and school inspectors.

8.8.1 Variables and Instruments Relevant to Pupils

As mentioned above, pupil data were collected at four time points: At the beginning of the 4th grade and at the end of 4th, 5th and 6th grade. The data collected at the beginning of 4th grade included three types of variables: (a) pupils' demographic information, school and family background; (b) pupils' cognitive characteristics, and (c) pupils' non-cognitive characteristics.

While the demographic, school- and family-related data were collected through a questionnaire administered to pupils, cognitive and non-cognitive characteristics were respectively assessed through tests based on school curricular programs and two different questionnaires (see Table 8.4).

The exploration of cognitive characteristics consisted of measuring pupils' achievement in reading, writing and mathematics. The survey included two reading dimensions: reading fluency and reading comprehension. Reading fluency consisted of word decoding. To test this ability, pupils were asked to read three categories of words: easy or mono-syllable words; words of medium difficulty or two-syllable words; and difficult words, each consisting of more than two syllables. For each category, consisting of 150 words, pupils were invited to decode as many words as possible within 1 min. To assess pupils' reading comprehension, they were asked to read a text taken from a 3rd grade textbook, and then to answer ten questions related to that text. Although the use of this textbook is imposed by the ministry on all schools in the DRC, the questions of the reading comprehension test were set up by the 'Service de Planification et d'Évaluation en Éducation' in cooperation with a couple of teachers of some of the city's primary schools. Writing ability was tested through the dictation of a paragraph of a text that had also been taken from the 3rd grade textbook. Pupils' initial level of achievement in mathematics was measured through a test containing 36 questions focusing on numbers, fractions, operations and measurement.

To assess non-cognitive aspects of learning, we used questionnaires related to self-concept and well-being at school. The first questionnaire was taken from the French version (Dierendonck, 2008) of the "Self-Description Questionnaire I"

Table 8.4 Variables and data collection instruments administered to pupils in 4th grade at the beginning of the year

Variables	Instruments
Sociodemographic characteristics	Pupil background questionnaire
Educational background	
Socio-familial characteristics	
Cognitive characteristics	Reading fluency
	Reading comprehension
	Writing test
	Mathematics test
Non-cognitive characteristics	Self-description questionnaire
	Questionnaire about well-being at school

Table 8.5 Variables and data collection instruments administered to pupils in the 4th grade at the end of the year

Variables	Instruments
Class climate	Pupil questionnaire on class and school
Perception of the school teacher as a teacher	
Feelings about school	
Cognitive characteristics	Reading fluency
	Reading comprehension
	Writing test
	Mathematics test
Non-cognitive characteristics	Self-description questionnaire
	Questionnaire about well-being at school

developed by Marsh (1988). The version used in 4th grade included 48 Likert-type items. The well-being questionnaire was taken from the Longitudinal Research in Secondary Education (LOSO) project conducted by the Center for Educational Effectiveness and Evaluation at KU Leuven. This scale included eight items.

Most of the instruments administered at the beginning of 4th grade were also administered at the end of the school year. However, the mathematics test administered at the beginning of the year covered content from 3rd grade, while the test at the end of the year was based on content from 4th grade. While the test based on the 3rd grade had 36 questions, that of the 4th grade had 40 questions. In addition, at the end of the year, the background questionnaire was replaced by a pupil questionnaire on class and school characteristics. The latter questionnaire explores the pupils' feelings about their classes, class climate and how they perceive the characteristics of their teachers (see Table 8.5).

The same characteristics were studied at the end of the 4th, 5th and 6th grade. To assess pupils' cognitive levels of achievement, the reading fluency test was the same as the one used in the 4th grade; the tests to evaluate pupils' achievement in reading comprehension, writing and mathematics were based respectively on the curricula of the 5th and the 6th grade. To test pupil non-cognitive outcomes, alongside the well-being questionnaire, only two dimensions of the self-description questionnaire (academic self-concept in mathematics and reading) have been administered in the 5th and 6th grade[5].

8.8.2 Variables and Instruments Relevant to the Class Level

The data related to the class level were obtained from the teachers (alongside the pupil questionnaire on class and school). For an overview, see Table 8.6. In addition

[5]The modification of the questionnaire dimensions of self-description was due to the fact that the questionnaire applied in 4th grade appeared to be too long (48 questions) for pupils.

Table 8.6 Variables and data collection instruments administered at the class level

Variables	Instruments
General information	Teacher questionnaire
Class climate	
Teaching practices	
Concepts about homework	
Learning opportunities	Questionnaire about learning opportunities for mathematics

Table 8.7 Variables and data collection instruments applied at school level

Variables	Instruments
Assessment of the school principal	School questionnaire for teachers
Perception of school climate	
Collaboration amongst teachers	
Pedagogical concepts of the teachers	
General characteristics of the school principal	School principal questionnaire
Views of the school principal about his job	
Views of the school principal about school	
Views of the school principal about the work of the teachers	
Perceptions about the collaboration amongst teachers	
Assessment of the pupils	
Pedagogical concepts of the school principal	

to the general characteristics of the class (class size, number of girls) and the teacher (qualification, experience, initial and continued training), the study considered class climate (16 items), teaching practices (26 items) and homework (10 items). These data were collected through a questionnaire that was administered to the teachers of the classes retained in the sample. Next, the study focused on learning opportunities for mathematics. These were also gathered through a questionnaire that was administered to teachers of classes involved in the study.

8.8.3 Variables and Instruments Relevant to the School Level

Most of the school-related data were obtained from teachers and school principals (see Table 8.7). The data collection included a school questionnaire which was administered to 5 teachers per school, asking them to assess their school principal (26 items), the way they perceived their schools (41 items), their perception of teachers (12 items) and pupils (7 items), collaboration amongst teachers (23 items), and pedagogical conceptions of teachers (11 items). Other school-related data were collected through a questionnaire submitted to school principals. In addition to general characteristics of the school principal (gender, age, qualification, initial and continued training, experience) and of the school (school size, number of qualified teachers, number of girls), this questionnaire explored the principal's point

of view regarding his/her job (26 items), his/her school (14 items), the work of the teachers (27 items), collaboration amongst teachers (24 items), pupil assessment (7 items) and pedagogical conceptions of principals (11 items).

8.9 School and Class Effects

In this section we answer successively both research questions presented above. To determine the size of school and class effects, we conducted a multilevel analysis of pupils' achievement in writing, reading (both reading fluency and reading comprehension) and mathematics. Using MLwiN 2.24 (Rasbash, Charlton, Browne, Hearly, & Cameron, 2011), we considered three levels: Pupil, class and school. To get the size of school and class effects, we simply estimated the empty model (see Tables 8.8, 8.9, and 8.10).

The analysis of the data from the end of 4th grade shows that 47% of the total variance in scores for writing and reading comprehension reflects variability in achievement between schools. In mathematics and reading fluency, however, the variance between schools represents 25% and 41% of the total variance, respectively. The achievement-related between-school variances are also significant at the end of the 5th grade, i.e. 34%, 51%, 40% and 28% of the scores obtained

Table 8.8 Empty model applied to the data of the end of 4th grade

	Writing	Reading fluency	Reading comprehension	Mathematics
Intercept	8.08*** (.90)	24.38*** (3.09)	3.29*** (.30)	11.61*** (.74)
Variance components				
Pupil level	47%	57%	45%	53%
Class level	6%	2%	8%	22%
School level	47%	41%	47%	25%

Note: ***p < .001

Table 8.9 Empty model applied to the data of the end of 5th grade

	Writing	Reading fluency	Reading comprehension	Mathematics
Intercept	12.91*** (1.09)	41.61*** (5.05)	3.19*** (.27)	12.14*** (.72)
Variance components				
Pupil level	57%	46%	56%	65%
Class level	9%	3%	4%	7%
School level	34%	51%	40%	28%

Note: ***p < .001

Table 8.10 Empty model applied to the data of the end of 6th grade

	Writing	Reading fluency	Reading comprehension	Mathematics
Intercept	14.75*** (1.44)	45.26*** (5.76)	3.62*** (.22)	16.22*** (.92)
Variance components				
Pupil level	50%	55%	69%	50%
Class level	6%	1%	8%	17%
School level	44%	44%	23%	33%

Note: ***$p < .001$

respectively in writing, reading fluency, reading comprehension and mathematics (see Table 8.9). The between-school variance component is also significant in the 6th grade, representing 44%, 44%, 23% and 33% of the total variance of the scores in writing, reading fluency, reading comprehension and mathematics (see Table 8.10).

The empty model results show that the variances between classes are systematically smaller than the between-school variances, ranging from 1% to 22% of the total variance. For writing and reading, the size of the class effect was always smaller than 10%.

Which variables explain the school and class effect? Until now, this issue has only been addressed in one study (Mokonzi, Van Damme, De Fraine, Gboisso, & Bela, 2017) which is still ongoing and focuses on the achievement in mathematics of pupils in 4th grade. After imputing the missing data[6], this study has revealed that, besides the individual variables (age, gender, initial knowledge in mathematics, language spoken at home, parent support and socio-economic status), which explain 40% of the between-class variance and 54% of the between-school variance, the considered class variables (composition effect in terms of the initial achievement in mathematics, class size, continued teacher training, teacher gender and experience, and learning opportunity for pupils) explain 30% of the residual variance between classes and 55% of the residual variance between schools (see Model 2 in Table 8.11). Only class composition in terms of prior mathematics achievement scores is a good predictor of achievement at the end of the same year ($\beta = 1.67$, $p < .001$). One can argue that the group composition effect could overlap with the effect of other class and teacher characteristics. When this composition variable is excluded (see Model 3), however, only the class size is (negatively) related to the achievement ($\beta = -0.71$, $p < .05$).

Similarly, the considered school-related variables (school size, continued training of the principal, principal's experience, school type) explain 54% of the residual variance between schools beyond the considered individual pupil characteristics and class-related composition effect (see Model 4 in Table 8.12). However, only the

[6]For the imputation of missing data, multiple imputation was applied using the R.3.05 software and the Amelia package.

Table 8.11 Effect of the class-related variables on mathematics of the end of the 4th grade

	Empty model (Null model)		Model 1 Individual variables		Model 2 Class variables (with effect of class composition)		Model 3 Class variables (without effect of class composition)	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
Fixed effects								
Intercept	12.59	0.51	12.58	0.42	12.54	0.58	11.94	0.66
Individual characteristics								
Pre-test mathematics			2.33***	0.10	2.23***	0.10	2.32	0.10
Age			−0.09	0.10	−0.08	0.10	−0.09	0.10
Gender (code: 0 = girls, 1 = boys)			0.45*	0.18	0.47*	0.18	0.45	0.18
Language (reference = French)								
Swahili			0.11	0.25	0.13	0.25	0.11	0.25
Lingala			−0.50*	0.25	−0.49*	0.25	−0.50	0.25
Other language			−1.08	0.80	−1.04	0.80	−1.10	0.80
Socio-economic status			0.00	0.08	−0.01	0.09	0.00	0.09
Parent support			−0.10	0.09	−0.10	0.09	−0.10	0.09
Repeaters			−0.30	0.17	−0.31	0.17	−0.30	0.17
Class composition								
Average pre-test mathematics					1.67***	0.30	–	–
Class size					−0.50	0.30	−0.71*	0.35
Characteristics teacher								
Gender (code: 0 = female, 1 = male)					0.14	0.54	0.43	0.62
Continued training					−0.08	0.58	0.94	0.64
Experience					0.06	0.28	0.32	0.32
Learning opportunity					0.43	0.28	0.58	0.33
Random effects								
Level 1 (pupil)	36.36	0.75	32.81	0.68	32.81	0.68	32.81	0.68

(continued)

Table 8.11 (continued)

	Empty model (Null model)		Model 1 Individual variables		Model 2 Class variables (with effect of class composition)		Model 3 Class variables (without effect of class composition)	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
Level 2 (class)	8.10	2.19	4.84	1.35	3.36	0.98	4.21	1.20
Level 3 (school)	6.89	2.86	3.12	1.56	1.40	0.97	2.75	1.38
−2*loglikelihood	30970.41		30441.81		30406.16		30431.91	

Note: *<.05; **<.01; ***<.001

Table 8.12 Effect of school-related variables

	Empty model (Null model)		Model 4 school variables (with effect of pre-test and class composition)		Model 5 school variables (without effect of pre-test and class composition)	
	Parameter	S.E.	Parameter	S.E.	Parameter	S.E.
Fixed effects						
Intercept	12.59	0.51	12.30	0.80	11.08	1.28
Individual characteristics						
Pre-test mathematics			2.23***	0.10		
Age			−0.09	0.10	−0.02	0.10
Gender (code: 0 = girls, 1 = boys)			0.49**	0.18	0.70**	0.19
Language (reference = French)						
Swahili			0.13	0.25	0.15	0.26
Lingala			−0.50*	0.25	−0.41	0.26
Other language			−1.03	0.80	−1.06	0.84
Socio-economic status			0.00	0.09	0.07	0.09
Parent support			−0.10	0.09	−0.15	0.09
Repeaters			−0.30	0.17	−0.59**	0.18
Class composition						
Average of mathematics pre-test			1.75***	0.29		
Class size			−0.09	0.32	−0.86	0.52
Teacher characteristics						
Gender (code: 0: f, 1: m)			0.10	0.55	0.67	0.79
Continued training			−0.15	0.57	1.78	0.82
Experience			0.17	0.28	0.28	0.58
Learning opportunities			0.21	0.28	0.62	0.45
School composition						
School size			−0.41	0.33	0.33	0.60
Characteristics principal						
Continued training			−0.14	0.76	−0.70	1.31
Experience			0.79*	0.29	0.28	0.58
Structural characteristics school						
Network (reference = non-conventioned public school)						
Conventioned public school			0.36	0.67	0.48	1.18
Private			0.76	1.06	1.75	1.80
Random effects						
Level 1 (pupil)	36.36	0.75	32.80	0.67	36.09	0.74
Level 2 (class)	8.10	2.19	3.61	1.03	6.89	1.89
Level 3 (school)	6.89	2.86	0.65	0.88	5.22	2.32
−2*loglikelihood	30970.41		30399.43		30920.60	

Note: *<.05; **<.01; ***<.001

principal's experience predicts pupil achievement in mathematics in 4th grade ($\beta = .79$, $p < .05$). We can argue that the pretest scores at the beginning of 4th grade and also the group composition based on these scores are influenced by the school. If we exclude these variables, none of the considered school characteristics is significantly related to the mathematics achievement (see Model 5).

8.10 Future Prospects

Our study showed that there are large differences in effectiveness between schools in DRC, as is usually the case in developing countries based on (limited) previous research. The reader will understand that we were rather disappointed because of the low number of class and school characteristics that significantly contributed to explaining the class and school effect on mathematics achievement in grade 4. Because we have yet to consider other variables, subject domains and grades, we consider it a bit early to focus on the few significant variables. Rather, we prefer to give some future perspectives as well as some additional observations and general reflections on doing educational effectiveness research in Central Africa.

As mentioned earlier, the 'Service de Planification et d'Évaluation en Éducation' of the University of Kisangani and the Center for Educational Effectiveness and Evaluation of KU Leuven (Belgium) have set up a database related to educational effectiveness in primary schools in the Oriental Province of the DRC. This database has allowed Masters and PhD students to carry out initial analyses which may be built upon in the future in order to identify the educational effectiveness factors of Congolese schools in general, and schools in the Orientale Province in particular. However, to optimize the exploitation of this database and improve the follow-up of the development of pupil achievement and the estimation of school effects, it will be important to calibrate instrument items before linking the different tests that have been successively used in the 4th, 5th and 6th grade of primary school.

Beyond the setting up of the database, this research -- which has now been extended to the secondary school level – also allows us to strengthen the skills of Congolese researchers within the field of educational effectiveness. This expertise could enable the set-up of larger research projects involving more provinces and schools in the future, and ultimately to fill the gap that the DRC has marked in this field.

The first research results emerging from this data analysis suggest the need for studying, in a later stage, both the direct and indirect relationships, and even the interrelationships, between the variables. More specifically, the use of structural equations may be appropriate.

To compare the effectiveness of Congolese schools with the effectiveness of schools in other countries, specifically Sub-Saharan countries, we intend to use in later research some of the instruments that have been used in comparative studies carried out in Africa (e.g. by the PASEC and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)) and internationally (e.g. PISA, TIMSS and PIRLS).

8.11 Reflections on Educational Effectiveness Research in Central Africa

The development of this study and the first results emerging from it give rise to some reflections on educational effectiveness research conducted in Central Africa in general, and in DRC in particular. These considerations concern more specifically the class and school characteristics, the teachers' achievement and knowledge, the selection of analytic models, the school sample, the researchers' training, and the funding of research related to educational effectiveness.

Indeed, while preparing this research in 2010, among other activities, we drew up an inventory of the research work on educational effectiveness realized between 1970 and 2010 at our 'Faculté de Psychologie et des Sciences de l'Education', visited some primary and secondary schools in Kinshasa and Kisangani, and organized a consultation workshop with school principals and inspectors. The visits to schools revealed that only a few of them meet the required national guidelines in terms of infrastructure (buildings, equipment, etc.) to guarantee good-quality education. In terms of material conditions, most Congolese schools are 'schools' in name only. As already indicated by the World Bank in 2005, some schools have partly damaged roofs or even no roof at all, as well as degraded walls. In practice, a lot of these schools cannot be used when it is raining or too hot; in those weather conditions, the pupils are sent back home (Banque Mondiale/World Bank, 2005).

Some of the schools organize two or more classes within one space separated by walls which are barely 150 cm high. A lot of schools lack the minimum resources and equipment required to accomplish high-quality education. It is not uncommon to encounter schools without benches, blackboards and teaching materials. In these schools textbooks tend to be the exception, rather than the rule. This implies that 'there are many pupils who leave primary and secondary school without ever having touched or read a book' (Mokonzi, 2009, p. 91).

Besides the lacking material conditions of schools and classes, we must also deplore the *insufficient quality of the teaching*. Lumeka (1985) compared teaching in the DRC to an unkept meadow or an unprotected pasture where anyone who feels like it, can have his cattle grazing. This comes as no surprise, considering the high rate of unemployment that afflicts a large number of Congolese graduates. They consider teaching as an 'in between job', a kind of 'waiting room' where they sit in expectation of something better to come (Mokonzi, 2009, p. 85).

Even the graduates having attended teacher-training institutions, such as the 'Humanités pédagogiques' and the 'Instituts supérieurs pédagogiques', where students are being trained to teach in primary and secondary school respectively, are not adequately equipped to provide high-quality education, as their own initial training is barely sufficient (UNESCO & IIEP 'Pôle de Dakar, 2014). On top of that, teacher professional development is a rarity in schools (Banque Mondiale, 2005).

The working conditions of Congolese teachers are poor. Indeed, a large percentage of teachers work without getting paid[7] and this for many years (Brandt, 2018). As Roller has pointed out so well in his foreword to Lumeka's book published in 1985 (p. 12), the simmering discontent, the muffled struggles, the weariness, and the abandonment of the teaching profession are perfectly understandable. In addition, teachers have to perform other duties and tasks beside their regular job just to survive, which leaves them at the end of the day with an enormous feeling of fatigue. The situation of Congolese teachers has not improved since the publication of Lumeka's book; on the contrary, it has even worsened. As Maroyi (2005) indicated, the Congolese teacher is stripped of his material prestige and has no time whatsoever to find an additional income elsewhere. So, he finds himself entirely demoralised and demotivated; his professional conscience is entirely undermined. In these conditions, we understand why young Congolese students, even the ones graduating from teacher-training institutions, are not longing to embrace a teaching career.[8]

This probably also has to do with poor initial and continued teacher training; teachers' demotivation with regard to the generally very poor achievements of their pupils; and the fact that the added value of attending school during 1 year often appears to be nil or even negative. This last fact has been demonstrated by exploratory analysis carried out on our data gathered in the fourth grade of primary school during the school year 2010–11. As shown in Table 8.13, the writing performance gains realized during the year are generally small, and within some categories of school type, even negative.

With regard to training and working conditions, the situation of the school principals is similar to that of teachers.

Apart from the general conditions related to the school, teachers and principals, we should also consider the study sample and the analytic model used in our research. Indeed, even if in our sample the conditions diverge by school, the differences between urban and rural schools are likely to be even more pronounced. In this respect, it would be interesting to also include rural schools in educational effectiveness research in Central Africa and the Democratic Republic of the Congo, which until now has always been restricted to urban schools.

As to the variables, the focus should be put on teacher-related characteristics, more particularly on their skills, their teaching practices and their motivation. Within the scope of an ongoing study, Maroyi (2015) tested teachers' skills by administering (to a sample of 202 4th grade teachers of primary schools in the city of Kisangani) a set of questions taken from textbooks used by these teachers to teach mathematics and French. So, this implies that they have been administered the same questions they pose to their pupils. In addition, Maroyi has asked these teachers to

[7]According to the teachers' syndicate, 3,00,000 teachers were not being paid by the Congolese government at the beginning of the school year 2014–2015. In a city such as Beni, only 52% of the public school teachers have been paid in 2017. There are teachers facing this situation for nearly 20 years now.

[8]We can't emphasize this enough: salary conditions have a significant influence on the attractiveness of the profession and the education quality.

Table 8.13 Writing test performances of pupils in 4th grade (maximum score = 37)

Network[(1)]		Beginning of the year	End of the year	Gain
Catholic	Average	6.52	7.55	1.03
	N	639	639	
Kimbanguist	Average	3.28	3.07	−0.21
	N	216	216	
Public (unconventioned)	Average	5.81	9.89	4.08
	N	336	336	
Private	Average	7.63	6.02	−1.61
	N	158	158	
Protestant	Average	4.58	8.31	3.73
	N	421	421	
Salvationist	Average	1.41	5.48	4.07
	N	54	54	
Total	Average	5.50	7.43	1.93
	N	1824	1824	

Note: The 'conventioned' public schools are split up in four religious subgroups

Table 8.14 Achievement of Grade 4 teachers in French language

Number	Lowest score	Average	Highest score	Maximum	SD	Q_1	Mdn	Q_3	Success rate
196	20	41.04	55	57	6.71	38	42	46	72%

Table 8.15 Mathematics achievement of Grade 4 teachers

Number	Lowest score	Average	Highest score	Maximum	SD	Q_1	Mdn	Q_3	Success rate
202	0	29.69	52	60	10.88	22	30	37	49%

describe the successive pedagogical steps of a mathematics and a French course. The results of Maroyi's study reveal that the teachers encounter difficulties in mastering the subjects they are teaching. None of the teachers involved in this study was able to correctly answer all of the 60 and 57 questions regarding mathematics and French, respectively. Worse still, the teacher with the lowest score in mathematics had all of the answers wrong, and the one with the lowest score in French only answered 20 out of 57 questions correctly (see Tables 8.14 and 8.15). In mathematics, 50% of the teachers did not succeed in obtaining a score of more than 50% (i.e. 30/60) and 75% have obtained scores lower than or equal to 37/60.

The results with regard to teachers' pedagogical content knowledge tests were also quite disastrous (see Tables 8.16 and 8.17). Out of 16 didactic questions about French classes, 75% of the teachers answered no more than 4 questions correctly. The same applies for questions asking them about their mathematics pedagogical content knowledge: Out of 10 questions, 75% of the teachers answered no more than 2 questions correctly. In other words, not only do teachers have an insufficient

Table 8.16 Pedagogical content knowledge of teachers in French language

Number	Lowest score	Average	Highest score	Maximum	SD	Q_1	Mdn	Q_3	Success rate
196	0	2.61	10	16	2.70	0	2	4	16%

Table 8.17 Mathematics pedagogical content knowledge of teachers

Number	Lowest score	Average	Highest score	Maximum	SD	Q_1	Mdn	Q_3	Success rate
202	0	1.46	9	10	2.05	0	0	2	15%

proficiency level with regard to the subjects they are teaching, they also lack the knowledge on how to teach mathematics and French. It is to be feared that the same may apply to the other subjects they teach.

But to highlight the class and school effectiveness factors, research should not only be focusing on teachers' skills and achievement, but also on their attitudes and motivation. In addition, the teaching materials and pedagogical resources such as textbooks for the pupils and methodology guides for the teachers, which are generally considered as significant predictors of pupil achievement in developing countries (PASEC, 2011; Van der Berg, 2008; Verspoor, 2006), have not been taken into consideration in the first year of this research.

The fact that educational effectiveness research in Central Africa has fallen far behind is partly due to the lack of training of researchers working within this domain within this region. In the case of DRC, this emerged very clearly from the seminar that the 'Service de Planification et d'Évaluation en Éducation' organized in 2013 for the researchers of the universities of Kinshasa and Kisangani. The training of researchers is therefore a crucial prerequisite to intensify educational effectiveness research in Central Africa.

The funding of this type of research by the governments of the countries of this sub-region is another important prerequisite. Unfortunately, this is not the case as of today.

References

ACER. (2015). *The Southern and Eastern Africa Consortium for Monitoring Educational Quality. Assessment GEMs n°.8*. Melbourne, Australia: ACER.

Banque Mondiale (World Bank). (2005). *Le système éducatif de la République Démocratique du Congo: priorités et alternatives* [The educational system of the Democratic Republic of the Congo: Priorities and alternatives]. Washington, DC: Banque Mondiale, Département du Développement Humain, Région Afrique.

Brandt, C. O. (2018). *Illegibility as a state effect. The Limits of governing teacher identification in the Democratic Republic of Congo* (Unpublished doctoral dissertation). Amsterdam, The Netherlands: University of Amsterdam.

Carrim, N., & Shalem, Y. (1999). School effectiveness in South Africa. *International Journal of Qualitative Studies in Education, 12*(1), 59–83. https://doi.org/10.1080/095183999236330

Cellule Technique pour les Statistiques de l'Éducation. (2008). *Annuaire statistique de l'enseignement primaire, secondaire et professionnel. Année scolaire 2006–2007* [Statistical yearbook of primary, secondary and vocational education. Schoolyear 2006–2007]. Kinshasa, Democratic Republic of the Congo: Ministère de l'EPSP.

Cellule Technique pour les Statistiques de l'Éducation. (2014). *Annuaire statistique de l'enseignement primaire, secondaire et professionnel. Année scolaire 2012–2013* [Statistical yearbook of primary, secondary and vocational education. Schoolyear 2012–2013]. Kinshasa, Democratic Republic of the Congo: Ministère de l'EPSP.

Cellule Technique pour les Statistiques de l'Éducation. (2017). *Annuaire statistique de l'enseignement primaire, secondaire et professionnel. Année scolaire 2014–2015* [Statistical yearbook of primary, secondary and vocational education. Schoolyear 2014–2015]. Kinshasa, Democratic Republic of the Congo: Ministère de l'EPSP.

Coleman, J. S., Campbell, E. Q, Hobson, C. F, McPartland, J., Mood, A. M, Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Conn, K. (2014). *Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluation* (Unpublished doctoral dissertation). New York, NY: Columbia University.

Creemers, B. P. M. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Abingdon, UK: Routledge.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in education effectiveness research*. London, UK: Routledge.

Christie, P. (1998). Schools as (dis)organisations: The breakdown of the culture of learning and teaching in south African schools. *Cambridge Journal of Education, 28*(3), 283–300. https://doi.org/10.1080/0305764980280303

Cronbach, J. L. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684. https://doi.org/10.1037/h0043943

De Herdt, T., Marivoet, W., & Muhigirwa, F. (2015). *Analyse de la situation des enfants et des femmes en RDC: vers la réalisation du droit à une éducation de qualité pour tous* [Analysis of the situation of children and women in the Democratic Republic of the Congo: Towards the realisation of the right to quality education for all]. Rapport final. Kinshasa, Democratic Republic of the Congo: UNICEF.

Dierendonck, C. (2008). Validation psychométrique d'un questionnaire francophone de description de soi adapté aux préadolescents [Psychometric validation of a French self-description questionnaire adapted to preadolescents]. *Mesure et évaluation en éducation, 31*(1), 51–91. https://doi.org/10.7202/1025013ar

Fleisch, B. (2007). History of the school effectiveness and improvement movement in Africa. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 341–350). Dordrecht, The Netherlands: Springer.

Gage, N. L. (1963). Paradigms for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 94–142). Chicago, IL: Rand McNally & Company.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal, 113*, F64–F98. https://doi.org/10.1111/1468-0297.00099

Harber, C., & Muthukrishna, N. (2000). School effectiveness and school improvement in context: The case of South Africa. *School Effectiveness and School Improvement, 11*(3), 421–434. https://doi.org/10.1076/sesi.11.4.421.3559

Howie, S. J. (2006). Multi-level factors affecting the performance of south African pupils in mathematics. In S. J. Howie & T. Plomp (Eds.), *Contexts of learning mathematics and science: Lessons learned from TIMSS* (pp. 157–174). London, UK: Routledge.

Hungi, N. (2011). *Accounting for variations in the quality of primary school education* (Working paper 7). Paris, France: Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., van Capelle, F., et al. (2010). *SACMEQ III Project results: Pupil achievement levels in reading and mathematics*. Paris, France: Southern and Eastern Africa Consortium for Monitoring Educational Quality. https://doi.org/10.13140/RG.2.2.27312.35841

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114*(2), 497–532. https://doi.org/10.1162/003355399556052

Lumeka, L. Y. (1985). *Auto-perception des enseignants du Zaïre. Contribution à la socio-psychologie professionnelle des enseignants dans les pays en développement.* [Self-perception of the teachers of Zaïre. Contribution to the professional socio-psychology of teachers in developing countries]. Kinshasa, Democratic Republic of the Congo: Editions et Culture Africaines.

Maroyi, S. (2005). La décadence du métier d'enseignant en République Démocratique du Congo [The decline of the teaching profession in the Democratic Republic of the Congo]. *Ecole Démocratique, hors série*, 45–48.

Maroyi, S. (2015). *Les compétences des enseignants de quatrième année primaire dans les écoles de la ville de Kisangani* [The competencies of teachers of the fourth grade in the schools of Kisangani]. Mémoire de Master inédit, Faculté de Psychologie et des Sciences de l'Éducation, Université de Kisangani.

Marsh, H. W. (1988). *Self-description questionnaire: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph*. San Antonio, TX: Psychological Corporation.

Ministère de l'Enseignement Primaire, Secondaire et Initiation à la Nouvelle Citoyenneté, Ministère de l'Enseignement Technique et Professionnel, Ministère de l'Enseignement Supérieur et Universitaire, & Ministère des Affaires Sociales, Action Humanitaire et Solidarité Nationale. (2015). *Stratégie sectorielle de l'éducation et de la formation 2016–2025* [Sectoral strategy of education and of the training 2016–2025]. Kinshasa. Retrieved from: https://www.globalpartnership.org/fr/download/file/fid/52992

Ministère de l'Enseignement Primaire, Secondaire et Professionnel. (2012). *Enquête sur la situation des enfants et adolescents en dehors de l'école* [Investigation into the situation of children and adolescents outside of school]. Kinshasa, Democratic Republic of the Congo: UNICEF-DFID-UNESCO.

Ministère du Plan. (2011). *Document de la stratégie de croissance et de réduction de la pauvreté de seconde génération* [Document on the growth strategy and reduction of poverty of the second generation]. Kinshasa, Democratic Republic of the Congo: Ministère du Plan.

Mokonzi, G. B. (2009). *De l'école de la médiocrité à l'école de l'excellence au Congo-Kinshasa* [From the school of mediocracy towards the school of excellence in Congo-Kinshasa]. Paris, France: l'Harmattan.

Mokonzi, G. B., & Issoy, A. (2002). Étude de la contribution de l'enseignement primaire à l'alphabétisme [Study of the contribution of primary education to literacy]. *Scientia Paedagogica Experimentalis, 39*(2), 181–195.

Mokonzi, G.B., Van Damme, J, De Fraine, B., Gboisso, O.A., Bela, J.P. (2017). *Effet des classes et des écoles sur les performances en mathématiques des élèves de quatrième année primaire de la Province Orientale en République Démocratique du Congo* [Effects of classes and schools on math achievement of pupils of fourth grade in the 'Province Orientale' in the Democratic Republic of the Congo]. Unpublished text, University of Kisangani, Democratic Republic of the Congo.

Palardy, G., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111–140. https://doi.org/10.3102/0162373708317680

PASEC. (2011). *L'enseignement primaire en République Démocratique du Congo. Quels leviers pour l'amélioration du rendement du système éducatif?* [Primary education in the Democratic Republic of the Congo. Which levers for the improvement of the output of the educational system?]. Dakar, Senegal: PASEC, CONFEMEN.

PASEC. (2015). *PASEC2014 – Performances des systèmes éducatifs en Afrique subsaharienne francophone: Compétences et facteurs de réussite au primaire* [PASEC2014 – Improvement of the education system in french subsaharian Africa: Success competencies and factors at the primary school level]. Dakar, Senegal: PASEC, CONFEMEN.

Présidence de la République. (2014). Loi-cadre n° 14/004 du 11 février 2014 de l'enseignement national [General Law nr. 14/004 of February 11th 2014 of the national education]. Journal officiel, numéro spécial.

Rasbash, J., Charlton, C., Browne, W. J., Hearly, M., & Cameron, B. (2011). *MLwiN version 2.24*. Bristol, UK: University of Bristol, Center for Multilevel Modelling.

Sammons, P., Sylva, K., Melhuish, E., Blatchford, I. S., Taggart, B., Barreau, S. and Grabbe, Y. (2008). *Effective pre-school and primary education 3–11 project (EPPE): The influence of school and teacher quality on children's progress in primary schools* [Research Report No DCSF-RR028]. London, UK: DCSF/Institute of Education, University of London.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61–80. https://doi.org/10.1080/0924345900010106

Scheerens, J. (1992). *Effective schooling: Research, theory and practice*. London, UK: Cassell.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.

Teddlie, C., & Reynolds, D. (Eds.). (2000). *The international handbook of school effectiveness research*. Abingdon, UK: Routledge.

UNESCO, & IIEP Pôle de Dakar. (2014). *République Démocratique du Congo: Rapport d'état du système éducatif. Pour une éducation au service de la croissance et de la paix* [Democratic Republic of the Congo: Report on the state of the educational system. Towards an education in the service of growth and peace]. Dakar, Senegal: UNICEF, UNESCO-IIEP Pôle de Dakar.

Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement, 13*(4), 383–397. https://doi.org/10.1076/sesi.13.4.383.10285

Van Damme, J., Opdenakker, M.C., Van Landeghem, G., De Fraine, B., Pustjens, H., & Van de gaer, E. (2009). Fondements et principaux résultats de recherche sur l'efficacité dans l'enseignement [Foundations and main results of the research on educational effectiveness]. In X. Dumay & V. Dupriez (Eds.), *L'efficacité dans l'enseignement. Promesses et zones d'ombre* [Effectiveness in education. Promises and gray areas] (pp. 19–34). Bruxelles, Belgium: De Boeck.

Van der Berg, S. (2008). Discussion papers. Retrieved April 30, 2014, from Center for European, Governance and Economic Development Research: http://wwwuser.gwdg.de/~cege/Diskussionspapiere/69.

Verspoor, A. (2006). *Conditions and factors of effective schools in Africa. Effective schools for Sub Saharan Africa*. Libreville, Gabon: Association for Development in Africa.

Chapter 9
School and Teacher Value Added Performance and the Relationship with Teacher Professional Development in Mainland China

Sally M. Thomas

9.1 Introduction

Mainland China is a huge country with a population of over 1.3 billion and comprises the largest state-run education system in the world, involving at the secondary level for example, over 76 thousand junior, senior and vocational secondary schools. The country is also very diverse both economically and geographically, demonstrated by the doubling of 2014 household average per capita income when comparing the poorest to the richest provinces (OECD, 2016). Moreover, expenditure of Chinese households on education has tripled over the past decade (Statistica, 2018) and this may well have contributed to increasing inequity in government and household funding spent on education, which varies widely between regions (China Daily, 2018; OECD, 2016). In this context of great disparity, the equity and effectiveness of education across urban and rural areas and Eastern and Western provinces in China is a hot topic for government policy makers as well as students, parents, teachers and the broader public. The 1986 Compulsory Education Law of China specified 9 years of government funded compulsory school attendance, including 6 years of primary school and 3 years of junior high school. However, it was not until the late 1990s that the Chinese government began to actively address educational inequality in order to counteract some of the consequences of earlier policies which focused resources on the best (key) schools (Feng, 2007). For example, in 1998 the government introduced changes in school funding, governance, enrolment policy, and teacher distribution and development to better support disadvantaged schools (Feng, 2007). Moreover, the 2001 curriculum reform

The findings presented are from a UK DFID/ESRC funded project "Improving Teacher Development and Educational Quality in China", Grant reference: RES-167-25-0428.

S. M. Thomas (✉)
School of Education, University of Bristol, Bristol, UK
e-mail: S.Thomas@Bristol.ac.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_9

directing schools to foster and assess students' all round development ("suzhi jiaoyu") refocused "quality education" on a broad range of student outcomes including not just academic examination results but also engagement, attitudes, critical thinking, problem solving and creativity, and encouraged more student-centred teaching methods (Ryan, Kang, Mitchell, & Erickson, 2009) thereby signalling a fundamental shift in approach. Essentially this key reform sought to replace the centuries-old system and culture based on selection by examination to identify an administrative elite to one designed to meet the needs of all students and the challenges of globalization. Given the radical nature of this reform the Ministry of Education found it necessary to formally reiterate the purpose of this broader concept of "quality education" 1 year later:

> . . .to enhance the national quality of students and this fundamental purpose is underpinned by national education policy based on the 'Education Law'. That is to focus on students' education and their long-term social development, to require comprehensive improvement to the basic quality of all students, to pay attention to train students' attitudes and abilities and to promote their moral, intellectual and physical development in a vivid, lively and active way. (CMOE, 2002)

The "Education Law" was updated in 2006, and equity and inclusion were emphasised as key priorities with requirements to reduce academic selection and provide more effective schooling for students with special educational needs, delinquent adolescents and poor families (CMOE, 2006). In 2010 the 17th National Congress reviewed progress and produced the Outline of China's National Plan for Education Reform and Development 2010–2020. This document concluded that although much had been achieved, "our education systems and mechanisms are not yet perfect", in particular curriculum content and teaching methodology are "relatively outdated", "school work burdens" on younger students are too heavy, and the development of "quality education" is not proceeding fast enough. In addition on-going issues of equity were highlighted, linked to the continuing unbalanced development of education provision in urban and rural areas, and in poorer and ethnically autonomous areas. Among the five executive principles outlined for 2010–2020 were requirements to "treat reform and innovation as a mighty driving force for education development", and to consider "quality improvement as a core task". To achieve these aims the 2010–2020 plan stated it is essential to "renovate school operation and education administration, overhaul quality evaluation, [and] examinations. . . revamp teaching . . . methods and approaches, and put a modern school system in place" (CMOE, 2010).

Thus, in the context of ongoing reforms in China to improve education quality and equity, and given the focus of this book, it is relevant at this point to review what researchers have argued work well in the Chinese education system before moving on to discuss more critical findings and present new evidence on the range and extent of educational effectiveness in mainland China. In terms of "what works", several researchers have noted the relative lack of empirical research on this topic (Reynolds, Caldwell, et al., 2016; Sun, Zheng, & Wang, 2014; Thomas, Kyriakides, & Townsend, 2016) especially given the enormous size and diversity of the country. A critical issue of available published evidence relates to the representativeness of the

research locations and samples employed, with many key studies only reporting findingss from major cities such as Shanghai, Beijing or more socio-economically developed areas in the East where typically students achieve the best examination outcomes. Most obviously this includes OCED PISA results, which only report the very high achievement of Shanghai students (Liang, Kidwai, & Zhang, 2016) but also other key studies such as that of Cheng and Wong (1996). As a consequence much of the discussion around "what works" in Chinese education is arguably only relevant to those schools serving more advantaged communities or high attaining selected groups. With this in mind, as well as the overwhelming influence of traditional Confucian culture that highly values examination success, it is not so surprising that current debates and evidence of "what works" in China paint a typically rosy picture of high academic achievement linked to established school effectiveness levers such as high expectations, time on task, opportunity to learn, whole class teaching, and regular homework (Reynolds, Caldwell, et al., 2016; Reynolds, Chapman, et al., 2016). Autonomy in leadership and management and a strong focus on teachers' professional development via the established system of teacher research groups are also claimed to be features that promote effective schooling in China (Cheng & Wong, 1996; Tan, 2013). However, the reality of school effectiveness across the full range and diversity of schools in China requires much further investigation and critical analysis, which is underlined by the Chinese government's own concerns about promoting school quality for all and the continuing focus on the need for educational reform (Zhou, 2017). For example, Feng (2007) has identified a number of expected but also new contributory factors that support improved outcomes, specifically in disadvantaged schools, including shared values and beliefs, research-based leadership decisions and policy, appropriate expectations and suitable pedagogy that creates experiences of success for all students, school based problem-centered teacher development and making full use of external opportunities and government initiatives for school development support. Liu (2017) has also pointed to the key role of external partnerships and collaboration between disadvantaged and key schools in China to support turnaround school improvement efforts. Other researchers have emphasised the important role of creating and sustaining schools as professional learning communities to enhance teachers practice and the quality of teaching and learning, which in turn is likely to improve student outcomes (Qiao, Yu, & Zhang, 2017; Thomas, Peng, & Triggs, 2017; Zhang & Pang, 2016; Zhang, Yuan, & Yu, 2017).

Regarding the wider context of 'what works' in education, Reynolds and colleagues (2016) have summarised the international evidence and identified the following key priorities and approaches to "optimise educational improvement" (p. 439). First, the need for a renewed focus on professional development to provide better links between initial pre- and in-service teacher training and to challenge established cultural norms that inhibit teacher learning and improved practice. Second, the need for more ownership and commitment to rigorous localised evidence-based research in order to establish a stronger contextualised evidence base of 'what works', which takes account of local priorities and models of practice in schools. Third, more joined-up public service provision which harnesses the combined efforts of within-, between-, and beyond-school improvement in order

to challenge and better address inequality in student outcomes. All three of these critical aspects are highly pertinent in the China mainland context to future educational improvement. However, interestingly in China what also does not seem to work so well is having rigorous systems of educational evaluation and assessment in place to support school improvement. Improving this aspect is seen as crucial by the Chinese government and many commentators, given current evaluation and assessment approaches are considered underdeveloped and unscientific (CMOE, 2013). Indeed, evaluation is fundamental to successful reform: there must be valid standards and frameworks, developed at local level, to evaluate and guarantee quality. Thus, in the next section a brief overview and critique of current evaluation and supervision systems is presented and this is followed by a detailed example of new research on this topic funded by DFID ESRC: "Improving Educational Evaluation and Quality In China" and "Improving Teacher Development and Educational Quality in China" projects (IEEQC, 2018).

9.2 Systems for Supervision, Evaluation and Inspection

To some extent the way educational quality is currently evaluated in China has been influenced by historical priorities which focused more on improving other key aspects of the system, notably curriculum reform, while continuing traditional supervision methods for the implementation of education legislation and the delivery of the service. However, this situation is changing rapidly given the Chinese Ministry of Education (CMOE) has recognised the need for improved evaluation and assessment methods to support reform efforts (CMOE, 2010). Overall monitoring of education provision and quality is provided by offices such as the National Institute for Education Sciences (NIES) and research centres located in universities, which are funded centrally by the CMOE, but much of the information they collect is not publicly available. Meanwhile, responsibility for delivery of education and quality assurance is essentially handed down from the CMOE to different levels of local government, and overseen centrally by CMOE departments such as the National Office of Educational Supervision. Schools are the end product of a complex and hierarchical system. The reality of planning, delivering and managing the work of education involves, in different ways, government at the level of the province, the city, the county, the district, the township, and the village. Evaluative activity is primarily grounded in law via a legal requirement for "supervision", at the appropriate level, of educational institutions, and this activity is incorporated into the systems that deliver education. The *Guidelines for Primary and Secondary Schools Supervision and Evaluation* (CMOE, 1991a, 1991b, CMOE, 1997) provide a framework and all provinces, 99% of the cities and almost all of the 2862 county-level divisions have established education supervision offices with a total of 46,245 full-time and part-time supervisors (Yang & Xiong, 2012). Thus, it is important to point out that school evaluation is essentially carried out by the same authorities that

provide education and so these administrative systems are not independent, potentially leading to conflicts of interest.

In brief, the legal liabilities for supervision are restricted to provision: construction and maintenance of buildings, financial management, factors that might impede school enrolment, ensuring compliance with specific rules and regulations. Local offices must report upwards on educational building work, budgets, financial accounts of educational expenditure and schools must submit to supervision (CMOE, 1995 – Education Law, Article 16). However, over the last 25 years, directives from the Ministry's Office of National Educational Supervision have expanded the range of tasks included in educational supervision beyond monitoring the fulfilment of legal and other duties. Supervision for legal compliance of easily quantifiable indicators is still strong and fits comfortably with the traditional focus on inspection for compliance. However, the 2001 and subsequent reforms in curriculum and assessment emphasising students' all-round development have raised new and ongoing challenges. Monitoring "quality-oriented education" and the improvement of school quality and effectiveness requires an approach aligned more closely with concepts of evaluation. Thus the approach to "evaluation" required in the twenty-first century has had to move beyond ensuring compliance and using students' performance in national examinations as a benchmark for effectiveness. It involves both assessments of the student in relation to all round development ("suzhi jiaoyu") and of the effectiveness of the school in delivering what the reforms require, of which student achievement is only a part. Administrators at provincial, city and county levels were and still remain faced with the challenge of working out how to discharge their supervision and evaluation responsibilities with respect to this new approach.

The radical shift in evaluation of student outcomes from an exclusive concern with examination results to a broader interest in all-round development creates tensions and dilemmas for schools and supervisors. The Gaokao universal Entrance Exam to Higher Education (EEHE), instituted in 1952, abolished in the Cultural Revolution and re-instated in 1977 as one of the first education-related acts of the then-new Deng Xiaoping administration, is firmly established as a significant marker in the creation of a meritocracy. Its significance is great for aspiring parents and students, and universities, and therefore for schools. In spite of this tension new approaches to more comprehensive individual student assessment have emerged, for example, the establishment of a standard student portfolio or Growth Record Report (CMOE, 2002). Use of portfolios is encouraged in schools with the aim of recording different dimensions of all round development including: moral quality, civic responsibility, learning ability, co-operation and interpersonal ability, sports and health, aesthetic and performance. The evidence includes quantitative assessments such as test scores and examination results and these sit alongside examples contributed by the student, which could include: creative work, their accounts of participation in activities or events, and volunteer work. All this is combined with comments from teachers, parents and peers to provide a basis for more qualitative or formative assessment. However, some researchers such as Wu (2013) argue there is little support for portfolios in practice. The CMOE has also recognised that the

system of school supervision and inspection needs to be more rigorous and systematic in order to support school improvement. Guidelines for a new framework for evaluating schools and the kinds of indicators Provincial and local inspection systems should use have now been developed (Chu & Cravens, 2012; CMOE, 2011) but this work is just beginning and indicates the ongoing need for further research and development in evaluating educational effectiveness in China. The Ministry has also acknowledged that "the tendency to evaluate education quality based simply on student test scores has not been fundamentally changed" (CMOE, 2013, cited in Zhao, 2013). Schools and local authorities are constrained by the demands of parents who continue to place high priority on university entrance examination results as indicators of school effectiveness. Moreover, researchers such as Zhang and Minxia (2006) argue that evaluation using "Quality" related outcomes, involving assessment of non-cognitive factors such as moral qualities, ideological awareness, and mental health was challenging and raised concerns about the validity and reliability of judgments based on teacher and peer assessment. These issues are ongoing and recognised as having a negative impact on the development of the "whole student". Nevertheless, the new framework draws on a model developed in Shanghai named "Green Indicators" (Liang et al., 2016) and research conducted on school supervision and evaluation in the context of the quality-oriented basic education by the CMOE National Assessment Center of Education Quality in Beijing Normal University (Chu & Cravens, 2012). The framework comprises five areas for student evaluation: academic development (knowledge, skills, creativity, thinking, application); moral development (behaviour, character, citizenship, ambition, beliefs); psychological and physical health; academic burdens and motivation (development of interest and unique talents). There are also indicators related to school processes including instruction and curriculum, teacher professional community, leadership, and management. Moreover, from 2015 to 2017, the CMOE National Assessment Center of Education Quality organized and implemented the first cycle of national compulsory education quality monitoring using a sample of fourth-grade and eighth-grade students in 31 provinces, to assess Chinese, mathematics, science, physical education, art, and moral education (Liang et al., 2016; Xin, 2017).

The new approaches being developed in China to evaluate schools and support school improvement are commendable, but interestingly do not include the key measure of school effectiveness, that is a measure of the relative progress of students at different time points, typically referred to as "value added" assessment in educational effectiveness research. This is problematic because without an a "value added" approach raw measures of students' academic outcomes and entrance levels to higher education are likely to continue to be viewed as the key indicators of school quality. As a result schools with disadvantaged intakes tend to be judged unfairly, while complacency is possible amongst schools with more able students, and it is difficult to identify best school practice. In contrast, value added measures provide an alternative, fairer approach to evaluating school performance than the "raw" examination results by adjusting for students' previous attainment and other relevant factors outside the control of the school, to estimate their progress in comparison to

students in other schools. The concept of value added is, therefore, both an indicator of a school's effectiveness and a tool for head teachers and their staff to use to analyse the extent to which they have effectively raised student achievement. Importantly, the CMOE has recognised the need to evaluate student progress (CMOE, 2013) but clearly there is a difficulty in developing new policy in this area given the scarcity of rigorous empirical research on the nature of school effectiveness in China on which to test out new "value added" evaluation methods. Given this lack of evidence, the remaining parts of this chapter focus on providing a detailed relevant example of school and teacher effectiveness research conducted in China as a way of demonstrating the potential of new approaches to school evaluation, as well as providing original estimates on the size and consistency of school and teacher effects in mainland China. This is especially important because international EER indicates that school and teacher effects can differ substantively across countries and regions (Sammons, Hillman, & Mortimore, 1995; Thomas, 2001; Zhang, 2016), in part reflecting the social, economic, cultural and political factors in different countries, and therefore new local studies are needed to provide more evidence and allow better consideration of context specificity when interpreting estimates of school effects.

9.3 School Effectiveness Research in China

Value added methods and empirical studies of school effectiveness have only rarely been reported in mainland China. Of the few limited studies conducted to date it has been suggested "that between 15% and 39% of the variance in students raw Higher Education Entrance Examination (EEHE) outcomes are attributable to differences between schools. Moreover, having taken account of student intake and background factors, the equivalent adjusted figures are 9–33%" (Thomas et al., 2016, p. 225). Zhang (2016) has also reviewed school effectiveness findings in China and reported a three-level multilevel modelling analysis conducted by Ding and Xue (2009) which suggests the percentages due to schools and classes were around 25% and 20% respectively in terms of raw EEHE total score. However, current evidence from Chinese school effectiveness research (SER) studies tends to be small-scale and in some cases the quality of the assessment measures used is unclear, highlighting a critical need for larger-scale empirical research in a vast, diverse landscape such as China.

In relation to factors that might explain observed differences in school effects, research has demonstrated that teacher quality and opportunities for teachers' professional development and learning play a crucial role in enhancing student attainment and progress (Bolam et al., 2005). Similarly, school improvement research has consistently reinforced the importance of the classroom level to enhance student performance (Hopkins, Ainscow, West, Harris, & Beresford, 1997) and research studies in a range of international contexts such as the USA and UK have also emphasised the value and role of professional learning communities (PLCs) to

enhance professional practice (Bolam et al., 2005; Stoll & Louis, 2007). These findings suggest a close relationship between school effectiveness and teacher professional development as well as teacher quality. However, in China research indicates that for many teachers in China professional learning is underdeveloped, inequitable and inconsistent (Peng et al., 2013; Robinson & Yi, 2008; Zhao, Zhou, & Zhu, 2009), and little empirical research exists on professional learning, although some researchers have argued that PLCs typically exist as part of the formal structure of the education system, largely in the form of school teaching and research groups (Qiao et al., 2017; Teddlie & Liu, 2008; Wang, 2015).

From this starting point, two linked Economic and Social Research Council (ESRC) and UK Department for International Development (UKAID) funded projects, "Improving Teacher Development and Educational Quality in China"(ITDEQC) and "Improving Educational Evaluation and Quality In China"(IEEQC) conducted by UK University of Bristol in collaboration with the Chinese National Institute for Education Sciences Beijing, sought to provide new evidence on school and teacher effectiveness in China, as well as the role of teachers professional development in explaining differences in school effects. The research aimed to extend current knowledge in China about the impact of student characteristics, school contextual factors and teacher professional development factors on students' attainment and progress at school by providing robust new evidence, in an area where empirical data is lacking, to support educational policy development.

9.4 ITDEQC and IEEQC Project Methodology

The research collected and analysed both quantitative and qualitative data from senior high schools in three district (local) education authorities (LEAs) across western and eastern China. Quantitative data were collected from four consecutive student cohorts that comprised students' 2009–2012 Entrance Examination to Higher Education (EEHE) scores matched to their 2006–2009 prior attainment scores (entrance exam to senior high school [EESHS] in Maths, Chinese and English) as well as student and school survey information. The student sample comprised 303,345 students in 134 senior secondary schools and data records included questionnaire items concerning student background and attitudes to school. The school questionnaires included items about school context, processes, culture, teacher development and attitudes to teaching and learning. Additionally, for the 2012 student cohort this data was also matched to teacher survey data collected in 2012 from 17,000+ teachers in the same schools. The teacher questionnaire was specifically designed to include questions about the existence and relevance of professional learning communities, as well as key items drawn from the OECD 2008 TALIS survey of teacher development in order to provide equivalent data. Although not the main focus of this chapter, qualitative data were also collected via interviews and focus groups conducted with key stakeholders including head teachers, teachers, students, national and local policy makers. The research using

the qualitative evidence is published elsewhere (see Peng et al., 2013; Thomas et al., 2017) but in brief, this sought to explore key themes and features of teachers' learning, approach to evaluation and professional development as well as professional learning communities in Chinese schools. Specifically, this chapter presents selected findings from the projects and focuses on analysing the quantitative data to provide estimates of school and teacher effects, and to explore time trends in schools' value added performance and the links between "value added" measures of school effectiveness and teachers' professional development. This was done by employing a variety of different multilevel models to estimate of the range and extent of school and teacher effects in three regions (East/West) over four cohorts, and also contrasting these findings against the extent of teacher development and the key features of PLCs in Chinese senior secondary schools. Various multilevel models were constructed, controlling for different sets of variables including (i) student prior attainment; (ii) student characteristics (e.g., gender, hukou[1], parental education, parental occupation, home possessions); (iii) school context variables (e.g. ratio of major arts and sciences, percent of mothers as migrant workers, school mean prior attainment); (iv) school input and process variables (e.g. ratio of students to teachers, relevant school culture and teaching and learning attitude factors); (v) student effort and attitude variables (e.g. time spent in self-study, time spent on the Internet, relevant teacher and learning factors); and (vi) teacher professional development variables (e.g. continuing professional development (CPD) time, CPD type, CPD via collaboration and feedback factor). See Thomas et al. (2012) for further details of variables.

9.5 Findings

9.5.1 Significant Differences in School Effectiveness

Extending previously-reported findings for a single year 2009 (Thomas et al., 2012; Thomas et al., 2016), the findings presented here indicate that in mainland China there are substantial and statistically significant differences between the estimates of senior secondary schools' value added effectiveness, and that these differences vary not only across regions and subject outcomes but also over the four student cohorts examined. It was found that, across four cohorts, three LEAs and four different subject outcomes (Total, Chinese, English and Maths EEHE scores), between 12% and 27% of the total variance in student raw scores was attributable to difference between schools (see Table 9.1 which illustrates the detailed results for LEA1 and LEA3). Moreover, the apparent performance of senior secondary schools changed significantly when comparing raw and contextualised value added measures. After

[1] Hukou refers to a system of household registration in China that records an individual's urban/rural (agricultural/non-agricultural) residency and place of origin.

Table 9.1 Comparison of raw and value added model results for four EEHE student outcome measures

(2-level individual year model)	LEA1 (East) (2009–2012)	LEA3 (West) (2009–2012)
Total EEHE score		
Intra-school correlation – Unadjusted (raw) model	22–26%	22–27%
Intra-school correlation – Adjusted (contextualized value added) model	4–7%	4–14%
School variance explained by adjusted (contextualized value added) model	88–93%	71–89%
Total variance explained by adjusted (contextualized value added) model	58–68%	34–48%
Chinese EEHE score		
Intra-school correlation – Unadjusted (raw) model	12–19%	17–26%
Intra-school correlation – Adjusted (contextualized value added) model	2–5%	8–22%
School variance explained by adjusted (contextualized value added) model	81–94%	31–66%
Total variance explained by adjusted (contextualized value added) model	39–50%	18–28%
English EEHE score		
Intra-school correlation – Unadjusted (raw) model	20–24%	23–27%
Intra-school correlation – Adjusted (contextualized value added) model	4–7%	3–12%
School variance explained by adjusted (contextualized value added) model	91–96%	72–94%
Total variance explained by adjusted (contextualized value added) model	66–77%	35–52%
Maths EEHE score		
Intra-school correlation – Unadjusted (raw) model	19–23%	19–24%
Intra-school correlation – Adjusted (contextualized value added) model	4–6%	2–10%
School variance explained by adjusted (contextualized value added) model	84–93%	71–93%
Total variance explained by adjusted (contextualized value added) model	45–61%	29–40%

Figures illustrate range of estimates across 4 cohorts from two LEAs with complete data
Note: Contextualised value added model includes controls for student variables (E*ESHS* prior attainment (English, Chinese, Maths), gender, age, major (science/arts), tuition fee, hukou, boarding status, previous school type, parents education, parents occupation, home possessions (e.g. number of books)) and context variables (% Hukou (city/town), % arts major, % fathers' education (college/university or above), School mean EESHS English prior attainment, School mean EESHS Math prior attainment, School standard deviation EESHS Total score)

controlling for student prior attainment on entry to senior secondary school and other student characteristics and school context[2] factors outside the control of the school, 18–77% of the total variance, and 31–96% of the school variance, in students' EEHE scores was explained (see Table 9.1). Subsequently, of the remaining total variance, 2–22% was still attributable to differences between schools, thereby demonstrating a school effect, particularly in western China (LEA3) where typically the largest school effects were observed (Thomas & Peng, 2011b). These findings indicate a better fit of the data and more robust estimates, in comparison to equivalent results from previous studies in China reported above. It seems that in China school effects in different subjects and regions range from similar to slightly larger than equivalent results in UK (see Munoz-Chereau & Thomas, 2016; Thomas, 2001; Thomas, Peng, & Gray, 2007).

Interestingly, the impact of school context employed in contextualised value added model varied considerably between the three LEAs, outcomes and years; explaining between 2% and 60% of school variance on top of that explained by individual student factors. This suggests a greater influence of context factors on school performance in some regions, particularly those in more rural disadvantaged areas (e.g. LEA3). Further details of the impact of a variety of student intake, and contextual factors on student and school performance are reported elsewhere (see Thomas, Massoud, Munoz-Chereau, & Peng, 2012).

9.5.2 Time Trends in School Performance

Another key research question concerned time trends in school performance. School and year effects have been estimated over 4 years using a three-level (school/year/student) model. Table 9.2 illustrates the percentages of variance in students' EEHE scores for raw unadjusted and value added measures that are attributable to differences between schools and to differences between years in LEA1 (East) and LEA3 (West). For the adjusted contextualised value added model, having taken into account the observed random variation over time (estimated at around 2–4% in the two LEAs), it appears that the percentages of variance in students' EEHE scores attributable to differences between schools reduced to 4–7% (LEA1) and 7–16% (LEA3). These estimates provide a more robust summary of the school effect in LEA1 and LEA3 respectively over 4 years. Overall, the findings demonstrate that to some extent there is a lack of consistency over time and some schools – relative to other schools in the LEA – are improving at a greater or lesser rate than the average trend. A similar pattern of results is also found in equivalent data from the UK

[2]Note that school context variables controlled for in individual 2009–2012 contextualized value added models employed for the current paper were slightly different from those controlled for in the 2009 contextualised value added model used in our previous study (see Thomas et al., 2012), and provided generally better fit in terms of percentage of total variance explained.

Table 9.2 Summary estimate of school effect for four EEHE student outcomes measures over the period of 2009–2012 in LEA1 and LEA3

	Intra School correlation				Intra Year correlation			
3-level model	Total	Chinese	English	Math	Total	Chinese	English	Math
LEA1 (2009–2012)								
Unadjusted (raw) Model	26.3	20.1	25.4	22.0	1.3	1.5	1.6	1.8
Adjusted (contextualised value added) Model	7.4	3.5	5.7	5.6	1.6	1.5	2.4	2.0
LEA3 (2009–2012)								
Unadjusted (raw) Model	22.7	19.7	23.4	20.3	1.6	2.0	1.7	1.4
Adjusted (contextualised value added) Model	10.4	16.4	10.4	7.0	3.5	3.4	3.0	2.4

Note: Contextualised value added model includes controls for student variables (E*ESHS* prior attainment (English, Chinese, Maths), gender, age, major (science/arts), tuition fee, hukou, boarding status, previous school type, parents education, parents occupation, home possessions (e.g. number of books)) and context variables (% Hukou (city/town), % arts major, % fathers' education (college/university or above), School mean EESHS English prior attainment, School mean EESHS Math prior attainment, School standard deviation EESHS Total score)

(Thomas et al., 2007). In practice, for school self-evaluation these kinds of results are particularly helpful for schools to reflect on what policies and practices may have had a significant impact on their value added performance from cohort to cohort. Further, the lack of stability over time highlights that reliance on a single year's results may be misleading.

9.5.3 School and Teacher Effects

School and teacher effects were also calculated for LEA1 (East) and LEA3 (West) which provide further evidence of the need to consider the full complexity of educational effectiveness when estimating school effects (see Table 9.3). For the contextualised value added model, when teacher as well as school effects are estimated, the percentage of variance in students' EEHE scores attributable to differences between schools is reduced to 3–4% (LEA1) and 0% (LEA3), while the equivalent figures for differences between teachers are 5–13% (LEA1) and 6–22% (LEA3). These findings strongly indicate that educational effectiveness is most apparent at the teacher level in Chinese senior secondary schools, in line with findings in other international contexts such as the UK (e.g. Thomas, 2001).

Table 9.3 Comparison of raw and value added school and teacher effects for three EEHE student outcome measures

	Intra school correlation (%)						Intra teacher correlation (%)					
	Chinese		English		Maths		Chinese		English		Maths	
	LEA1	LEA3	LEA1	LEA3	LEA1	LEA3	LEA1	LEA3	LEA1	LEA3	LEA1	LEA3
Raw Model	14.8	15.1	16.0	26.0	18.3	21.7	18.5	9.6	30.3	20.7	30.2	15.7
Contextualised Value Added Model	3.0	0.0	3.6	0.0	4.2	0.0	5.1	5.9	11.0	21.7	13.4	11.0

Figures illustrate range of estimates from two LEAs with complete data for the 2012 cohort

Note: Contextualised value added model includes controls for student variables (E*ESHS* prior attainment (English, Chinese, Maths), gender, age, major (science/arts), tuition fee, hukou, boarding status, previous school type, parents education, parents occupation, home possessions (e.g. number of books)) and context variables (% Hukou (city/town), % arts major, % fathers' education (college/university or above), School mean EESHS English prior attainment, School mean EESHS Math prior attainment, School standard deviation EESHS Total score)

9.5.4 *Differential School Effects*

The findings from individual years' contextualised value added models (2 level, school/student) also revealed within school differential effects for EEHE Total score in relation to different groups of students. For example, in LEA1, the correlation between high and low prior attainment groups ranged from 0.43 to 0.69 over the 4 years (see Table 9.4). This indicates that the issue of differential school effectiveness may be hidden if only one overall measure is used to evaluate value added performance, and is especially pertinent in the Chinese context where many senior school students are taught in mixed ability classes. Clearly, evidence of differential effects within a school may help schools and teachers identify when less able students are struggling and when more able students are not being sufficiently challenged by their academic work. However, equivalent differential effects were found to a lesser extent across the three different curriculum subjects: Chinese, mathematics and English (correlations between any of two subjects ranged from 0.57 to 0.89). These subject correlations are also higher than may be expected on the basis of equivalent findings in the UK (e.g. Munoz-Chereau & Thomas, 2016; Sammons, Thomas, & Mortimore, 1997), thereby suggesting that whole school

Table 9.4 Differential school effects in *LEA1* correlations between school residuals for different student groups (Total EEHE scores) and subject outcomes over 2009–2012

LEA1	2009	2010	2011	2012
Gender	0.95	0.98	0.99	0.94
Prior attainment:				
Upper vs lower	0.43	0.60	0.69	0.65
Major (sciences vs arts)	0.82	0.81	0.80	0.79
Hukou (city vs village)	0.77	0.91	0.91	0.89
Subject:				
Chinese vs English	0.61	0.89	0.83	0.85
Chinese vs Maths	0.57	0.80	0.78	0.81
English vs Maths	0.66	0.81	0.86	0.80

Note that Leckie (2018a) argues that simple correlation estimates between school residuals from MLwin software are potentially biased, partly due to the shrinkage factor used in calculating the residuals, especially applicable in the case of very small samples. However, this issue is somewhat overstated in my view, especially given for larger samples any difference between correlations estimated by the model and those estimated via shrunken residuals is minimal and typically irrelevant when making substantive interpretations of the findings, as demonstrated for UK and Dutch secondary school residuals (Thomas, 2001). Moreover, there are good reasons why simple correlation estimates between school residuals are preferable, precisely because these relate directly to the more conservatively estimated (shrunken) school residuals and provide a standard method of comparing residuals from different outcomes and model specifications. Interestingly, Leckie (2018b) agrees that different model specifications (eg including student background factors or not) could also lead to differences in correlation estimates and notes this point could be made more strongly

policies and teacher collaboration across subject departments may be stronger in China than the UK, which is plausible due to the widespread practice of teachers meeting and discussing their work regularly in timetabled research and subject groups.

9.5.5 *The Impact of Teacher Professional Development on Student Value Added Performance*

Teachers professional development, Professional learning communities (PLCs) and related issues were also examined in detail via an individual teacher questionnaire administered in 2012. These data have fed into an analysis of the impact of PLC and teacher development factors on student "value added" progress. The findings indicate some impact of teacher professional development variables on student progress and achievement in higher education entrance examination results. For example, for LEA1 based on a 3-level (school/teacher/student) model controlling for prior attainment, student characteristics, and school context, as well as teacher professional development variables, the analyses show a significant ($\alpha = 0.05$ level) positive association of the following variables on one or more of the 2012 student outcomes (English, mathematics, Chinese): Total formal CPD hours, participation/engaging in informal dialogue with your colleague, participation in research project/writing paper, participation in training on instructional strategies and teaching methods, teacher collaboration and feedback, and educational resources/support for CPD. However, further research is needed with larger datasets in more regions to support these findings.

Other findings from the teacher survey can be summarised by the following four key points:

1. Opportunities for teacher development vary across regions. On average, teachers reported that their entitlement to attend formal continuing professional development (CPD) training activities ranged from, on average, 34 to 67 h over an 18-month period across the 3 regions.
2. CPD activities engaged in by teachers vary across regions (see Table 9.5). These activities also vary across teacher status, with junior teachers engaging to a greater extent in teaching competitions and observation of model lessons.
3. A need for further CPD opportunities on particular topics was also emphasised, particularly concerning academic subject knowledge, understanding of instructional strategies and psychological development of students.
4. The concept of PLCs was considered highly relevant and useful to promote teachers professional development in China, with 84% of teachers agreeing that the concept was meaningful. Teachers also reported, however, that some PLC features were seen as challenging in the Chinese context, such as promoting individual reflective professional inquiry and inclusion of all school staff.

Table 9.5 Percentage of teachers reporting that they have participated in the listed professional development activities during the past 18 months – based on 2012 Teacher survey

2012 Teacher survey – percentages of teachers participating in any of following kinds of professional development activities during the past 18 months (%)	LEA1	LEA2	LEA3
Courses/workshops on subject matter/methods	79.50	78.12	82.88 (74.97)
Education conferences or seminars	77.54	75.82	76.50 (67.64)
Qualification programme	52.47	45.01	44.80 (37.98)
Participation in a network of teachers	70.97	68.37	73.08 (62.84)
Individual or collaborative research	65.12	60.39	67.18 (57.88)
Mentoring and/or peer observation	91.33	84.81	91.56 (80.10)
Reading professional literature	77.56	72.05	82.99 (71.83)
Engaging in informal dialogue with colleagues	91.79	83.04	91.77 (80.49)
Research project, writing papers	76.46	70.55	77.33 (67.70)
Engaging in activities in teaching research groups	93.40	85.74	94.40 (83.68)

Note: (i) Calculated at individual level excluding missing values; (ii) Sample size (teachers/schools) = LEA1 (8093/41), LEA2 (7608/45), LEA3 (1814/21) with percentage of individual item missing values less than 2% for total sample (17,515); (iii) However, % item missing for LEA3 ranged from 9% to 16%, hence for LEA3 figures in brackets also show equivalent percentages not excluding missing values

Further, some interesting differences have been identified by comparing similar teacher survey findings from China and England. For example, the extent of teachers' team work and classroom observation seem to be reported more frequently in China, whereas sharing information with parents and community, the involvement of non-teaching support staff in PLCs, and target setting for individual students seem to be reported more frequently in England (Thomas & Peng, 2011a). Also, some aspects of professional learning may be unique to China and other similar East Asian countries, for example, the widespread existence of teacher research groups and model classes, and the common practice of teacher research publications, both of these being typically required for promotion in the Chinese system (Thomas & Peng, 2011b).

Interestingly, from a wide range of school input and process variables additionally tested in the contextualised value added models for the three Chinese LEAs, only a few were found to be statistically significant in 2009. Subsequently, for the 2010–2012 analyses some variation in the significance and direction (positive/negative) of the same associations was observed across cohorts and LEAs suggesting to some extent a lack of stability over time in the results. Nevertheless, regarding Total EHEE score, three key input or process factors were found to have a significant positive association for one or more of the cohorts examined for all 3 LEAs (student-teacher ratio, head teacher observation of class teaching, and teachers participate in decision making).

These selected findings from ITDEQC and IEEQC projects have clear implications for how teacher development, school effectiveness and educational quality may be better understood and evaluated, and the kinds of evidence that could be

potentially used by practitioners in the Chinese context to improve teachers' practice and student outcomes. The findings are also relevant to inform the development of educational policy and practice in mainland China and internationally, and in relation to issues of context specificity, school improvement and accountability.

9.6 Discussion

Improving the quality of education is a major goal in both developed and developing countries, given the clear links drawn between better student access and outcomes as well as poverty reduction and stronger economic growth (UNESCO, 2004). In this context, school effectiveness research has stimulated and focused educational policy makers' attention on the potential to raise overall levels of educational standards and student achievement. For example, Western governments such as the UK have placed a strong focus on encouraging schools and teachers to innovate and use new evaluation methods and data to inform their own evaluations of the education they provide as well as to feed into accountability and inspection frameworks, and these approaches have been linked to improved student outcomes. However, there is very little comparable empirical research evidence on the range and extent of school effectiveness in mainland China and how this links to teacher development issues. This chapter provides an example of original large-scale evidence that begins to address this crucial gap in order to inform new policy developments in mainland China; especially given the far-reaching educational reforms currently in progress (CMOE, 2010). Moreover, the findings regarding school effects in China across different regions, and the role of teacher development in enhancing student outcomes, critically develops and extends the international knowledge base of school effectiveness research.

In conclusion, this research provides some robust evidence that educational effectiveness research methods and findings are validated to some extent in the Chinese context, although differences from other settings also seem to exist in some aspects. For example, broadly comparable patterns of contextualised value added results and time trends were found between the UK and China, even though the extent of school effects may be larger in less affluent Chinese regions such as LEA3. Moreover, the conceptualisation of schools as "professional learning communities", which emerged from research in the USA and UK, seems to be relevant in China, although findings also indicate that the extent of some teacher development practices seems to vary considerably in different international contexts. Overall, this research provides a useful illustration of the kind of large-scale school effectiveness research project that would be possible if the appropriate datasets and evidence were systematically collected at the regional or national level in China. One important finding of the research over four cohorts is the apparent differences in the impact of contextual factors on school performance between regions within China, indicating that local context issues may operate in different ways and to varying degrees. Similar findings were also observed to some extent regarding the impact of input and process factors.

This suggests that separate regional evaluation systems may be most meaningful, and may be essential, in addition to any national evaluation system. Nevertheless, it is important to emphasise that only three LEAs were examined in this research, and it is clear that across a country as vast and diverse as China, more large-scale and representative educational effectiveness studies are needed, including further examination of the contribution of teacher development, school context and different levels within the education system to educational quality and effectiveness.

References

Bolam, R., et al. (2005). *Creating and sustaining effective professional learning communities* (Research report 637). London, UK: DfES and University of Bristol.

Cheng, K. M., & Wong, K. C. (1996). School effectiveness in East Asia: Concepts, origins and implications. *Journal of Education Administration, 34*(5), 32–49.

China Daily. (2018). *Household education spending totals $296 billion*. 16th Jan 2018. http://www.chinadaily.com.cn/a/201801/16/WS5a5d5e8aa3102c394518f794.html

Chinese Ministry of Education (CMOE). (2002). *Notice to actively promote the reform of primary and secondary school evaluation and examination system.* No26

Chinese Ministry of Education (CMOE). (2010). 国家中长期教育改革和发展规划纲要 *(2010–2020年)* The guidelines of the national plan for medium- and long-term educational reform and development (2010–2020). Retrieved July 30, 2010, from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_838/201008/93704.html

Chinese Ministry of Education (CMOE). (2011). *Methods of inspection and evaluation of quality education in primary and secondary schools*.

Chinese Ministry of Education (CMOE). (2013). *The Ministry of Education's suggestions on the implementation of 'Outlines for the reform of the evaluation of China's primary and secondary schools*. Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s7054/201306/xxgk_153185.html

Chinese Ministry of Education (CMOE) National Committee for Education (国家教育委员会). (1991a). 教育督导暂行规定(国家教育委员会第15号令). Provisional Regulations on Education Supervision (Decree No. 15). Retrieved 20 March 2011. from http://www.edu.cn/20041216/3124533.shtml.

Chinese Ministry of Education (CMOE) National Committee for Education (国家教育委员会). (1991b). 普通中小学校督导评估工作指导纲要. Guidelines for the Supervision and Evaluation of Ordinary Primary and Secondary Schools.

Chinese Ministry of Education (CMOE) National Committee for Education (国家教育委员会). (1995). 中华人民共和国教育法 Education Law. Retrieved March 22, 2011, from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_619/200407/1316.html

Chinese Ministry of Education (CMOE) National Education Committee. (1997). *Guidelines on supervision and evaluation of primary and secondary schools* (revised), No 4

Chinese Ministry of Education (CMOE). (教育部). (2006). 中华人民共和国义务教育法*(*修订*)* Compulsory Education Law of the People's Republic of China (Amendment). Retrieved March 22, 2011, from http://www.gov.cn/flfg/2006-06/30/content_323302.htm

Chu, H., & Cravens, X. C. (2012). Principal professional development in China: Challenges, opportunities, and strategies. *Peabody Journal of Education, 87*(2), 178–119.

Ding (丁延), Y. Q., & Xue (薛海平), H. P. (2009). 高中教育的一个生 函数研究 A study on the education production function with high school data. 中 范大学学 *(*人文社会科学版*)* Journal of Huazhong Normal University(Humanities and Social Sciences), *48*(2), 122–128.

Feng, D. 2007, School effectiveness and improvement in mainland China, In Avalos, B. & Townsend pp. 207–306). International handbook of school effectiveness and improvement. Springer, New York, NY.

Hopkins, D., Ainscow, M., West, M., Harris, A., & Beresford, J. (1997). *Creating the conditions for classroom improvement*. London, UK: David Fulton Publishers.

IEEQC. (2018). *Improving educational evaluation and quality in China*. Website accessed 12 Jan 2018. http://ieeqc.bristol.ac.uk

Leckie, G. (2018a). Avoiding bias when estimating the consistency and stability of value-added school effects using multilevel models. *Journal of Educational and Behavioral Statistics, 43*, 440–468.

Leckie, G. (2018b). personal communication 7th December 2018.

Liang, X., Kidwai, H., & Zhang, M. (2016). *How Shanghai does it: Insights and lessons from the highest-ranking education system in the world.* World Bank. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/24000/9781464807909.pdf

Liu, P. (2017). Transforming turnaround schools in China: A review. *School Effectiveness and School Improvement, 28*(1), 74–101. https://doi.org/10.1080/09243453.2016.1239641

Munoz-Chereau, B., & Thomas, S. M. (2016). Educational effectiveness in Chilean secondary education: Comparing different value added approaches to evaluate schools. *Assessment in Education, 23*(1), 26–52.

OECD. (2016). *Education in China: A snapshot*. Retrieved from https://www.oecd.org/china/Education-in-China-a-snapshot.pdf

Peng, W.-J., et al. (2013). Emerging perceptions of teacher quality and teacher development in China. *International Journal of Educational Development, 34*, 77–89. https://doi.org/10.1016/j.ijedudev.2013.04.005

Qiao, X., Yu, S. and Zhang, L. (2017) A review of research on professional learning communities in mainland China (2006–2015): Key findings and emerging themes. *Educational Management Administration & Leadership*. Epub ahead of print on May 11, 2017. DOI: 1741143217707523.

Reynolds, D., Caldwell, Cruz, Miao, et al. (2016). Chapter 10: Comparative educational research. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The international handbook of educational effectiveness: Research, policy and practice* (pp. 246–282). New York: Routledge.

Reynolds, D., Chapman, C., Clarke, D., Muijs, D., Sammons, P., & Teddlie, C. (2016). Chapter 17: Conclusions: The future of educational effectiveness and improvement research, and some suggestions and speculations. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The international handbook of educational effectiveness: Research, policy and practice* (pp. 408–439). New York: Routledge.

Robinson, B., & Yi, W. (2008). The role and status of non-governmental ('daike') teachers in China's rural education. *International Journal of Educational Development, 28*(1), 35–54.

Ryan, J., Kang, C., Mitchell, I., & Erickson, G. (2009). China's basic education reform: An account of an international collaborative research and development project. *Asia Pacific Journal of Education, 29*(4), 427–441. https://doi.org/10.1080/02188790903308902

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. Report prepared for Ofsted. Retrieved from http://www.highreliabilityschools.co.uk/_resources/files/downloads/school-effectiveness/psjhpm1995.pdf

Sammons, P. Sally M. Thomas & Peter Mortimore (1997). Forging links: Effective schools and effective department. London, UK: SAGE

Statistica. (2018). *Education in China – Statistics & facts*. Retrieved from https://www.statista.com/topics/2090/education-in-china/

Stoll, L., & Louis, K. S. (2007). *Professional learning communities: Divergence, depth and dilemmas*. Maidenhead, UK: Open University Press/McGraw-Hill.

Sun, H., Zheng, H., & Wang, Y. (2014). *Educational effectiveness research in Mainland China*. 5th World Conference on Educational Sciences. Procedia – Social and Behavioral Sciences 116, 4547–4551, 1877–0428. https://doi.org/10.1016/j.sbspro.2014.01.983.

Tan, C. (2013). *Learning from Shanghai: Lessons on achieving educational success*. Dordrecht, The Netherlands: Springer.

Teddlie, C., & Liu, S. (2008). Examining teacher effectiveness within differentially effective primary schools in the People's Republic of China. *School Effectiveness and School Improvement, 19*(4), 387–407.

Thomas, S. M. (2001). Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement, 12*(3), 285–322.

Thomas, S. M., Kyriakides, L., & Townsend, T. (2016). Chapter 9: Educational effectiveness research in new and emerging contexts. In Chapman, Muijs, Reynolds, Sammons, & Teddlie (Eds.), *The International handbook of educational effectiveness: Research, policy and practice* (pp. 220–245). London, UK/New York, NY: Routledge.

Thomas, S. M., Massoud, S., Munoz-Chereau, B., & Peng, W.-J. (2012). Educational quality, effectiveness and evaluation: Perspectives from China, South America and Africa. In C. Chapman, D. Muijs, A. Harris, D. Reynolds, & P. Sammons (Eds.), *School effectiveness and improvement research, policy and practice: Challenging the orthodoxy?* (pp. 125–145). London, UK/New York, NY: Routledge.

Thomas, S. M., & Peng, W. -J. (2011a). *Evaluating schools as effective professional learning communities*. Conference paper presented at 2011 AERA symposium, New Orleans.

Thomas, S. M., & Peng, W.-J. (2011b). Chapter 10: Methods to evaluate educational quality and improvement in China. In J. Ryan (Ed.), *Understanding China's education reform: Creating cross cultural knowledge, pedagogies and dialogue* (pp. 75–91). Routledge.

Thomas, S. M., Peng, W.-J., & Gray, J. (2007). Modelling patterns of improvement over time: Value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education, 33*(3), 261–295.

Thomas, S. M., Peng, W.-J., & Triggs, P. (2017). Professional learning communities in Chinese senior secondary schools. In A. Harris, M. Jones, & J. Huffman (Eds.), *Teachers leading educational reform: The power of professional learning communities*. Abingdon, UK: Routledge.

Thomas, et al. (2012). 学校效能增值评量研究 (Research on Value Added Evaluation of School Effectiveness). *Educational Research*, *7*, 29–35. National Institute for Education Sciences: Beijing.

UNESCO. (2004). *EFA global monitoring report 2005* (Education for all: The quality imperative). Paris, France: UNESCO.

Wang, T. (2015). Contrived collegiality versus genuine collegiality: Demystifying professional learning communities in Chinese schools. *Compare, 45*(6), 908–930.

Wu, X. (2013). *School evaluation for quality improvement in China, 30 years policies and practices* (Working paper). Beijing, China: National Institute for Educational Sciences.

Xin, T. (2017). *A glance of National Assessment of education quality in China*. Presentation. Retrieved from http://tcg.uis.unesco.org/files/resources/meetings/3rd/Session%208a_China.pdf

Yang, M., & Xiong, S. (2012). Supervision, driving force in the development of education. *People's daily, 7*, 6.

Zhang, J., & Pang, N. S. K. (2016). Exploring the characteristics of professional learning communities in China: A mixed-method study. *The Asia-Pacific Education Researcher, 25*(1), 11–21.

Zhang, J., Yuan, R., & Yu, S. (2017). What impedes the development of professional learning communities in China? Perceptions from leaders and frontline teachers in three schools in Shanghai. *Educational Management Administration & Leadership, 45*(2), 219–237.

Zhang, L. (2016). *School and teacher effectiveness of senior high schools in Western China***.** A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of doctor of philosophy (PhD). University of Bristol. October 2016

Zhang, T., & Minxia, Z. (2006). Universalizing nine-year compulsory education for poverty reduction in rural China. *International Review of Education, 52*(3–4), 261–286.

Zhao, M.-R., Zhou, J., & Zhu, X.-D. (2009). A survey on the participation and needs of primary and secondary school teachers' professional development. *Teacher Education Research, 21*(1), 62–67.

Zhao, Y. (2013). *Green evaluation: China's latest reform to deemphasize testing*. Retrieved http://zhaolearning.com/2013/06/24/green-evaluation-china%E2%80%99s-latest-reform-to-deemphasize-testing/

Zhou, J. (2017). China's path to achieve world-class 0065ducation. *ASIANetwork Exchange, 24*(2), 27–55. https://doi.org/10.16995/ane.239

Chapter 10
Three Decades of Educational Effectiveness Research in Belgium and the Netherlands: Key Studies, Main Research Topics and Findings

Marie-Christine Opdenakker

10.1 Introduction

In their state-of-the-art review on educational effectiveness research (EER), Reynolds et al. (2014) mention that EER addresses two core, foundational questions, namely "What makes a 'good' school? and "How do we make more schools 'good'?" (p. 197). To answer these questions, all the factors within schools and the educational system that might affect learning outcomes of students in both their academic and social-emotional development are foci of EER. The ultimate goal is to understand existing practices, and to establish and test models and theories in order to explain why some schools, learning environments and teachers are more effective than others, and in what way they are more effective (Creemers, 2007).

About three decades ago, these questions were picked up by a scholar from the Netherlands, namely Bert Creemers, who summarized around the mid-1980s the literature on educational research and, in particular, school effectiveness research done so far to support the creation of an outline for a new structure of secondary education in the Netherlands as well as to enhance the quality and effectiveness of Dutch education (Creemers, 1983; Creemers & Schaveling, 1985). A few years later, he started to collaborate and write papers with Jaap Scheerens, a Dutch scholar experienced in, among other things, educational evaluation and school organization research. At the end of the 1980s, a scholar from the Flemish-speaking part of Belgium (Flanders), namely Jan Van Damme, got triggered by the same questions too, and laid the foundations for educational effectiveness research in Flanders (Belgium) when he started a longitudinal research project on the school careers of

M.-C. Opdenakker (✉)
Department of Education, University of Humanistic Studies, Utrecht, The Netherlands

Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands
e-mail: m.opdenakker@uvh.nl; mariechristineopdenakker@gmail.com

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_10

secondary education students in Flanders financed by the Ministry of Education (Van Damme, De Fraine, Van Landeghem, Opdenakker, & Onghena, 2002).

In this chapter, an overview of the developments in the area of EER conducted in Belgium and the Netherlands over the past three decades (1987–2018) is given. After a brief historical overview of the origins and evolution of educational effectiveness research in both countries, with special attention to some important (large-scale, longitudinal) key studies, an overview of dominant research topics and research issues will be addressed. Furthermore, main findings of EER studies in both countries will be provided and discussed, as well as illustrated with results of a selection of key studies. Of course, it is not possible to present a full and comprehensive overview of all the (results of the) EER studies of the past three decades in a chapter of limited length such as this: our search with EBSCOhost, additional search in the School Effectiveness and School Improvement (SESI) Journal, and in Pedagogische Studiën (a SSCI-ranked Dutch Journal on educational sciences), alongside some additional use of the snowball method, resulted in a large amount of studies of which, after reading the summary, 189 were suitable for inclusion in this review. Therefore, first a general overview about research topics and addressed issues throughout the three decades will be given, and in particular, attention will be paid to developments and changes over time and differences between EER in the Netherlands and Belgium. Next, a summary of the main findings of the studies will be presented and key findings will be discussed. Since international EER has established that teachers and classes are crucial for student outcomes and learning and are important layers for variation (Scheerens, 2017), and EER in Belgium and the Netherlands is in agreement with these findings of international EER, the focus will be mostly on teacher behaviour and what students experience in the classroom. In addition, effects of group composition (also in relation to what students experience in the classroom) will be addressed, as well as differential effectiveness and relationships between school and learning environments at class level. The presentation of findings will be restricted to empirical findings of content-related studies. Hence, exclusively theoretical or methodological articles are not or only occasionally addressed in this chapter.

10.2 Educational Effectiveness Research in the Netherlands and Belgium: An Historical Overview of Developments

10.2.1 Origins and First Developments of Educational Effectiveness Research in the Netherlands

First of all, it must be mentioned that, in contrast to some leading EER countries like the United States, ‘school effectiveness research’ – which was the first name to address EER – was in the Netherlands strongly routed in research on classroom studies, teacher behaviour and teacher effectiveness. In addition, it was inspired by

the desire to improve schools and education in general. This was not a coincidence, since the early work of one of the two founders of EER in the Netherlands, namely Bert Creemers (1974), concentrated on the effects of teaching styles on initial reading proficiency of pupils.

Furthermore, in the second half of the 1980s, a few preliminary Dutch studies were carried out in primary education to find evidence for the so-called 'school effectiveness enhancing factors of Edmonds five-factor model' (Edmonds, 1979) in relation to students' academic achievement. However, these first studies were not very successful. For example, the translation of the instruments used in the US to measure the effective school factors failed to deliver reliable measurements of the factors (with the exception of the factor 'orderly atmosphere aimed at the stimulation of learning', which additionally proved to have a positive relation with average achievement; Vermeulen, 1987). Further, the factor 'educational leadership', which was a clear school effectiveness enhancing factor in the US, failed to demonstrate a clear positive relationship with average achievement in the first Dutch studies of van de Grift (1987). Apart from these problems, these first studies did not take into account the nested structure of education in their data analyses.

However, these flaws were recognized a few years later: At the end of the 1980s, when the first and second International Congress of School Effectiveness and School Improvement (ICSEI) took place (in 1988: London; in 1989: Rotterdam) the Dutch country reports mentioned the importance of good statistical techniques to improve research and to get more decisive conclusions. In addition, these reports contained a plea for more theory development to help generate explanations as to why some variables (at school level) seemed to be associated with differences in achievement (Creemers & Lugthart, 1989). This methodological and theoretical interest guided EER in the Netherlands from the late 1980s with the development of several conceptual multilevel models of educational effectiveness, e.g. the contextual multilevel model for school effectiveness (Scheerens & Creemers, 1989), the integrated model of school effectiveness (Scheerens, 1990, 1992) focusing on organizational factors in relation to the instructional level, the comprehensive model of educational effectiveness (Creemers, 1994) focusing on classroom instruction (teacher behaviour, grouping procedures, curriculum) and the classroom-school interface. More recently, these models were updated in the dynamic model of educational effectiveness of Creemers and Kyriakides (2008) and the recent conceptual framework of Scheerens (2016) including system effectiveness. From the development of these frameworks, several studies were set up to test (parts of) these models (e.g. de Jong, Westerhof, & Kruiter, 2004). Furthermore, research has taken place to address methodological topics concerning foundational issues in EER such as stability of school effects over time and consistency of effectiveness between subjects (e.g. Luyten, 1994). These studies were not only important for the establishment of the conceptual integrity of the field, but were also important for practical purposes such as the assessment of excellence in an accountability context. In addition, new statistical programs for multilevel modeling (taking into account the nested structure of education with e.g. students belonging to schools) such as VACL, HLM and MLN (later developed for Windows as MLwiN) were used in research studies (e.g. Brandsma & Knuver, 1989).

In addition to the already-mentioned topics in the Dutch country reports of ICSEI, Creemers and Lugthart (1989) urged for a distinction between school effectiveness and effectiveness at the level of the learning environment in the class (instruction) in order to facilitate teacher effectiveness and learning models. The adoption of this suggestion is noticeable in the Dutch conceptual models of educational effectiveness already mentioned.

Since the start of the ICSEI conferences, Dutch educational effectiveness researchers got along quite easily with international scholars and played an important role in the international EER community.

10.2.2 Origins and First Developments of Educational Effectiveness Research in Belgium

The onset of interest in EER in Belgium was a bit later than was the case in the Netherlands. In Belgium, Jan Van Damme can be seen as the founding father of EER in Flanders (the Flemish-speaking part of Belgium). Coming from an interest in school careers of students in different school types in secondary education and having a background in aptitude-treatment interaction research, he got inspired by American, British and Dutch EER scholars he met at the annual AERA conference in the US. He convinced the Ministry of Education of Flanders to give him a grant to make a start with the first longitudinal EER study in Flanders, the 'Longitudinal Research in Secondary Education' (LOSO: 'Longitudinaal Onderzoek Secundair Onderwijs') project (Van Damme et al., 2002). In this project, which started in 1989, he combined all of his aforementioned interests. The project resulted in longitudinal data collection in which school careers, and cognitive and the non-cognitive student outcomes of a cohort of students (and their classmates) belonging to almost all schools of three areas in Flanders were collected during their careers in secondary education, as well as characteristics of their secondary education schools (90 schools, to a certain extent representative of the Flemish secondary schools in general). Pioneers of the research team were, among some others, Alexander Minnaert and Marie-Christine Opdenakker, who played an important role in the construction of the achievement tests and other instruments for the data collection, such as questionnaires to measure characteristics of educational practice. The construction of (multiple choice) achievement tests was quite an endeavor, since in Flanders no standardized central exams existed at the time of data collection (and still do not exist), and schools of different denomination/school sector and/or (academic) tracks had their own curricula, though luckily they seemed to have much in common. In addition, this project marked the first time that common achievement tests for different school sectors in Flanders were constructed. This paved the way for comparing the effectiveness of school sectors, though this was a very delicate political topic in the 1990s (and, in fact, it still is). Therefore, tests were very carefully constructed in close collaboration with experienced subject matter experts

within the inspectorate, and with teachers of both school sectors (Catholic and public), paying much attention to content and curriculum validity. Opdenakker also played an important role in the operationalization of the concepts of school characteristics, paying attention to school organization and school climate factors as well as to learning environment and teacher characteristics (related to instruction, climate and teacher styles) at the class and teacher level. The LOSO project was quite innovative in collecting a variety of outcomes (several cognitive and non-cognitive outcomes), in taking a multi-informant perspective with regard to collecting data of school and learning environment characteristics (school leader, a stratified sample of teachers and subject teachers, students) and in being a longitudinal EER study. In addition, an elaborate data collection was carried out among the parents of the students in order to obtain information on student background characteristics and their home environment. The project resulted in a very rich database that delivered input for several scientific reports, PhD theses[1] (e.g. Opdenakker, 2003; De Fraine, 2003; Van de gaer, 2006; Pustjens, 2008; Vanwynsberghe (partially), 2017) and many articles.[2] After the end of the project, a follow-up was realized which made it possible to consider long-term effects of secondary schools on passing on to post-secondary education, the transition to and success in higher education,[3] and the transition to and success at the labor market. (A new follow-up focusing on success at the labor market is recently started and is still in progress.)

The main interest of EER in Flanders during the 1990s was to give an answer to the questions, what works in education, and what makes secondary schools and learning environments in Flanders effective. Theoretical work was not much a focus in Flemish EER, which is in sharp contrast to the Netherlands. An exception is the work of Opdenakker (2004) who developed her own multilevel multifactor model of educational effectiveness building on the already existing (international) models of educational effectiveness and international EER in the 1990s.

After the LOSO project finished in 2001, Van Damme started, together with some other colleagues of the University of Leuven, Antwerp and Ghent, on a new longitudinal EER study that was an initiative of the Flemish Ministry of Education, namely the "Educational careers throughout primary education" project (SiBO: 'Schoolloopbanen in het BasisOnderwijs'). In this study, about 6000 students of 196 primary schools were followed from the age of 4–5 throughout their primary education. Research topics were, among others, individual differences in the development of students and their school careers, and effects of individual and environmental background characteristics referring to the home environment as well as the school and class environments. When the students were 17 years old, a follow-up study was started. The SiBO project also resulted in several PhD theses

[1]Most of them consisted of a collection of articles which were part of this review study.

[2]E.g. Van Damme et al. (2002), De Fraine et al. (2003), De Fraine et al. (2007), Opdenakker and Van Damme (2000a, 2000b, 2001, 2006a, 2006b, 2007), Opdenakker, Van Damme, De Fraine, et al. (2002), Pustjens et al. (2008), Pustjens et al. (2007), Pinxten et al. (2012), and Van de gaer et al. (2006, 2007, 2009).

[3]See e.g. Pustjens et al. (2004).

(e.g. Verachtert, 2007; Goos, 2013; Speybroeck, 2013; De Bilde, 2013; Boonen, 2013; Vanlaar, 2015; Vandecandelaere, 2015; Belfi, 2015; Vanwynsberghe (partially), 2017) and articles (mostly of the just mentioned authors)[4] in which the developments of students and effects of school and learning environments were the core of the research focus.

After the LOSO project, a few other scholars in Flanders started some (small) EER studies as well. Flanders also participated in the 'International System for Teacher Observation and Feedback (ISTOF) project (Teddlie, Creemers, Kyriakides, Muijs, & Yu, 2006). In this project, a unique teacher observation instrument in the EER field was developed by a team from 20 countries, using an iterative Delphi process to ensure cross-cultural relevance and validity. The instrument focused on teacher behaviour, with an interest in investigating which features were effective across educational systems. Van Damme and Opdenakker coordinated the Flemish team of the ISTOF project and developed, in addition to the teacher observation instrument, a student questionnaire in close alignment with the observation instrument. Opdenakker adapted this questionnaire for use in the Netherlands (Opdenakker & Minnaert, 2011).

In Wallonia, the French-speaking part of Belgium, interest in EER came about much later than in Flanders and was highly influenced by possible effects of group composition on student outcomes, a topic that was introduced first in Flanders (Opdenakker & Van Damme, 2001). Xavier Dumay and Vincent Dupriez were among the first researchers of the French-speaking part of Belgium who addressed EER topics (see e.g. Dumay & Dupriez, 2007, 2008).

10.3 An Overview and Comparison of (Trends in) Dominant Research Topics and Research Issues

As was already clear from the previous section, interest for EER began earlier in the Netherlands than in Belgium. As a consequence, the number of articles reporting on EER was much higher in the Netherlands compared to Belgium in the first decade of EER in both countries (1987–97). In fact, only a small number of articles on EER in Belgium were published during this period. Afterwards, in both countries, the number of studies increased markedly from one decade to the next with an accelerated increase in Belgium. As a result, about 55% of the studies between 1987–2018 refer to Dutch studies[5] and about 45% are related to Belgian, mostly Flemish

[4] An extensive selection of these articles are discussed in Sect. 10.4 'The knowledge base'.

[5] Some illustrative studies for the first decade (1987–97) of EER in the Netherlands are Hofman (1994), Knuver and Brandsma (1993) and Luyten (1994). For the second decade (1998–2007) illustrative studies are Driessen and Sleegers (2000), de Jong et al. (2004) and Hofman et al. (1999), and for the last decade (2008–18) studies to be mentioned are Opdenakker et al. (2012), Opdenakker and Minnaert (2011) and ten Bruggencate et al. (2012).

studies.[6] This indicates that Belgium (Flanders) caught up very quickly once the interest in EER was established.

10.3.1 School Level

Taking the articles of the two countries as a whole, about half of the relevant studies refer to the effectiveness of primary schools and about half of them address the effectiveness of secondary schools. It is striking that almost no study addresses the effectiveness of institutions of higher education. Comparing the evolution of EER over time within both countries with regard to their focus on the effectiveness of primary schools versus secondary schools, different patterns are noticed across the two countries to some extent. While reporting on EER in scientific articles from both countries started with a stronger focus on the effectiveness of secondary schools, this changed to an almost equal attention to both levels from the second decade on (1998–2007) in the Netherlands, whereas, in Flanders an evolution was visible from a stronger focus on the effectiveness of secondary schools in the first two decades (1987–97, 1998–2007) to a stronger focus on the effectiveness of primary schools in the last decade (2008–18). Largely, this focal shift can be explained by the availability of data from the LOSO project on the effectiveness of secondary schools from the second decade on (1998–2007) in Belgium, and of the availability of data on the effectiveness of primary schools of the SiBO project from the last decade on (2008–18).

10.3.2 Outcome Criteria

Looking at the outcome measures used in the effectiveness studies, it is clear that cognitive outcomes are (most) often used (in about half of the studies), referring mainly to math and language achievement.[7] This is true for effectiveness studies on primary as well as on secondary schools. However, language outcomes are a bit more dominant as effectiveness criteria compared to math outcomes with regard to

[6] Some illustrative Belgian Flemish studies for the second decade (1998–2007) are Opdenakker and Van Damme (2000a, 2001, 2006a, 2007), De Fraine et al. (2003) and Pustjens et al. (2004). For the third decade (2008–18), the study of Dumay and Dupriez (2008) (French speaking part of Belgium) and the Flemish studies of Boonen, Van Damme and Onghena (2014), Belfi et al. (2014), Vanwynsberghe et al. (2017a, 2017b, 2017c) and Agirdag et al. (2012) are illustrative.

[7] A selection of illustrative examples with regard to primary education in which language achievement is used as the outcome criterion are van der Slik et al. (2006) and Dumay and Dupriez (2008). In studies of, for example, Hofman et al. (1999) and Dumay (2009) math achievement is used as the outcome criterion. Illustrative studies in secondary education using language as the outcome criterion are, among others, De Fraine et al. (2003) and den Brok et al. (2004). Math achievement is used as outcome criterium in studies on secondary education such as de Jong et al. (2004) and Opdenakker and Van Damme (2001, 2006a, 2007).

primary schools, although these are accompanied by math achievement as an effectiveness criterion as well in about 1 out of 2 studies.[8] On the other hand, math is a bit more prevalent as the effectiveness criterion in studies on secondary schools' effectiveness, and is only rarely accompanied by language achievement as an additional effectiveness criterion.[9] Furthermore, only a few studies used science achievement, or looked at an indicator of the school career of students such as their educational position or their study success[10] at a particular moment in their school career. Most of these effectiveness studies referred to the effectiveness of secondary schools. There was no substantial change in the relative prevalence of particular cognitive effectiveness criteria over time for both primary and secondary schools' effectiveness studies.

In addition to cognitive outcomes, non-cognitive outcomes are frequently used in EER studies in both countries: in more than one third of the studies, non-cognitive outcomes were investigated as effectiveness criteria.[11] Educational effectiveness studies on secondary schools are a bit more represented in these studies compared to studies on primary schools. With regard to non-cognitive outcome indicators, motivation for school or motivation for a specific subject, academic engagement, attitudes towards school or attitude to a subject, and indicators of well-being are the

[8]Examples of studies in primary education in which both language and math achievement are used as outcome criteria are Brandsma and Knuver (1989), Driessen and Sleegers (2000), and Peetsma et al. (2006).

[9]An exception is, for example, the study of Opdenakker and Van Damme (2000a) conducted in secondary education in Flanders. In this study, both math and language achievement were used as effectiveness criteria.

[10]The terms educational position and study success refer in this context to outcomes referring to the educational track or educational/curricular program students are in at a particular moment in time during their educational career (for example, an academic track, a general track or a (pre-)vocational track; it can also refer to different educational/curricular programs). It can also refer to the grade students attend after a specified period of time. While the educational position can be the result of a choice for a track or an educational/curricular program, study success always includes a judgement of the successfulness of a certain period of time in education. For example, this can be a judgment on the degree to which the track or educational/curricular program the student attends after a certain period of time is more academic-oriented or less academic-oriented (e.g. vocation-oriented), or the grade the student attends after a certain period of time (referring to no grade retention, or one or more years not passed).

[11]For the Netherlands, work of Opdenakker and colleagues, mainly in secondary education, (e.g. Maulana et al., 2013; Opdenakker, 2013, 2014; Opdenakker & Maulana, 2010; Opdenakker & Minnaert, 2014; Opdenakker et al., 2012; Stroet et al., 2015, 2016; van der Werf et al., 2008) is quite illustrative. Opdenakker and Minnaert (2011, 2014) also did some work in Dutch primary education. Furthermore, work of Peetsma and colleagues (e.g. Peetsma et al., 2006; Hornstra et al., 2015a, 2015b) and work of den Brok and colleagues (e.g. den Brok et al., 2005, 2010) respectively in primary and secondary education could be mentioned. Also Reezigt and Weide (1990) and Meelissen and Luyten (2008) paid attention to non-cognitive outcomes such as student attitudes. In Flanders, early work of Opdenakker and Van Damme (2000a) as well as work of Van Landeghem (e.g. Van Landeghem et al., 2002), of Van de gaer (Van de gaer et al., 2009), of Van Petegem (e.g. Van Petegem, Aelterman, et al., 2008) and of Vanwynsberghe (Vanwynsberghe et al., 2017a) addressed non-cognitive outcomes.

most commonly-investigated outcomes. However, there are some differences between studies referring to primary-school versus secondary-school effectiveness. While motivation and well-being are often addressed in effectiveness studies focusing on secondary schools, effectiveness studies focusing on primary schools often use social-emotional aspects as criterion of effectiveness. Additionally, academic engagement has often been studied at both school levels.[12]

10.3.3 Short-Term Versus Long-Term Effects

In agreement with the international EER literature (Creemers, Kyriakides, & Sammons, 2010) almost all studies in Belgium and the Netherlands address the short-term effects of schools, teachers or classes. Short-term effects of schools, teachers and classes refer to effects over a brief time period, for example, over a few or several months within a certain school year. The term "long-term effect" is used in the international literature in different ways. For some (e.g. Bressoux & Bianco, 2004; Kyriakides & Creemers, 2008) it can refer to the effect of schools, teachers or classes after at least one or several school years, while others (e.g. Teddlie & Reynolds, 2000) use the phrase "long-term" as a synonym for continuing effects and refer to this when effects of schools, teachers or classes on students from one phase of schooling to another is addressed (constituting a more narrow definition). Vanwynsberghe, Vanlaar, Van Damme and De Fraine (2017a) refer to long-term effects as "*the effects of schools (or teachers) at a particular moment of a student's educational career, on the outcomes of a student after at least one year, during which the student experienced another influence (e.g. another school or teacher)*" (pp. 84–85). In this chapter, we will use the concept long-term effect in keeping with the definition of Vanwynsberghe et al. (2017a).

It must be stated that, although also scarce in Belgium and the Netherlands, Belgium in particular has taken a leading position in the international literature on long-term effects of primary schools, addressing long-term effects on cognitive and

[12]Some illustrative examples of studies on motivation and/or well-being in secondary education are Opdenakker and Van Damme (2000a), Van Landeghem et al. (2002), Van Petegem, Aelterman, et al. (2008), den Brok et al. (2004), Opdenakker et al. (2012), and Stroet et al. (2015). Examples of studies in primary education on social-emotional aspects are Hornstra et al. (2015a) and Karssen et al. (2016). Some examples of studies on student engagement are de Bilde et al. (2013), Lietaert et al. (2015), Opdenakker (2014), and Opdenakker and Minnaert (2011).

non-cognitive outcomes as well as on educational positions and other school career outcomes.[13]

10.3.4 Explaining Factors: Level, Type and Attention to Differential Effects Versus Genericity

In general, most effectiveness studies in Belgium and the Netherlands (about 2 out of 3) focus on school factors in an attempt to explain differences between students and about a quarter of them include class factors as well. About 30% of the studies included in this review exclusively investigate the effects of class factors. This observation is similar for Belgium and for the Netherlands. However, there are some differences between countries as well. For example, in Belgium, effectiveness studies on secondary schools changed focus from an exclusive focus on school factors or on a combination of school and class factors in the second decade (1998–2007) to an almost exclusive focus on class factors in the last decade (2008–18), while studies on the effectiveness of primary schools evolved from an almost exclusive focus on class factors to an exclusive focus on either class or school factors.[14] These differences in the evolution of the focus between primary and secondary schools effectiveness studies are not visible in the Netherlands. For both primary and secondary levels, an evolution was visible from an almost exclusive focus on school factors in the first decade to an almost exclusive focus on either school or class factors in the last decade.[15]

[13]See for example work of Pustjens and colleagues (2007), (2008) and work of Vanwynsberghe and colleagues (Vanwynsberghe et al., (2017a,) (2017b), (2017c)).

[14]Illustrative studies with regard to secondary education in the second decade are, for example, Opdenakker and Van Damme (2006a, 2007), Van Houtte (2004), Pustjens et al. (2007) (school level factors) and Opdenakker and Van Damme (2000a), Opdenakker, Van Damme, De Fraine, et al. (2002), De Fraine et al. (2003), Van Landeghem et al. (2002) (school and class factors) and, for the third decade, examples of studies focusing exclusively on class factors are work of Van Petegem and colleagues (e.g. Van Petegem, Aelterman, et al., 2008), Lietaert et al. (2015), Vause et al. (2010), and Denies et al. (2016). For primary education, an illustrative study of the second decade focusing exclusively on class factors is the study of Gadeyne et al. (2006). Examples of studies related to the third decade focusing exclusively on school factors are, among others, Dumay et al. (2013), Agirdag et al. (2012), Boonen, Speybroeck, et al. (2014), de Bilde et al. (2013), and Belfi and colleagues (2014). A selection of examples of studies of the same decade focusing solely on class factors are Boonen, Van Damme and Onghena (2014), Vanlaar and colleagues (2014), and work of De Smedt and Van Keer (e.g. De Smedt & Van Keer, 2018; De Smedt et al., 2016).

[15]Illustrative studies of the first decade (1987–97) are Brandsma and Stoel (1987), van de Grift (1987), Luyten (1994), Hofman (1994, 1995), Roeleveld and Dronkers (1994) and van Batenburg and Lokman (1991). Some illustrative studies of the third decade (2008–18) focusing exclusively on school factors or school effects are van der Werf et al. (2008), Guldemond and Bosker (2009), Dijkstra et al. (2015), and Hofman et al. (2013). A selection of illustrative studies focusing exclusively on class factors and class/teacher effects are Opdenakker and colleagues (Opdenakker,

Looking more carefully into the investigated **school factors**, it becomes clear that school leadership, policy and organizational characteristics are the most frequently studied school factors in both countries. While there is no difference between primary school and secondary school EER studies in the extent to which these factors appear in the Netherlands, there is a difference between primary and secondary school EER studies in Belgium with regard to this aspect. Specifically, the mentioned factors are more often studied in secondary than in primary schools in Belgium. In addition, with regard to organizational and policy characteristics, policy and governance structure is most often studied in the Netherlands, while in Belgium this mostly refers to an investigation of teacher collaboration. Furthermore, there is less attention devoted to school climate in the Netherlands than in Belgium, which is almost exclusively related to secondary school effectiveness in Belgium.[16]

With regard to **class and teacher factors**, indicators of (the quality of) instruction in the learning environment are far more often studied than indicators of the quality of learning and relational climates in classes of both countries. In addition, effects of class climate are – in both countries – more often investigated in secondary schools than in primary schools. A few studies addressed grouping or organizational arrangements within classes.[17]

Some studies in both counties also paid attention to **school or class context** variables like denomination, school/class size and the educational offerings (e.g. tracks[18]) in schools.[19] Denomination was most frequently studied in both countries.

2014; Opdenakker & Minnaert, 2011; Opdenakker et al., 2012; Maulana et al., 2013, 2016), de Haan et al. (2013), Hornstra et al. (2015a, 2015b), and Veenman et al. (2013).

[16]Illustrative studies of the Netherlands in which attention is devoted to effects of school leadership/school organizational characteristics (in primary education) are van de Grift and Houtveen (1999), Hofman, Hofman and Gray (2015) and (in secondary education) Kruger et al. (2007) and Hofman et al. (2001). The studies of Opdenakker and Van Damme (2001, 2006a, 2007) are examples of Belgian studies in which attention is paid to effects of teacher collaboration. In studies of, for example, Hofman et al. (1999) and Opdenakker and van Damme (2001, 2006a) and Dumay and Dupriez (2007) effects of school climate were addressed.

[17]For example, the study of Boonen, Van Damme, and Onghena (2014) pays attention to this kind of class factors.

[18]Tracks refer in this context to curricular offerings in (secondary) schools which are for students with different academic abilities (and/or interests). By means of tracks, students are separated by academic ability and/or interests into different groups/classes. In some countries like the UK, this refers to streams. In Belgium and the Netherlands, schools can differ with regard to the curricular programs they offer. More information on the Belgian (Flemish) educational system with regard to tracks and curricular offerings of schools can be found, for example, in Opdenakker and Van Damme (2006a). Information on tracks and class compositions related to this topic in the first grades of Dutch educational system can be found, for example, in Opdenakker et al. (2012).

[19]Some examples of studies on effects of denomination are De Fraine et al. (2003), Opdenakker and Van Damme (2006a), Agirdag et al. (2017), Hofman et al. (2002) and Dronkers and Robert (2003). Illustrative studies on effects of school size are, Luyten (1994), Opdenakker and Van Damme (2006a, 2007), and ten Bruggencate et al. (2012) and examples of studies investigating the effects of

In addition, in about 30% of the studies reviewed, attention was paid to **school or class composition**. This occurred in relatively more studies related to Belgian schools and was most pronounced for effectiveness studies on Belgian secondary schools. While cognitive composition as well as composition related to SES and ethnicity was often addressed in Belgium, the cognitive composition was relatively less addressed in Dutch studies.[20]

Attention to **differential effects** of schools, teachers or classes was already visible from the first decade of EER (1987–97). However, this consideration was apparent in only a very small number of studies at that time. In the following decades (1998–2007, 2008–18) attention was paid to differential effects, but still in a rather small number of studies. In general, it can be concluded that, in EER studies in both countries, in about 1 out of 5 studies attention is paid to differential effects. There are no differences in this respect between studies on primary and secondary schools. A striking difference between the two is, however, the specific nature of differential effects on which the relevant studies focus. While in studies addressing secondary education the main focus of differential effects is on schools and school characteristics (about 2 out of 3 studies), more than 55% of the EER studies on primary schools with an interest in differential effects focus only on class or teacher characteristics and more than 35% solely on school characteristics. This is more or less the same across the two countries, although the focus of differential effects is more pronounced in the Netherlands and attention to differential effects of both schools and classes/teachers within the same study on secondary schools is almost solely a Belgian (Flemish) undertaking.

With regard to which characteristics of schools, teachers and classes are explicitly focused on to investigate differential effects, and in relation to which student characteristics, of the *school characteristics* school composition is most frequently studied (56% of relevant reviewed studies), followed by school context (33%). Of the school context variables, denomination is the most frequently-studied characteristic, and this is often operationalised in relation to student SES. The most-often studied school composition characteristics are related to student SES and ethnic origin/nationality/language at home. These are often studied in relation to student SES or students' cognitive abilities.

---

schools with different educational offerings are Opdenakker and Van Damme (2006a) and Timmermans et al. (2014).

[20]Example studies from the Netherlands paying attention to school composition are Driessen (2002) and Guldemond and Bosker (2009), and example studies from Belgium are work of Opdenakker (Opdenakker & Van Damme, 2001; Opdenakker, Van Damme, De Fraine, et al., 2002), De Fraine et al. (2003), Agirdag et al. (2012), and of Dumay and Dupriez (2008). Illustrative studies paying attention to class composition are, from the Netherlands, Driessen and Sleegers (2000), van der Slik et al. (2006), and Peetsma et al. (2006). Illustrative studies from Belgium are the already mentioned studies of Opdenakker and colleagues (Opdenakker & Van Damme, 2001; Opdenakker, Van Damme, De Fraine, et al., 2002), Opdenakker and Van Damme (2006b), Opdenakker, Van Damme and Minnaert (2002a, 2002b, 2005), Van Landeghem et al. (2002), Dumay and Dupriez (2007), Van de gaer et al. (2006), Vanwynsberghe et al. (2017a), and the already mentioned study of De Fraine et al. (2003).

With regard to *class characteristics*, class process factors are studied in about 75% of the reviewed studies on class characteristics; and class composition in the remaining 25%. Only a few studies paid attention to differential effects of class process and composition together. Almost all studies on the differential effects of class process characteristics focus on instructional factors; about half of them also pay attention to differential effects of class climate indicators. The differential effects of class instructional factors are often studied in relation to student gender, cognitive abilities or ethnic origin/language spoken at home; class climate is almost always studied in relation to student gender.[21]

10.4 The Knowledge Base

10.4.1 The Importance of Schools, Teachers and Classes

10.4.1.1 Short-Term Effects on Status and Growth of Cognitive, Non-cognitive and School Career Indicators

Based on all reviewed EER studies conducted in Belgium and the Netherlands in the last three decades, it is clear that schools, teachers and classes – or, otherwise stated, students' experiences in schools and classes – matter: students' cognitive and non-cognitive functioning and their school careers seems to be affected by the schools they attend, the teachers they have and the classes they belong to.

Furthermore, in agreement with the international literature, it is evident that the effects of schools, teachers and classes together seem to be larger on cognitive outcomes than on non-cognitive outcomes (see e.g. Opdenakker & Van Damme, 2000a for a comparison). For example, while for cognitive outcomes like math and language achievement (e.g. reading comprehension, text writing) the amount of variance above the student level ranges, depending on the EER study, from 34% to 56% in secondary education in Belgium (language: De Fraine, Van Damme, Van Landeghem, Opdenakker, & Onghena, 2003; math: Opdenakker, Van Damme, De Fraine, Van Landeghem, & Onghena, 2002; Opdenakker & Van Damme, 2000a, 2000b, 2001; Dumay & Dupriez, 2007; Van den Broeck, Opdenakker, & Van Damme, 2005) and from 10% to 31% in primary education in Belgium (Boonen, Van Damme, & Onghena, 2014; Dumay, 2009; Dumay, Dupriez, & Maroy, 2010; De Smedt, Van Keer, & Merchie, 2016; Vause, Dupriez, & Dumay, 2010), it ranges from 5% to 20% for non-cognitive outcomes in primary and secondary education in

[21] A selection of illustrative studies investigating effects of school and/or class composition are, from Belgium, Opdenakker, Van Damme, De Fraine, et al. (2002), De Fraine et al. (2003), Van de gaer et al. (2006), Lietaert et al. (2015) and, from the Netherlands, Peetsma et al. (2006), den Brok et al. (2010), and Opdenakker and Minnaert (2014).

Belgium (e.g. Gadeyne, Ghesquière, & Onghena, 2006; Opdenakker & Minnaert, 2011; Opdenakker & Van Damme, 2000a; Van Damme, Opdenakker, & Van den Broeck, 2004; Van Landeghem, Van Damme, Opdenakker, De Fraine, & Onghena, 2002; Van Petegem, Aelterman, Van Keer, & Rosseel, 2008).[22] Furthermore, there is some evidence that teachers and classes are more important for math than for language in primary education (e.g. Boonen, Van Damme, & Onghena, 2014; Gadeyne et al., 2006; Meijnen, Lagerweij, & de Jong, 2003) and that some non-cognitive outcomes are more school/teacher/class related than others. Motivational aspects like interest in learning tasks, controlled motivation and academic self-concept seem to be less related to classes and schools compared to wellbeing at school, (student perceptions of) relationships with teachers, academic engagement, amotivation, autonomous motivation, performance avoidance behaviour and attitudes towards math, and psychosocial adjustment (internalizing and externalizing problem behaviour) (Gadeyne et al., 2006; Van Damme, Van Landeghem, De Fraine, Opdenakker, & Onghena, 2000[23]; Opdenakker & Van Damme, 2000a; Opdenakker & Minnaert, 2011; Stroet, Opdenakker, & Minnaert, 2015; Van Damme et al., 2004). Recently, a few studies (e.g., Dijkstra, Geijsel, Ledoux, van der Veen, & ten Dam, 2015) also paid attention to citizenship competences (in primary education) in the Netherlands, and found that schools mattered for this outcome – which referred to students' knowledge, skills, attitudes and reflection that they need to adequately address the everyday social challenges of living in a pluralistic and democratic society – as well.

From the preceding section, one might get the impression that secondary schools, teachers and classes matter more for the cognitive functioning of students than primary schools, teachers and classes. However, it is necessary to interpret the mentioned percentages (which are mostly based on raw or gross school and class effects) with caution. With regard to the percentages related to secondary education, one must bear in mind that part of these percentages is, perhaps more in secondary than in primary education, related to student recruitment and selection differences between schools (and/or classes within schools). In general, student recruitment/selection differences between secondary schools (and classes) can be quite large: depending on the outcome variable, they are sometimes responsible for about 50% or even somewhat more of the school (and/or class) variance in the outcomes (e.g. see Opdenakker, Van Damme, De Fraine, et al., 2002; De Fraine et al., 2003). An important explanation for this is, among other things, the difference in educational track offerings at Belgian secondary schools. Some schools offer only academic tracks, others only vocational tracks, while the rest offers both kind of track. In the Netherlands, the same trends as reported for Belgium are noticeable, although the percentages of variance attributable to classes and schools (i.e. above the student level) seem to be often somewhat lower than results from studies in Belgium (see e.g. Creemers, 2007).

[22] A similar difference between effects of schools, teachers and classes together on cognitive and non-cognitive outcomes is visible in Dutch studies.

[23] Some of the results are also cited in Van Landeghem et al. (2002).

Next to effects of schools, teachers and classes on cognitive and non-cognitive outcomes, educational positions in secondary education could be related to schools and to experiences in schools as well. For example, Pustjens, Van de gaer, Van Damme, and Onghena (2008) found that educational positions of students at the start of secondary education seemed to be quite related to the primary school the students went to: about 25% of the differences in positions could be related to primary schools. They also found evidence for a relationship between positions in secondary education and the secondary school a student attends: between 15% and 24% of the variance could be attributable to the secondary school. However, Luyten (2004) found smaller effects of primary schools in relation to track choice in the Netherlands. In his study only 11% of the variance in track choice could be related to primary schools. Van Batenburg and Lokman (1991) found large differences between schools with regard to passing or failing the first grade of senior vocational education in the Netherlands, and Timmermans, Bosker, de Wolf, Doolaard, and van der Werf (2014) also found large effects of schools on educational positions in secondary education.

In the preceding part, the described effects of schools, teachers and classes referred to effects measured at a particular moment in time, mostly referring to a measurement of outcomes at the end of a particular school year. However, together with further methodological development of methods and programs to study growth and development, some EER studies in Belgium and the Netherlands began to pay attention to *effects of schools and class experiences on growth* in outcomes as well. In general, empirical evidence is found for effects of schools and classes on the evolution over time in cognitive outcomes. For example, research of Dumay, Boonen, and Van Damme (2013) and Belfi, Haelermans, and De Fraine (2016) demonstrated effects of primary schools on growth in math achievement in Belgium, and Guldemond and Bosker (2009) found the same for growth in achievement in Dutch primary schools. In addition, Dumay et al. (2013) found the effect of schools on growth and on initial status of cognitive outcomes to be rather comparable (respectively 18% and 22% of the variance between schools). This finding is in line with the results of De Fraine, Van Damme, and Onghena (2007) in studying secondary school effects on status and growth in language and academic self-concept in Belgium. Furthermore, Vanlaar et al. (2014) investigated effects of Belgian primary classes on growth in reading comprehension and found evidence for effects of class practices, which is in line with findings on primary school class effects in the Netherlands on reading comprehension (Peetsma, van der Veen, Koopman, & van Schooten, 2006; Hornstra, van der Veen, Peetsma, & Volman, 2015a, 2015b) and language proficiency (van der Slik, Driessen, & De Bot, 2006. Meijnen et al. (2003), Hornstra et al. (2015b), and Peetsma et al. (2006) found clear effects of Dutch primary classes on growth in math achievement, and de Jager,

Jansen, and Reezigt (2005) found effects of (manipulated)[24] learning environments on metacognitive knowledge and metacognitive skills.

In addition, effects of schools and classes/teachers on the evolution over time in non-cognitive outcomes such as motivation and academic self-concept are found as well. Several studies investigated effects of secondary schools and classes on growth in these outcomes. Findings indicate that effects of schools on growth in non-cognitive outcomes range from 17% to 50%. In addition, there is some evidence that schools have more effect on growth than on status of the mentioned outcomes. For example, Van de gaer et al. (2009) found, in their study on effects of schools on growth in motivation towards learning tasks and academic self-concept in Belgium, that schools have strong effects on growth over time in both outcomes (respectively 17% and 50% of the linear growth was situated at school level), which were much higher than the school effects on status (which were respectively 9% and 8%). Research of Opdenakker, Maulana and colleagues (Maulana, Opdenakker, & Bosker, 2016; Opdenakker, 2014; Opdenakker & Maulana, 2010; Opdenakker, Maulana, & den Brok, 2012) found evidence for differences between secondary school classes (seventh grade) in the growth over time of motivational constructs such as academic engagement, autonomous and controlled motivation, test anxiety, self-efficacy and intrinsic value during one school year in the Netherlands. In addition, Opdenakker (2013) and Stroet, Opdenakker, and Minnaert (2016) investigated effects of secondary school prevocational classes (seventh grade, math and Dutch classes) in the Netherlands and found differences in developmental trends of motivational constructs such as intrinsic and identified motivation, intrinsic value and performance avoidance orientation. The study of Stroet et al. (2016) demonstrated also that differences between math classes were more pronounced than differences between Dutch language classes. Research of van der Werf, Opdenakker, and Kuyper (2008) found evidence for effects of secondary schools on growth in achievement motivation and educational positions during students' school careers in secondary education. In addition, a few studies addressed effects of primary school classes in the Netherlands on growth in motivation (e.g. Hornstra et al., 2015a, 2015b) and academic engagement (Opdenakker & Minnaert, 2011), finding significant effects. Peetsma et al. (2006) found a significant amount of variance at the class level in the growth over time of wellbeing at school (feeling comfortable at school, having a good relationship with teachers), however, they did not find evidence for class effects with regard to changes in academic self-concept (cognitive self-image) from Grades 4 to 6.

[24]In this study, three learning environment conditions were created by regular teachers in an intervention study in daily classes: a direct instruction, a cognitive apprenticeship, and a control group learning environment. The assignment of the teachers to a condition was based on voluntary participation and the teachers in the direct instruction and cognitive apprenticeship condition received a 15-h training and coaching session. In addition, they received exemplary lessons specifically designed to enhance the implementation of either a direct instruction or cognitive apprenticeship learning environment. Teachers in the control condition practiced no specific instructional model and received no training.

10.4.1.2 Long-Term Effects on Cognitive, Non-cognitive and School Career Indicators

In the preceding part, merely short-term effects of schools, classes and teachers on a particular moment in time or on growth within one year were addressed. However, from an educational effectiveness point of view, it is also interesting to know whether schools and the experiences students have had in their classes and with their teachers have long-term effects on students' cognitive and non-cognitive functioning related to school or school career. In the international EER literature, studies on this topic are quite scarce (Creemers et al., 2010). However, a few studies in Belgium and the Netherlands examined such effects of primary schools.

One of the first studies in Belgium that addressed this topic was research by Pustjens, Van de gaer, Van Damme, and Onghena (2004). They investigated, among other things, long-term effects of secondary education on success in higher education and found differences between secondary schools with regard to the extent to which students graduated from secondary education. In addition, Pustjens, Van de gaer, Van Damme, Onghena, and Van Landeghem (2007) investigated long-term effects of primary schools on cognitive outcomes (math and language) and demonstrated that, although primary schools have considerable short-term effects on math and language achievement, their long-term effects on cognitive outcomes seem to be rather small and decrease quickly over time. Also, Vanwynsberghe, Vanlaar, Van Damme, and De Fraine (2017c) studied long-term effects of primary schools on math achievement and found evidence for long-term (continuing) effects on math achievement at the age of 17. Unlike Pustjens et al. (2007), they followed students already in primary education and measured math several times during and at the end of primary education, while Pustjens et al. only measured cognitive outcomes at the start of secondary education. In addition, there is evidence that the sample of Vanwynsberghe et al. (2017c) is more heterogeneous and they studied a larger time period, namely 5.5 and 11.5 years compared to 1 and 2 years in the study of Pustjens et al. This may explain the larger long-term effects in the study of Vanwynsberghe et al. The findings of Vanwynsberghe et al. are in line with those of the study of Goldstein and Sammons (1997) in the UK: Goldstein and Sammons found evidence for substantial long-term effects of primary schools on cognitive outcomes at age 16.

Furthermore, there is some evidence that primary schools can have long-term effects on educational positions during secondary education. Luyten (2004) found a small continuing net effect of primary schools (0.3%) on career success 4 years after starting secondary education in the Netherlands above the differences already existing between students at the end of primary education. Also, Vanwynsberghe, Vanlaar, Van Damme, and De Fraine (2017b) found evidence for (small) long-term effects of primary schools on educational positions 2 and 4 years after the start of secondary education in Flanders. However, Pustjens et al. (2008) did not find long-term effects on the educational positions of students at the end of the second year of secondary education in Flanders.

There is also evidence of long-term effects of primary schools on non-cognitive outcomes. Research of Vanwynsberghe et al. (2017a) mentions small but significant long-term effects on non-cognitive outcomes such as mastery goal orientation,

students' social integration in the class and students' self-concept (in math), even at the age of 17. Also, long-term effects of particular school characteristics were found on autonomous and controlled motivation, performance avoidance orientation, and general and math self-concept. No long-term effects of primary schools or primary school characteristics were found with regard to wellbeing at school, performance approach orientation and interest in learning tasks. Finding some long-term effects of primary schools and school characteristics is not in agreement with a study in the UK of Sylva, Melhuish, Sammons, Siraj-Blatchford, and Taggart (2012), in which no long-term effects of (the academic effectiveness of) primary schools were found on social-behavioural outcomes (such as pro- and anti-social behaviour, self-regulation and hyperactivity) at the age of 14. It must be mentioned, however, that in this study by Sylva et al., other non-cognitive outcomes were studied than the ones in the study of Vanwynsberghe et al. Most likely, contextual differences between the UK respectively Belgium might be an explanation as well.

Overall, it seems that there is some evidence for long-term effects of primary schools, but effects seem rather small and often more visible when a long time period is taken into account. In addition, one must bear in mind that in the mentioned studies the teacher effect (of all teachers the students had during primary education) and often also the class effect (of all classes to student belonged to in primary education) were not taken into account. The study of Kyriakides and Creemers (2008) demonstrates that this underestimates the long-term school effect.

Lastly, a few studies of Luyten and colleagues (Luyten, 2004, 2006; Luyten, Tymms, & Jones, 2009) and of Verachtert, Van Damme, Onghena, and Ghesqiere (2009) found evidence for absolute effects of schooling[25] on math and differences in learning rates between schools.

10.4.2 The Importance of Schools Versus Learning Environments at Class Level (Teachers/Classes)

In general, there is clear-cut evidence in (multilevel) EER studies that take into account the level of schools and classes/teachers, that classes/teachers matter at least equally with regard to cognitive and non-cognitive outcomes compared to the school a student attends (Dumay, 2009; Opdenakker & Van Damme, 2000a, 2000b; Van Landeghem et al., 2002; Van Damme et al., 2004). Methodological work of Opdenakker and Van Damme (2000b) and Van den Noortgate, Opdenakker, and Onghena (2005) stresses that all relevant educational levels (e.g. school/teacher/class) should be included in multilevel models in order to, among other things, avoid

[25]In the study on the absolute effect of schooling, the effect of schooling versus no schooling is addressed. This is in contrast to the typical approach on the study of school effects in which the relative effects of schools are addressed: differences between schools. For more information on this approach, see Luyten (2006).

an overestimation of the variance attributable to levels included in the models and – in case a level above the student level is omitted (e.g. class level) – an underestimation of the overall effect of schools, teachers and classes together. This implies that studies that take into account only the school and student level in their multilevel models overestimate the variance at the school and student levels and underestimate the overall effect of going to a particular school and belonging to a particular class since the level of class and teacher is not included in the models. Similarly, Creemers and de Jong (2002) argued – and provided evidence for – the importance of classes in relation to (math) achievement.

10.4.3 Effects of School Factors on Students' Cognitive and Non-cognitive Outcomes

10.4.3.1 Context Factors

The picture that arises when looking at studies that address effects of *denomination* of schools in both countries is a mixed one. Some studies find positive effects of Catholic (and Protestant) schools compared to public schools, others find no effects of denomination, and even some find a positive effect of being a public school on student outcomes. Reasons have to do with, among other things, the time period the study was conducted, whether an adjustment for to student intake and group composition was made, and the amount and kind of adjustment (related to student intake and group composition characteristics) that was made.

For example, De Fraine et al. (2003) found that without adjusting for student background, students in Catholic secondary schools in Flanders performed better than students in public schools on language achievement in the eighth grade. However, when adjusting for a diversity of individual student characteristics, student recruitment and class composition, students in public schools performed best. Opdenakker, Van Damme, De Fraine, et al. (2002) did not find an additional effect of denomination of secondary schools on math achievement in the eighth grade after school process variables were included. At the end of the seventh grade, however, Opdenakker and Van Damme (2006a) found a small effect of denomination on math achievement that disappeared when student background was taken into account. Furthermore, they found that Catholic schools seemed to have a more favourable school composition (higher average cognitive ability and higher average educational level of the mothers of the students) in comparison to public schools and had a higher score on almost all school practice characteristics (teacher cooperation and participation in decision making, relational climate, study-orientedness of math classes, focus on individual student development). Furthermore, the differences in social climate between Catholic and public schools seemed to be most pronounced. The study findings are in agreement with the study of Elchardus and Kavadias (2000) and Brutsaert (1998): in these studies, the effect of denomination on non-cognitive outcomes also disappeared once school composition (related to SES) was taken into account.

Brutsaert (1998) studied also the effect of denomination in primary schools in Belgium on students' cognitive outcomes and found that Catholic schools influenced high achievement in low-SES students to a larger extend than public schools. Vanwynsberghe et al. (2017c), studying long-term effects of primary schools on math achievement, found that students from Catholic schools performed better at the end of secondary education than students from public schools. Agirdag, Driessen, and Merry (2017) investigated the effect of the denomination of primary schools (Catholic versus public schools) on math and reading achievement growth, and compared the learning growth of native Belgian and Muslim immigrant students within these two groups of schools. They found that Catholic schools outperformed public schools in math growth, but this difference disappeared once student background at student and at school level was taken into account. They did not find a differential effect of denomination related to the country of origin of the students. This implies that they did not found support for the Catholic school advantage hypothesis, nor the common school effect hypothesis when they took student background at student and school level into account.

Effects of denomination were also investigated in the Netherlands. Differences between private and public schools with regard to cognitive outcomes were found by Dronkers and Robert (2003), Hofman, Hofman, and Guldemond (2002), Meelissen and Luyten (2011), and Hofman, Hofman, and Gray (2015). Hofman et al. (2002) found that the higher mathematics achievement of Catholic schools can be largely explained by the educational culture and governance structure of these schools (e.g. parents have more influence on the policy of the school board). The coherence between governance, school leader, teachers and school community seemed to produce a sense of community that shaped conditions in these schools, which in turn had a positive effect on students' achievement. Hofman et al. (2015) also found an effect of denomination on math achievement. In addition, Dronkers (2004) mentions somewhat more pronounced differences in achievement between private and public schools in primary education compared to differences between private and public schools in secondary education, suggesting some evidence for a stronger denomination effect in primary education compared to secondary education. Driessen and van der Slik (2001) did not find an effect of denomination of Dutch primary schools on cognitive and non-cognitive outcomes (well-being, self-confidence) of kindergartners, which is in line with a study of Avram and Dronkers (2011) in secondary education. Avram and Dronkers did not find such an effect on the degree to which students felt integrated in secondary schools in the Netherlands. However, they found that students felt better integrated in publicly supported private schools in Belgium (which are mostly Catholic schools). School climate is one of the explanations researchers often give for the differences in effectiveness between public and private, often religious schools, which is in line with the already mentioned findings of Opdenakker and Van Damme (2006a). Also, the finding of ten Bruggencate, Luyten, Scheerens, and Sleegers (2012) that in Catholic schools school leaders show more human relations behaviour (in addition to rational goals behav-

iour)[26] can be seen as an outcome pointing in the same direction. In addition, Bryk, Lee, and Holland (1993) mention in their comprehensive analysis of the effects of US Catholic schools that Catholic schools are more often characterized by both an academic and communitarian climate. Roeleveld and Dronkers (1994) found that secondary schools of a denomination with either a clear minority or majority position in the region are less effective than schools that enroll about 40% of the students in the region. Driessen, Agirdag, and Merry (2016) found that students from Catholic (and Protestant) primary schools outperformed students from public and Islamic schools on reading in Grade 5. After adjusting for student background (parental education and parental ethnicity), only students of Catholic schools still outperformed students of public schools. However, when an additional adjustment for student background at school level (parental education and parental ethnicity) was taken into account, the effect of Catholic schools also disappeared. In addition, they found with regard to a variety of cognitive and non-cognitive outcomes of students in Grade 2, 5 and 8 no evidence of a denomination effect once they took the already mentioned input differences at student and school level into account, suggesting that perhaps the apparent effects of school denomination may relate instead to student intake composition in terms of family background factors.

Opdenakker and Van Damme (2006a) also investigated the effect of *school type,* referring to the number of grades and the study programs the school offers, on math achievement. Controlling for student background, results revealed a significant effect of school type in favour of the six-grades multitrack school (compared to unitrack and middle school types). Their study also revealed that these kind of school types not only differed with regard to school composition, but also with regard to processes within these schools. Timmermans et al. (2014) investigated the effect of school type on educational positions in secondary education in the Netherlands, and found differential effects of school type (based on the kind of tracks provided in the school) in relation to SES and prior achievement of students.

A few studies also investigated effects of *school size*. Most of them were conducted in secondary education (e.g. Luyten, 1994; Opdenakker & Van Damme, 2006a, 2007; Opdenakker, 2004; ten Bruggencate et al., 2012; Hofman, de Boom, Meeuwisse, & Hofman, 2013; Kruger, Witziers, & Sleegers, 2007), one in senior vocational education (van Batenburg & Lokman, 1991) while a few studies in primary education used school size as control variable (Belfi, Gielen, De Fraine, Verschueren, & Meredith, 2015; Belfi et al., 2016). In general, there is only little evidence of the importance of school size to achievement (Luyten, 1994). However, there is some indication that school size matters with regard to school processes. For example, Opdenakker and Van Damme (2006a, 2007) found evidence of a significant positive direct effect of school size on teacher collaboration in secondary

[26]Human relations behaviour of the school leader refers in this study to behaviour with an emphasis on participation and collaboration, conflict management, and consensus building. Rational goals behaviour refers to behaviour with a focus on the clarification of goals, rational analysis, and decisive action.

schools, and in turn an indirect effect of climate characteristics in schools, which resulted in better math achievement.[27] However, the study of Hofman et al. (2013) suggests an advantage of rather small (innovative) schools. Also, ten Bruggencate et al. (2012) found effects of school size (and other school context variables) on several school (process) characteristics. With regard to school size, they found a moderately positive effect on open-systems school leader behaviour and on development orientation (with regard to culture and practice of the school), a small negative effect on human relations school leader behaviour, and a small positive influence on performance orientation (with regard to culture and practice of the school) and on teachers' work (referring to interactions with students, student support, the working climate in classes and the organization of subjects and lessons).[28]

10.4.3.2 School Management and Organization

Several studies paid attention to the effects of *management and leadership* of schools. In general, effects are non-existent or rather small and often only indirect on student outcomes.

For example, van de Grift (1987) did not find an effect of school leadership on achievement in primary education in the Netherlands, but later, when educational leadership had grown as a field, he did (van de Grift & Houtveen, 1999). Witziers and Bosker (1997) did not find effects of leadership on achievement as well, nor did Kruger et al. (2007) find direct or indirect effects of school leadership on student commitment in secondary schools in the Netherlands. However, their study found evidence of reciprocal relationships between strategic leadership and student commitment. In addition, Hofman, Hofman, and Guldemond (2001) found effects of leadership styles on achievement in secondary education as well as Brandsma and Stoel (1987). Brandsma and Stoel also found an effect on students' well-being in secondary schools. Hofman and Hofman (2011) also studied effects of management in secondary education in the Netherlands. Their study made clear that the managerial capacities of schools were of importance for the effectiveness of schools, in particular, when the schools' position got under pressure at the local (student) market. In addition, the study of Hofman et al. (2015) indicated the importance of primary school characteristics related to school governance, policy on evaluation and monitoring of student performance, cooperation between teachers and clarity of rules with regard to math achievement.

Opdenakker and Van Damme (2006a) studied the effects of secondary school practice (organization and management, school relational and learning climate, and

[27] One has to bear in mind that the largest schools in Flanders and the Netherlands refer to a medium size school in American literature.

[28] Open-systems school leader behaviour refers to behaviour that is focused on political adjustment, creative problem solving, innovation, and change management. An open-systems focused school leader emphasizes flexibility and change and embraces ongoing adjustment to the external world in order to acquire and maintain external resources.

opportunity to learn mathematics at school) in Flanders. Evidence was found for positive effects of cooperation between teachers, relational and learning climate and opportunity to learn mathematics. However, no effect was found for school leadership (referring to teacher participation in decision-making and to the number of educational tasks the school leader is involved with). When all significant school variables, including school composition, were included in one model, only school composition and learning climate remained significant, indicating that there were joint effects of the mentioned school characteristics. Also, Opdenakker and Van Damme (2007) investigating the relationships between school context, school leadership (participative, educational/professionally oriented relationship), school composition and school practice and outcomes in secondary education in Flanders, did not find an effect of school leadership, neither on school practices nor on student outcomes. A possible explanation could be the rather limited existence of such a leadership in secondary education in Flanders at the time of the investigation. More recently, Dumay and Galand (2012) studied effects of transformational leadership in primary education on teachers' collective efficacy and organizational commitment and found evidence for small positive effects. However, ten Bruggencate et al. (2012) found small effects of school leadership on the mean promotion rate of schools. The effect was mediated by a development-oriented school organization and favourable class practices. However, they did not find a direct effect on achievement.

The meta-analysis by Witziers, Bosker, and Kruger (2003) gives an explanation for inconsistent findings with regard to the effect of school leadership: effects of leadership tend to be smaller when school context and school composition or teacher variables are included in the model. This is in line with findings of Dumay et al. (2013), who did not find direct effects of leadership on growth in math achievement. However, they did find effects of leadership on teacher collaboration and teachers' collective efficacy. Research of Opdenakker and Van Damme (2007) revealed that teacher collaboration has a positive effect on school climate, which is positively related to student outcomes such as math achievement. In addition, the study of De Maeyer, Rymenans, Van Petegem, van den Bergh, and Rijlaarsdam (2007), investigating the relationship between educational leadership (integrated leadership) and math and language achievement in primary education (grade 4 and 6) in Flanders, revealed that whether integrated leadership had an effect on achievement was dependent on the choice of the conceptual model underlying the analysis. Taken together, these finding indicate that effects of leadership should be investigated with complex models[29] taking into account school context and school composition (thus taking into account the context-specificity of the setting) and focusing on effects it has on school practice and in particular teacher collaboration, collective efficacy and school climate. Since, findings indicate that what is conceived as good leadership qualities in EER seems to operate in the reality of education in an indirect way by

[29]In the study of De Maeyer et al. (2007) evidence was found for the importance of an antecedent model which takes into account school context/composition and assumes the existence of direct and indirect effects of school leadership on student outcomes.

means of its effect on teachers (and climate) to enhance the educational effectiveness of schools, in a direct way on student outcomes, and is also influenced by school context/composition. However, as De Maeyer et al. (2007) states: the use of such models needs to be accompanied by well-grounded theory.

10.4.3.3 School Processes

Several studies investigated effects of school characteristics, referring to school processes, in relation to student outcomes. For example, van der Werf (1997) examined differences between low, average and high-achieving primary schools and looked, among other things, at differences in instructional characteristics. She found that in highly effective schools for math more time was spent on learning and evaluation, more whole-class instruction was provided and more common goals for all students were formulated.

De Bilde, Van Damme, Lamote, and De Fraine (2013) investigated the effects of alternative education (Waldorf and Freinet) compared to traditional education on students' school engagement from the last grade in kindergarten to their third grade in primary education. They did not find evidence for a positive effect of alternative education on school engagement (school enjoyment and independent participation rated by the teacher). However, they discovered differential effects on both outcomes with regard to initial language achievement, indicating that only in traditional schools, initial language achievement is relevant for school engagement. No differential effects were found with regard to socioeconomic status (SES). Stroet et al. (2016) studied effects of secondary schools with a traditional, a social-constructivist and a combined approach on (growth in) motivational aspects (intrinsic and identified motivation, intrinsic value, performance avoidance) and also found no differences between traditional and social-constructivist schools. However, they found that in schools with a combined approach (which were often schools in transition) students scored significantly lower on intrinsic, identified and intrinsic value. De Bilde et al. (2015) investigated effects of kindergartens' experiential practices, which are related to a child-centered educational framework that is popular in kindergarten in Flanders, on school adjustment. They looked at five practices: autonomy support, stimulation, emotional support, time of choice activities and interest-based activities). They found that not all experiential practices were related to a positive school adjustment. A stimulating teaching style was related to greater gains in cognitive achievement, but an autonomy supportive style was related to smaller gains, especially among low-achieving children. Heers, Ghysels, Groot, and Maassen van den Brink (2015) studied the effectiveness of community schools for care students (i.e. students with additional educational needs) and students whose parents have low educational attainment. Their study showed that both types of students benefited from going to a community school in terms of a reduced underachievement. The duration of attendance in such schools mattered when the community school subsidy was taken into account.

Opdenakker and Van Damme (2000a) studied effects of secondary school characteristics related to the support of (cognitive) development/achievement as well as non-cognitive development on cognitive and non-cognitive outcomes in Flemish secondary education. Two characteristics also refer to supportive conditions for cognitive development, but also to school climate namely 'teaching staff cooperation in relation to teaching methods and student counselling' (which is an operationalisation of cooperation, consensus and cohesion among staff) and 'orderly learning environment' (which is an operationalisation of the school climate indicator orderly atmosphere). Results indicated that school characteristics related to the supportive conditions for cognitive development had positive effects on cognitive outcomes, but school characteristics related to non-cognitive aspects of education had no effect on cognitive outcomes (with the exception of a focus on education and personality development which seemed to have a negative effect on math achievement). In addition, supportive conditions for cognitive development as well as characteristics related to non-cognitive aspects of education seemed to have an effect on non-cognitive outcomes. An overwhelming positive effect of teaching staff cooperation was visible on language achievement and on almost all non-cognitive outcomes.

In addition, the study revealed that high intelligent students (and for some non-cognitive outcomes also initial high achievement motivated students) seemed to benefit the most from the school characteristic teaching staff cooperation. Furthermore, a clear positive effect of a high orderly learning environment was found on both cognitive outcomes as well on non-cognitive outcomes. Again, evidence for some differential effects related to intelligence and motivation were found. Attention to student differences and development seems to have a negative effect for high intelligent students with regard to their cognitive outcomes, and for all students with regard to their attentiveness in the classroom and attitude towards homework. In addition, it has a clear negative effect on the attitude towards homework and social integration of initial low achievement motivated students. No evidence was found for the effectiveness of a strong focus on discipline and subject matter acquisition, but it seemed to have a positive effect on the well-being of initially high achievement motivated students and a negative effect on the well-being of initially low achievement motivated students. A focus on non-cognitive aspects of education like a focus on social and moral education and personality development seems to have a positive effect on the interest in and motivation towards learning tasks of all students and on the attitude towards homework of low achievement motivated students. However, it also has a negative effect on the attitude towards homework of initially high achievement motivated students and on the mathematics achievement of all students. A focus on cultural education and creativity seems to be beneficial for some non-cognitive outcomes (well-being at school, attentiveness in the classroom, interest in learning tasks and social integration in the class) of initially high achievement motivated students, but negative for the mentioned outcomes of low achievement motivated students. In general, these findings seem to suggest that cognitive and non-cognitive outcomes are to be seen as rather distinctive outcomes, and that schools effective for cognitive outcomes, are not necessarily effective with regard

to non-cognitive outcomes. This is in line with research of Knuver and Brandsma (1993) who found that there is only a small positive correlation between the relative positions of schools with regard to cognitive and non-cognitive outcomes and research of Hofman, Hofman, and Guldemond (1999) who stressed the importance of climate factors.

Opdenakker and Van Damme (2001) also studied the effectiveness of school practice characteristics and found also evidence for some differential effects. They indicated a higher sensitivity to school processes of highly able students. Differential effects were found for teacher cooperation and orderly environment. Teacher cooperation seemed beneficial for highly able students, but negative for less able students, while an orderly school environment seems to be most beneficial for highly able students. Furthermore, it was found that in particular, highly-able students from low-SES families, were most sensitive. In addition, it was found that schools with a heterogeneous ability composition can enhance the achievement of their students when they pay a lot of attention to student differences and development.

Opdenakker and Van Damme (2007) studied the effectiveness of secondary school practice indicators as well. They tested a theoretical model (based upon Opdenakker, 2004) and their study confirmed the hypothesized relationships to a large extent. Among other things, they found evidence for positive effects of teacher collaboration, indicators of school climate (relational and learning climate) and opportunity to learn math on math achievement (and effort for learning). Effects of teacher collaboration were mediated by school climate. Climate had, although also related to school composition, its own positive effect on math. In addition, opportunity to learn math had its own positive influence on math as well. "In schools were teachers teaching the same subject as well as teachers teaching the same class frequently collaborate, talk and make agreements about teaching, methods, curriculum and student counselling, and where students as a group feel well integrated in their classes and experience their teachers as warm, fair and understanding, students are willing to invest effort for learning and perform better" (Opdenakker & Van Damme, 2007, p. 198). The study indicated that in order to advance our understanding of educational effectiveness, the complex relationships between school composition, school context, school functioning, teaching and learning should be investigated in concert. Lomos, Hofman, and Bosker (2011) found evidence that school departments within secondary schools that focus on reflective dialogue, collaboration, a shared vision and student achievement are successful schools which have higher achievement.

Studies addressing effects of school climate on cognitive and non-cognitive functioning often found significant effects. For example, Opdenakker and Van Damme (2001) and Dumay and Dupriez (2007) found such an effect on the math achievement of secondary school students in Belgium and the Netherlands. Van Batenburg and Lokman (1991) found that school characteristics related to student monitoring and the availability of school resources were significantly related to differences between senior vocational schools with regard to passing or non-passing in the first year of senior vocational education. Furthermore, the study of Meelissen and Luyten (2011) found evidence for the importance of a positive

climate in primary schools (and classes) for raising science achievement. Keuning, van Geel, Visscher, and Fox (2016) found positive effects of differentiated instruction on math achievement and spelling in primary education and found that this was in particular effective in low-SES schools.

10.4.3.4 School Composition

Effects of school composition were found as well, although not in every study. Maslowski, Scheerens, and Luyten (2007) found effects of school composition characteristics such as mean SES[30]) on reading literacy of students in secondary schools in OECD countries (including the Netherlands, Flanders and the French-speaking part of Belgium). Also other studies done in the Netherlands and the Flemish and French-speaking part of Belgium separately, found evidence for school composition effects on student outcomes. In the next sections, findings of studies in secondary as well as in primary education addressing effects of school composition on a diversity of student outcomes will be discussed. For the sake of clarity, studies will be discussed for both countries separately, starting firstly with secondary education.

School Composition in the Netherlands

Hofman (1995) was among the first to pay attention to effects of school composition in studying the school careers of students and cross-relationships with schools in Dutch *secondary schools*. He did not find that the average cognitive level of the school population or the proportion of minority students at school had an effect on the school career of students, neither in general nor respectively on the careers of higher or lower ability students nor on the careers of ethnic majority or minority students. However, low school average SES was positively related to the school careers of low-SES students. On the contrary, Driessen (2002) found positive effects of a high-SES composition and a negative effect of a high proportion of ethnic minority students on math achievement in Dutch *primary education*. Also, Meelissen and Luyten (2011) found an effect of primary school SES composition on science achievement. Weide (1993) found small differences in educational quality between primary schools of different ethnic composition, with larger differences for language compared to math. In schools with a high proportion of ethnic minority students, the educational quality was lower due to lower educational provisions. However, Guldemond and Bosker (2009), investigating differences between primary schools in achievement growth, did not find an effect of school composition related to students' background on growth. Their explanation is the extra funding for staff for schools with a disproportional number of students from low-SES families,

[30]In addition, this effect seemed to go together with positive effects of autonomy of personnel management on reading literacy.

especially if these students are also from ethnic minorities. Also, Karssen, van der Veen, and Volman (2016) did not find an effect of ethnic diversity at primary schools on educational outcomes, social-emotional functioning and citizenship competences. However, they did find evidence for differential effects indicating that mono-ethnic majority students (i.e. students of which both parents are of Dutch origin) scored lower on reading comprehension and wellbeing and higher on citizenship orientations in ethnic diverse schools compared to the same type of students in less ethnically diverse schools.

School Composition in the Flemish and French-Speaking Part of Belgium

Opdenakker, Van Damme and colleagues were among the first to study effects of school composition in *Flanders*. For example, Opdenakker, Van Damme, De Fraine, et al. (2002) and De Fraine et al. (2003) studied the effects of school composition on math and language achievement, respectively, in *secondary education* in Flanders. They investigated the effects of a variety of school composition indicators (e.g. mean cognitive ability, mean SES, mean achievement motivation, mean immunity to stress, heterogeneity in the mentioned variables, proportion of girls,[31] proportion of students speaking Dutch at home) and school process factors (in addition to class composition and class process factors). They found important student recruitment differences between schools (and classes) in Flanders, which is not quite surprising since there is already some tracking in the first grades in secondary education in Flanders and schools differ from each other with regard to the tracks and the number of tracks they offer. However, they also found that school composition (level and heterogeneity) did not explain additional variance in math and language achievement respectively once class composition was taken into account. An exception was found for the proportion of girls at school, which had a positive effect on math achievement. Positive effects of mean SES and mean ability were also found by Opdenakker and Van Damme (2001) on math achievement in secondary education in Flanders. However, the effect of mean SES was a joint effect with mean ability level. In addition, their study revealed a differential effect of mean ability in relation to students' individual ability: a high average ability level of the school seems to be most beneficial for highly able students. Furthermore, it was found that in particular, talented students from low-SES families were most sensitive to the effect of mean ability.

Van Damme et al. (2004) paid attention to the effects of school composition characteristics on students' attitudes towards math in Flemish secondary education. They found a negative effect of school composition referring to the amount of

[31] Some Belgian studies in the second (or beginning of the third) decade of EER paid attention to the gender composition of schools or classes (i.e. the proportion of girls). At that time, schools and classes could consist of 100% of girls, since Flemish schools could be only for girls, only for boys or mixed-sex schools.

problem behaviour at school. In addition, their study revealed a differential effect of the gender composition of schools indicating that in equally-mixed schools, a larger effect of gender was found compared to other gender composition school types (indicating that girls have more negative attitudes and boys more positive attitudes in equally-mixed schools). In addition, a differential effect was found in relation to students' comfort at home as well.

A few studies (e.g. Brutsaert, 2006; Brutsaert & Van Houtte, 2002) investigated the effect of the gender composition of secondary schools (single-sex schools versus coeducational schools) on school well-being (sense of belonging, peer group acceptance). In general, these studies found differential effects of schools' gender composition, indicating that female students experienced a higher sense of belonging and felt more socially accepted by their peers in single-sex schools compared to coeducational schools. Boys, on the other hand, experienced an equal level of school belonging in both school types, but felt more accepted by their peers in coeducational schools.

Pinxten, De Fraine, Van Den Noortgate, Van Damme, and Anumendem (2012) studied the effect of school composition (with regard to gender, math, and SES) on the curriculum choices (option choice) in the academic track in secondary education, namely in the transition from Grade 8 to Grade 9, and from Grade 10 to Grade 11. They found hardly any effects of school composition; the curriculum choices in the academic track were mainly determined by individual student characteristics.

Agirdag, Van Houtte, and Van Avermaet (2012) studied school composition effects in Flemish *primary schools* and found a positive effect of SES composition on math achievement. Boonen, Speybroeck, et al. (2014) also studied effects of primary school composition (mean prior achievement, SES, and ethnic and gender composition) in Flanders, related to math (in the second grade). They did not find direct effects, but they found a few small indirect effects of school composition on math achievement referring to a positive relationship between a favourable school composition (high average achievement, high mean SES) and regular contact between the school and parents, which was beneficial for math achievement. In addition, they found evidence for a few small differential effects: the average achievement of schools positively affected initially-high achievers (which is in line with findings of Opdenakker & Van Damme, 2001) and the proportion of ethnic minority students negatively affected students not speaking Dutch at home (with the exception of students speaking Turkish, Arabic or Berber at home). The study seems to indicate that school composition hardly matters in the early years of primary education. However, one must bear in mind that the Flemish government enhanced the Act of Equal Opportunities in September 2002, which provides schools with high proportions of disadvantaged students with, among other things, additional teaching periods. Boonen, Pinxten, Van Damme, and Onghena (2014) also investigated effects of mean school SES and mean prior achievement and found effects on math and reading comprehension in primary education. However, they also found that much of these effects are mediated by academic optimism, a latent construct referring to academic emphasis, collective efficacy and faculty trust.

Furthermore, Belfi et al. (2014) studied effects of SES and ethnic school composition in primary education in relation to several aspects of language education

(reading comprehension, fluency and spelling) and found effects on the level, but not on the growth in language achievement. Belfi et al. (2016) investigated long-term differential effects of primary school SES composition om math achievement growth in relation to different SES groups and found clear effects indicating more positive growth in math in high-SES schools. In addition, the lesser growth in mixed-SES schools was more pronounced for high-SES students.

Vanwynsberghe et al. (2017a) investigated long-term effects of Flemish primary schools' composition. They found that a high proportion of high-risk students in schools had a positive effect on non-cognitive outcomes such as autonomous motivation and general self-concept and a negative effect on controlled motivation, which is often seen as a low-quality motivation (Ryan & Deci, 2000; Vansteenkiste, Lens, & Deci, 2006; Vansteenkiste, Soenens, Sierens, Luyckx, and Lens (2009). Vanwynsberghe et al. (2017b) investigated long-term effects of primary schools on math achievement but did not find an effect of this primary school composition characteristic on math achievement at the end of secondary education.

Dumay and Dupriez (2008) investigated relationships between a variety of school composition characteristics and language achievement in *primary education* in the *French-speaking part* of Belgium and found small effects of all composition characteristics except for gender composition. School composition related to the sociocultural background of the students was most important. Also, Dumay et al. (2010) found evidence for effects of school composition (related to academic achievement and sociocultural background of the students) on language achievement in primary education in the French-speaking part of Belgium. In addition, they also found evidence for effects of the proportion of students speaking French at home. However, they did not find an effect for the proportion of girls at school.

10.4.3.5 Relationships Between School Composition and School Processes

Several researchers also paid attention to the relationship between school composition and school process characteristics. For example, Opdenakker, Van Damme, De Fraine, et al. (2002) and De Fraine et al. (2003) found evidence for relationships between school composition and a variety of school process characteristics related to school leadership and general quality of class processes. Also, Opdenakker and Van Damme (2001, 2006b, 2007), Opdenakker, Van Damme, and Minnaert (2002a, 2002b, 2005), De Fraine et al. (2003), and Belfi et al. (2015) paid attention to relationships between school composition (e.g. average ability and SES) and school process variables (e.g. teacher cooperation, school climate, orderly environment, attention to student differences, focus on discipline and subject matter acquisition) and found relationships with school process variables (specifically, orderly learning environment and cooperation between teachers). Opdenakker and Van Damme (2001) explored the relationships between five school composition characteristics (mean numerical intelligence, heterogeneity of numerical intelligence, mean SES, proportion of girls, and proportion of students speaking Dutch at home) and six school process characteristics (based on the reports of a representative sample of

teachers of the school), namely teaching staff cooperation (with a focus on teaching methods and student counselling), focus on discipline and subject matter acquisition (which refer to a more traditional style of teaching and vision on learning and has also to do with a strong achievement orientation), orderly learning environment (which refers to the school climate indicator orderly atmosphere), attention to student differences and development, and two characteristics referring to non-cognitive aspects of education namely focus on cultural education and creativity and focus on moral and social education and personality development. Results indicated significant relationships between school composition and school process characteristics (medium sized correlations), for example between mean ability and mean SES, and teacher cooperation and orderly environment, as well as large joint effects on math achievement. In addition, Belfi et al. (2015) found a relationship between SES composition and collective teacher efficacy, which could be explained by teacher staff perception of school-based social capital. The effect remained even when average achievement of the students and ethnic composition of the school were taken into account.

In addition, Opdenakker and Van Damme (2006a) found evidence for a positive relationship between school composition and school practice and for relationships between school context (e.g. denomination, school type), school composition and school practice. However, Dumay and Dupriez (2008) found only weak associations between school composition and school organization processes.

Several studies of Opdenakker and colleagues (e.g. Opdenakker & Van Damme, 2001; Opdenakker, Van Damme, De Fraine, et al., 2002) revealed that, although medium to strong relations were found between school composition and school processes in secondary schools in Flanders, significant effects of school processes on student outcomes remained even when controlling for school composition.

10.4.4 Effects of Learning Environment Factors at Class Level: Teacher Behaviour and Students' Class Experiences on Students' Cognitive and Non-cognitive Outcomes

With regard to class and teacher factors, indicators of (the quality of) instruction in the learning environment are far more often investigated than indicators of the quality of the learning and relational climate in classes. In addition, effects of class climate are more often studied in secondary schools than in primary schools in both countries.

10.4.4.1 Effects of Instruction and Instructional Support

A lot of EER studies in Belgium and the Netherlands investigated the effects of classroom instructional practices, and in particular the quality of instruction and

instructional support. In addition, factors related to instruction – such as opportunity to learn and amount of instructional time – were also investigated, as was classroom management (which can be seen as a tool to organize and structure the learning environment).

Evidence was found for the effectiveness of opportunity to learn (math secondary education: e.g. Opdenakker, Van Damme, De Fraine, et al., 2002; Dumay & Dupriez, 2007), amount of instructional time (reading primary education: Boonen, Van Damme, & Onghena, 2014; math primary education: Meijnen et al., 2003), time on task (math primary education: Vause et al., 2010), and class practices and teacher behaviour that refer to structured, well-organized and supportive instruction or learning environments. Evidence was found for the effectiveness of well-organized, structured and attractive instruction on the level and growth of reading comprehension (primary education: Vanlaar et al., 2014). Clarity of instruction, guidance and structure support in the sense of the provision of optimal challenging tasks, adequate help, encouragement after failure, praise, and communication of clear guidelines and expectations can be seen as aspects of this kind of approach as well. Research of Opdenakker (2013, 2014) has demonstrated that structure support is, among other things, of importance for students' psychological basic need satisfaction, in particular the need to feel competent. Competence is widely seen as a core element in motivated actions (Bandura, 1989; Harter, 2012) and research has also demonstrated its importance to growth in academic engagement (Opdenakker & Maulana, 2010; Opdenakker & Minnaert, 2014). According to self-determination theory (Ryan & Deci, 2017), the fulfillment of students' psychological basic needs (feeling competent, related and autonomous) are important for positive development and growth, effective functioning, and wellbeing at school. Evidence for the effectiveness of these aspects of the learning environment can be found in studies on preschool and kindergarten arrangements (math, language: e.g. de Haan, Elbers, Hoofs, & Leseman, 2013; Vause et al., 2010), primary schools (metacognitive knowledge and skills: de Jager et al., 2005; motivational aspects: Opdenakker, 2014; Opdenakker & Minnaert, 2011), secondary schools (motivational aspects and academic engagement: Lietaert, Roorda, Laevers, Verschueren, & De Fraine, 2015; Maulana et al., 2016; Opdenakker, 2013, 2014; Stroet et al., 2015) and even in studies of learning and working in groups in vocational education (situational interest: Minnaert, Boekaerts, de Brabander, & Opdenakker, 2011). Denies et al. (2016) found positive effects of clear communication of expectations on writing in French (second language) in secondary education. There is some evidence that a structured approach is a bit more beneficial for students at risk, e.g. students with a Moroccan/Turkish ethnic background[32] (Meijnen et al., 2003: reading

[32]There are several reasons why students with a Moroccan or Turkish ethnic background are at risk. First of all, the language spoken in the school is often not the language spoken at home. In addition, the socioeconomical position (educational level, income level, occupational status) of the parents is often lower than that of Dutch or Belgian classmates. Furthermore, the achievement level of students with a Moroccan or Turkish ethnic background lag a long way behind even at an early age in primary education (e.g. at the age of 5/6 year) (see e.g., Herweijer, 2009).

comprehension and math primary education) or low-SES (Meijnen et al., 2003: math primary education).

With regard to classroom management, evidence is found of an effect on cognitive and non-cognitive outcomes, although this effect is often less pronounced than the effect of the previously-discussed characteristics of learning environments. In addition, the quality of classroom management often goes (to a certain extent) hand in hand with structured and clear instruction (see e.g. research of Opdenakker & Minnaert, 2011; Maulana et al., 2016). Effects of classroom management were found with regard to language and academic engagement as well as level and growth in reading comprehension in primary education (Vanlaar et al., 2014; engagement: Opdenakker & Minnaert, 2011), and motivational aspects such as intrinsic value, text anxiety and self-efficacy in secondary education (Maulana et al., 2016). Vause et al. (2010) found a positive effect of the climate of discipline in classes on math achievement in primary schools.

In addition, teaching learning strategies seemed to be important for realizing growth in reading comprehension in primary education (Vanlaar et al., 2014). In line with this, Droop, van Elsäcker, Voeten, and Verhoeven (2016) found positive effects of sustained strategic reading instruction on knowledge of reading strategies and reading comprehension of primary school students. Vause et al. (2010) found a positive effect of teachers' attention to meta-cognition on math achievement of primary school students.

Some studies, however, did not find effects of class practices. For example, De Smedt et al. (2016) did not find an effect of teaching writing skills and teaching writing strategies on fifth- and sixth-grade primary education students' informational and narrative text quality. A possible explanation might be the low occurrence of writing assignments and writing instruction in Flemish daily class practice. Indeed, when explicit instruction of writing was manipulated in an intervention study in primary education (De Smedt & Van Keer, 2018), it seemed to cause positive effects on students' writing quality: students in the intervention performed better than students who got writing instruction and activities as usual in class.

In contrast to approaches referring to structured instruction and support, class practices that often go along with less structured learning environments such as discovery learning do not always indicate positive effects. With regard to discovery learning, Vanlaar et al. (2014) found negative effects on the level of reading comprehension of high-risk primary school students and no effects on their growth, however, positive effects were found on the growth of low-risk students and no effects for high-risk students. With regard to cooperative learning, negative effects were found for high-risk students and no effects for low-risk students. However, positive effects were found on the growth in reading comprehension of both types of students.

Furthermore, Hornstra et al. (2015b) investigated effects of innovative learning in primary school classes (authentic learning, collaborative learning, focus on self-regulation) on math, reading comprehension and motivational aspects (task orientation, academic self-efficacy, school investment) and found mixed results on the development in the mentioned outcomes and also evidence for differential effects

in relation to gender and student background. Meijnen et al. (2003) did not find evidence for differences in math development between primary school students (first grade) who received math instruction with a traditional approach (with a lot of direct instruction and a lot of rote learning) and those who received instruction with an innovative approach ('realistic' math instruction with a focus on contextual learning). However, their research demonstrated that when teachers used an approach between traditional and innovative, students' growth in math was lower than when the teacher used a traditional or an innovative approach alone. Denies et al. (2016) found positive effects of the use of authentic learning material in French second language classes for French writing quality.

Besides, research on instructional approaches and teacher behaviour directed at supporting the autonomy of students (e.g. by offering choice, trying to avoid controlling language and coercive behaviour, showing respect by acknowledging students' opinions, ideas, feelings etc., and by providing meaningful rationales for doing learning activities) has demonstrated positive effects on students' academic engagement in secondary education (Lietaert et al., 2015; Opdenakker & Maulana, 2010; Opdenakker, 2014).

Also, with regard to approaches more in line with (social-)constructivist theories, characterized by shared teacher-student control or loose teacher control (i.e. more student control) some evidence has been found for positive effects on motivational aspects such as intrinsic value, test anxiety and self-efficacy in secondary education (Maulana et al., 2016). Van Damme et al. (2004) found effects of the degree to which the class experienced the teaching of mathematics as constructivist on attitudes towards math. In addition, positive effects were found for the stimulation of active learning on academic engagement in primary education (Opdenakker & Minnaert, 2011). De Jager et al. (2005) found evidence for positive effects of using cognitive apprenticeship as instruction method on metacognitive knowledge and skills of primary school students.

Some of these characteristics, mainly referring to constructivist practices, seemed to have negative effects on high-risk (mainly low-SES) students' reading comprehension e.g. discovery learning and cooperative learning; well-organized instruction had a less positive effect and differentiation no effect; discovery learning had no effect on the growth in reading comprehension of high-risk students, while more traditional practices have positive effects on all students, although sometimes to a lesser extent on high-risk students (Vanlaar et al., 2014).

With regard to differentiated instruction, which can be seen as a means to accommodate to students' differential needs, results are mixed. For example, Vanlaar et al. (2014) found that differentiated instruction was only effective for the level and growth of reading comprehension of low-risk students in primary education; for high-risk students differentiated instruction seemed to have only an effect on their growth in reading comprehension, and the effect was also smaller for high-risk than for low-risk students. De Fraine et al. (2003) studied effects of class practices (teacher perception) – the effects, among others, of differentiated instruction and attention to individual differences on language achievement in secondary education – and did not find effects of a focus on individual development, special

attention to low or high achievers, or giving individual feedback to students on their academic results. However, Opdenakker and Minnaert (2011) found positive effects of teachers who stimulated active learning and made use of differentiation during class time on students' academic engagement at the end of primary education.

With regard to mastery learning, which can be seen as a form of adaptive learning and instruction, results are mixed. For example, Reezigt and Weide (1990) found only a few effects of mastery learning in relation to achievement (language, math) with a positive effect on language achievement in Grade 6 and a negative effect on the growth from Grades 5 to 6 in Dutch primary education.

Almost no studies investigated the effects of the opposite dimensions of structure and autonomy (chaos and control) in relation to cognitive or non-cognitive outcomes. The only study that explored this in relation to motivational and self-regulation aspects (autonomous motivation, academic engagement, academic self-efficacy, use of meta-cognitive strategies and procrastination) and psychological basic need satisfaction in secondary education found that dependent on the outcome, the positive (support) or the negative dimension (thwart) were equally important, or one was more important than the other (Opdenakker, 2015). In general, thwart mattered most for maladaptive student behaviour and support mattered most for adaptive behaviour (although thwart had often also a negative effect).

Several studies also found evidence that it is important to ask students about their experiences instead of asking teachers about their teaching (e.g. Van Damme et al., 2004). Evidence was found that how students experienced their learning environment could explain more variation in student outcomes than teachers' reports about their teaching.

10.4.4.2 Effects of Class Climate

Class climate can refer to the learning climate in classes as well as to the relational climate (teacher-student and student-student relationships). Several studies in secondary education indicate the importance of the *learning climate* in classes in relation to cognitive or non-cognitive functioning. For example, De Fraine et al. (2003) and Opdenakker, Van Damme, De Fraine, et al. (2002) found a positive effect of the degree to which the class was study-oriented on language and math achievement respectively at the end of the eighth grade. Opdenakker and colleagues (Opdenakker, 2004; Opdenakker et al., 2005), and Dumay and Dupriez (2007) found the same with regard to math achievement and effort for math, and for math achievement alone, respectively. Van Landeghem et al. (2002) found positive effects of learning climate on a diversity of non-cognitive and wellbeing indicators, and Van Damme et al. (2004) found evidence of a positive effect on attitudes towards math.

Clear evidence has also been found for the importance of a good *relational climate* in classes as well. The relational climate in classes can refer to the relationships between teacher and students or the relationships between the students in the same class. Most studies have investigated the teacher-student relationship, or pay attention to teacher involvement towards their classes/students or teachers'

proximity in their relationships with their class. Opdenakker (2004) found evidence for the importance of a good relationship between teachers and students/the class group, as well as students' integration within their own class on effort for math and math achievement in secondary education in Flanders. In addition, Van de gaer, Pustjens, Van Damme, and De Munter (2006) found that secondary education class group perceptions of the quality of the relationships with their teachers, and the degree to which the students in a class group feel integrated in their class, have a positive effect on language achievement of boys and – particularly with regard to integration in the class – also of girls. Work of Opdenakker, Maulana, Stroet and colleagues in the Netherlands (Opdenakker, 2013, 2014; Opdenakker & Maulana, 2010; Stroet et al., 2015; Maulana, Opdenakker, Stroet, & Bosker, 2013) indicate the importance of teacher involvement – which is important for students to meet their psychological basic need to feel related to significant others – in relation to motivational aspects and academic engagement of students. Lietaert et al. (2015) also found positive effects of teacher involvement on academic engagement in Flanders.

Cottaar (2012) found a positive effect of the pleasantness of secondary teachers on students' interest in physics. Den Brok and colleagues (2004, 2005, 2010) studied effects of teacher proximity (as perceived by their students) on their students' achievements, motivation and attitudes, and found positive effects on all type of outcomes. In line with this, Opdenakker et al. (2012) found positive effects of teachers' interpersonal behaviour on motivation. Also, Van Petegem, Creemers, Aelterman, and Rosseel (2008) and Van Petegem, Aelterman, Rosseel, and Creemers (2007) investigated effects of teachers' interpersonal behaviours on students' well-being at school (secondary education) and found evidence for positive effects of the tolerant/authoritative teacher compared to the authoritarian teacher and the dominant cooperative teacher, as perceived by the students.

Positive effects of class climate are also demonstrated in studies on primary education. The study of Dewulf, van Braak, and Van Houtte (2017) found evidence for the important role teacher trust plays in language achievement and development of students in (socially and ethnically) segregated schools. Teacher trust seemed to be a key factor in explaining growth in reading comprehension, and their study demonstrated that it could mediate the (positive) effect of ethnic diversity of classes.

Almost no studies investigated the effects of the opposite dimension of teacher involvement in relation to students (neglect). Opdenakker (2015) explored this in relation to motivational and self-regulation aspects and psychological basic need satisfaction in secondary education, and found that dependent on the outcome, teacher involvement or teacher neglect was equally important, or one was more important than the other. In general, teacher thwart[33] mattered most for maladaptive

[33]Thwart refers to the opposite of support. For example, with regard to teacher involvement it refers to teacher neglect, with regard to autonomy support it refers to controlling behaviour and controlling language of the teacher towards the students, and with regard to structure support it refers to chaos in the classroom.

student behaviour and teacher support mattered most for adaptive behaviour (although teacher thwart had often also a negative effect on adaptive behaviour).

10.4.4.3 Effects of Group Composition (Also in Relation to What Students Experience in the Classroom)

In general, clear-cut effects of group composition are found with regard to cognitive functioning in secondary education. For example, De Fraine et al. (2003), Opdenakker, Van Damme, De Fraine, et al. (2002), Opdenakker, Van Damme, and Minnaert (2002a, 2002b, 2005), and Dumay and Dupriez (2007) found positive effects of the mean cognitive ability/achievement of classes and mean social-economic status (SES) of classes on respectively language and math achievement in the lower grades of secondary education in Belgium and the Netherlands. They also found evidence for a (medium to strong) positive relationship between class composition and class process variables. One of the strongest relationships was found between mean cognitive ability and class learning climate, which resulted in a joint positive effect on achievement. Also, strong relationships were found between opportunity to learn and class composition. In the study of De Fraine et al. (2003), only an additional unique effect of group composition remained on language achievement after the inclusion of student background, school and class practice characteristics. Research of Opdenakker, Van Damme, De Fraine, et al. (2002) clearly demonstrated that next to net effects of group composition (mean cognitive ability and SES) and process variables, a large common effect of group composition and process was visible. In addition, it appeared that the net effect of group composition was considerably larger than the net effect of the process variable (learning climate), which was much smaller than the common effect. Also, Opdenakker and Van Damme (2001, 2006b), Opdenakker et al. (2005), and Opdenakker (2004, 2014) found evidence of relationships between class composition and class process, indicating a positive relationship between the cognitive ability of classes and the amount of instructional support, structure in the learning environment, teacher involvement, and a smaller decrease in autonomy support during the school year. In addition, Opdenakker and Van Damme (2006b) found that the degree to which parents are involved in the learning and the school career of their children in the class group, is positively related to climate characteristics in that group, which is also positively influenced by a teacher that uses a learner-centered teaching style. In addition, they found that the mean motivation of the class group influences the degree to which the learning climate in the class is optimal. Maulana et al. (2016) found evidence that students in high-ability classes perceive more clarity of instruction and stronger teacher control than their peers in heterogeneous mixed-ability classes. Also, other studies found evidence for a relationship between composition and student achievement (e.g. de Jong et al., 2004, who found a positive link between the average level of prior math achievement in the class and later math achievement of students) and between composition and practice.

A few studies (e.g. Van Landeghem et al., 2002; De Fraine, Van Damme, & Onghena, 2002) investigated the effects of class group composition on school well-being in secondary education. Both mentioned studies addressed the effects of the average cognitive ability level of the class and found a positive effect of this characteristic on students' school well-being (over and above their initial achievement). School well-being referred in these studies to relationship with teachers, liking of the school, interest in learning tasks and attentiveness in the classroom.

A few Flemish studies (mainly in the second decade of EER)[34] paid also attention to the gender composition of classes (i.e. the proportion of girls) or characteristics of boys and girls in classes in relation to cognitive outcomes, and found evidence for effects of these kinds of class characteristics. For example, De Fraine et al. (2003) investigated the effect of the proportion of girls in the class, and found that this class characteristic had a positive effect on language achievement in secondary education, even after controlling for a diversity of student characteristics.

Furthermore, Van de gaer et al. (2006) investigated the effect of attitudes of peers in the class and same-sex students' attitudes in the class (with regard to the relationship with teachers and integration in the class) on math achievement of boys and girls in secondary education. They found that the attitudes of the class group, and especially the attitudes of the same-sex students in the class, had a stronger impact on the math achievement of boys compared to that of girls. This seems to indicate that boys are more sensitive to the attitudes of boys in their class or that peer pressure seems to be higher for boys. In addition, they found that boys seemed to perform less well compared to girls in classes with no good relationship with the teacher, in classes where students were not very motivated and in classes where students felt poorly integrated. The findings of the studies of Van Houtte (2004) and Van de gaer, Pustjens, Van Damme, and De Munter (2007) related to (language) achievement point in the same direction.

Effects of class composition on cognitive outcomes have also been studied in primary education. Compared to secondary education, effects of class composition are (often) more modest in size or do not exist at all. In addition, findings seem sometimes inconsistent. For example, Boonen, Van Damme, and Onghena (2014) investigated the effects of mean math/language achievement and heterogeneity, mean SES and percentage of minority students in relation to math, spelling and reading. They did not find effects on math, small effects of the heterogeneity in language on reading, and small positive effects of mean SES and mean language achievement on spelling. Also, Boonen, Speybroeck, et al. (2014) did not find direct school composition effects related to SES, ethnicity, and initial achievement on mathematics achievement at the end of the second grade. However, they did find two small indirect effects of mean prior achievement and mean SES, suggesting that schools scoring high on these composition characteristics keep in regular contact with the parents of their students, and this seems to enhance the mathematics

[34] At that time, schools and classes in Flanders could consist of 100% of girls, since Flemish schools could be only for girls, only for boys or mixed-sex schools.

achievement of their students. Vandecandelaere, Vanlaar, De Fraine, Van Damme, and Verhaeghe (2011) found an effect of ethnic diversity on math achievement. However, they did not find an effect of this composition characteristic on growth in math. Vanlaar et al. (2014) studied the effects of the percentage high-risk students on the level of reading comprehension at the end of the fifth grade and on the learning gain between the fourth and the fifth grade. They found a negative effect of the percentage of high-risk students on the level of reading comprehension, but no effect on growth.

Driessen and Sleegers (2000) found a clear negative effect of a high proportion of disadvantaged students in the class (based on students' socio-ethnic background) on language and math achievement et the end of primary education; the research of van der Slik et al. (2006) also found a negative effect of a high proportion of ethnic minority students in classes on language achievement. Hornstra et al. (2015a) found a negative effect of low-SES classes on the reading comprehension of low-SES students. However, their research also revealed a positive effect on the reading comprehension of these students when they belonged to a class with a high proportion of minority students. In addition, in classes with a high proportion of low-SES or ethnic minority students, all students showed more positive developments in motivation. This seems to be in line with findings of Vanwynsberghe et al. (2017a), who discovered a positive effect of a high proportion of at risk students on motivation. Research of Vause et al. (2010) revealed an effect of the mean cultural capital of the class on math achievement, and also found evidence of an interaction effect between composition and teaching practice, indicating, among other things, that math in disadvantaged classes can be improved by maximizing time on task.

Peetsma et al. (2006) investigated development in math and language achievement, and found that in classes with a large number of ethnic minority students, the increase in math achievement is smaller, and in classes with a large number of low-SES students the increase in language proficiency is smaller. However, they did not find effects of class composition on developments in wellbeing, which is in line with findings of Vanwynsberghe et al. (2017a). There is also some evidence that the diversity of the class group in terms of ethnicity matters in a negative way: Veenman, Van de Werfhorst, and Dronkers (2013) found a negative effect of ethnic diversity on reading comprehension in Dutch primary education. Weide (1993) and Agirdag, Merry, and Van Houtte (2016) found evidence for a (negative) association between class composition (related to ethnic diversity or proportion of minority students) and school quality and class practice, respectively. Luyten and Van der Hoeven-Van Doornum (1995) found evidence of a positive effect of the intellectual level of the class group on student achievement.

Furthermore, a few studies paid attention to a comparison between effects of class composition and school composition, and revealed that class composition is often more important than school composition (or is solely important). Evidence for this can, for example, be found in the study of De Fraine et al. (2003) and Opdenakker, Van Damme, De Fraine, et al. (2002) with regard to cognitive outcomes, and in the study of Van Landeghem et al. (2002) with regard to non-cognitive outcomes.

10.4.4.4 Effects of Configurations of Learning Environments as a Holistic Way to Look at Effects of Learning Environments

It is remarkable that, although much EER research reveals clear indications of covariance of characteristics of schools and classes (e.g. combinations of process characteristics as well as combinations of process characteristics and composition), that only a few studies addressed schools and classes in a holistic way and investigated the existence of configurations or constellations.[35] Hofman, Hofman and colleagues (Hofman, Hofman, & Guldemond, 2003; Hofman & Hofman, 2011) explored configurations in the Netherlands. Hofman et al. (2003) investigated the existence of configurations related to the social context of learning of secondary schools in the Netherlands and studied their effects on cognitive outcomes and social outcomes. Hofman and Hofman (2011) studied configurations of management styles in secondary education based on the assumption that more formal management should fit with the specific, more cultural management elements and studied their relationship with math achievement. Opdenakker (2004) investigated the existence of configurations of classes and of secondary schools in Flanders in relation to math achievement. Hofman et al. (2003) distinguished four configurations empirically based on the social context of school, peer groups and family and found effects on math achievement and social outcomes. Hofman and Hofman (2011) found strong differences in effective management styles between schools which differed in school composition and their study delivered evidence that the configurations had added effects on math achievement in addition to the effects of individual management characteristics. Opdenakker (2004) used characteristics of process and composition as input for empirically based configurations (of classes and of schools) and found evidence for three configurations related to classes and four configurations related to schools. The class configurations were (1) the class with an favourable climate, (2) the class with an unfavourable learning environment, and (3) the 'everything goes well' class (favourable learning environment and climate). The school configurations referred to (1) the 'everything goes well' school (referring, among other things, to a school with a favourable student composition, climate, opportunity to learn and teacher cooperation), (2) the school with frictions at the class level (referring to a school which has a rather low quality of learning climate in the classes and an average level of opportunity to learn), (3) the unfavourable school (referring to a school which has an unfavourable student composition, low quality of relational climate, average learning climate, limited amount of opportunities to learn, limited cooperation between teachers, a school leader scoring low on participative professionality-oriented leadership), and (4) de cramped unfavourable school (referring to a school with an unfavourable student composition, small school size, low

[35]Configurations or constellations refer to clusters or groups which are constructed based on a variety of characteristics which are more or less the same between the members of a group and quite different between groups. Because a variety of characteristics of groups is simultaneously considered, it addresses groups or clusters in a holistic way.

quality of relational and learning climate, very limited cooperation between teachers, quite high amount of curriculum offerings and a school leader scoring rather high on participative professionality-oriented leadership). The mentioned class and school configurations were able to explain differences in math achievement between classes or schools even when controlling for a variety of class composition characteristics. In another study (Hofman et al., 2013), characteristics of innovative secondary schools in the Netherlands were explored to find configurations of innovative schools. Evidence was found for five clusters in secondary education which differed significantly on quality assessments of the inspectorate related to quality of time and teacher-learning processes. These clusters were related to student outcomes. Results indicated that, in the lower educational tracks in particular, more innovative schools – which were in general a bit more often private nondenominational schools of relatively small size (500 students or less) – obtained good results with regard to the school careers of their students. The clusters did not differ with regard to the percentage of ethnic minority students in the schools.

10.4.4.5 The Connection Between School-Level Factors and Learning Environment (i.e. Teacher/Classroom) Characteristics

Studies on the relationship between school-level factors and learning environment characteristics are quite scarce. Opdenakker, Van Damme, De Fraine, et al. (2002) paid attention to relations between school process and class process factors (mainly based on teaching staff and school leader perceptions) in Flanders and found only weak relationships, where any significant relationship was found. De Jong et al. (2004), testing some of the main components of Creemers' comprehensive model of educational effectiveness in the Netherlands, did not find evidence for relations between the school and class level.

Opdenakker and Van Damme (2006a) studied the school- (and teacher-) relatedness of a variety of class practices in math classes and found evidence for such a school-relatedness, although the size of the relatedness depended on the class practice characteristic investigated. For example, evidence was found for clear school-relatedness in the quality of the relationship between students and their math teacher (26% of the variance at school level) and the degree to which a focus on individual development was visible in the math classes (28% of the variance at school level). A somewhat weaker school-relatedness was found for the degree to which students were integrated in their classes (11% at school level) and opportunity to learn math (10% at school level), and a rather weak school-relatedness was found with regard to the degree to which an optimal learning climate was present in the math classes (4% of the variance at school level). The most comprehensive study on the relation between school and learning environment levels was conducted by Opdenakker (2004), addressing the connection between a variety of secondary school-level factors, teacher characteristics and learning environment characteristics, referring to instructional support and climate factors in secondary school math classes, and addressing relationships between characteristics of the same level

(school or class) as well. The study was based on a multilevel multifactor framework and parts of this study were published in journal articles e.g. Opdenakker and Van Damme (2006b, 2007). Findings of this study reveal clear evidence for more differences between classes within schools than differences between schools with regard to the studied class practices. In general, evidence was found for connections between school characteristics and characteristics of the learning environment and climate in classes. For example, it was found that the quality of the climate in classes (relational climate, learning climate) is, among other things, positively connected with the amount of cooperation between teachers in a school. In addition, the degree to which the school leader acts as a participative professionality-oriented leader is positively connected with the degree to which students are integrated in their classes and with the amount of instructional support students receive from their teachers. Furthermore, connections between teacher characteristics and instructional support and climate in classes were found. For example, a learner-centered teaching style was positively linked to the degree to which classes received instructional support and played a significant role in the quality of the relationship between the teacher and class.

Dumay (2009) studied the relationship between schools' organizational culture characteristics (transformational leadership, teachers' collective decision making about pedagogical aspects, cultural homogeneity, cultural values) and students' mathematics achievement in the French-speaking part of Belgium. He found that cultural homogeneity was positively related to transformational leadership and collective decision making, but it did not show a clear moderation effect on the relationship between cultural values and students' achievement. In addition, the relation between school culture and student achievement was weak. Hulpia, Devos, and Rosseel (2009) investigated the relationship between school leadership and teacher commitment and job satisfaction and found a strong positive relationship between the cohesion of leaders (the leadership team), the amount of leader support and participative decision-making with teacher commitment and, to a lesser degree and also indirectly, with the job satisfaction of teachers. In addition, they found that decentralization of leadership functions was only weakly related to teacher commitment. Aelterman, Engels, Van Petegem, and Verhaeghe (2007) also investigated the school–teacher relationship. They found that, although teachers' general professional wellbeing was most strongly related to teachers' own self-efficacy in primary and secondary education, support from the principal mattered as well through experienced support from colleagues. Ten Bruggencate et al. (2012) found a relationship between school leadership on one hand, and a development-oriented school organization and favourable class practices on the other hand.

Also, Van Gasse, Vanlommel, Vanhoof, and Van Petegem (2017) investigated relations between school characteristics and teacher practice and found evidence for a positive relation between teacher collaboration and teachers' individual data use (based on student performance) to improve their own practice. It must be said, however, that teachers' general self-efficacy also influenced their collaboration and data use.

A few studies (e.g. Opdenakker, 2004; Agirdag et al., 2016; Vervaet, Van Houtte, & Stevens, 2018) paid attention to links between (school) composition and learning environment characteristics such as teaching practice or climate indicators. All of these studies found evidence for links between group composition characteristics and characteristics of the learning environment. Opdenakker (2004) found links between a variety of composition characteristics (referring to cognitive ability, motivation, SES, and parental involvement in the schooling of their child) and climate and instructional support characteristics of learning environments. Vervaet et al. (2018) investigated the relationship between school composition (proportion of ethnic minorities) and teaching practice (multicultural teaching) and found evidence for a positive relationship. Also, Agirdag et al. (2016) found an association between the ethnic composition of a school and a specific dimension of multicultural teaching, namely, multiethnic content integration.

10.4.4.6 Generic Versus Differentiated Educational Effectiveness

As mentioned before, only one out of five studies paid attention to differential effects of (characteristics of) schools, teachers or classes. This implies that for most studies we do not know if differential effects are present. Differential effects related to the learning environment at the class level will be discussed here, as differential effects of school characteristics were already discussed in an earlier section. In general, based on the studies that address differential effects, it seems to be the case that initially low performers or students at risk (because of their low SES or foreign ethnic background) are most sensitive in primary as well as in secondary education. Evidence of this can be found, for example, in the studies of De Fraine et al. (2003) and Vanlaar et al., (2014) on language achievement, or of de Bilde et al. (2013) in relation to school engagement. However, there is also some evidence for the opposite. For example, Opdenakker, Van Damme, De Fraine, et al. (2002) found that initially high performing students in math were more sensitive to the average cognitive level of their class group and benefited most from a high cognitive ability class group. Furthermore, they found that initially high-achievement-motivated students were more sensitive to characteristics of their learning environment. In addition, they found that girls performed as a group more heterogeneously in classes with a negative learning climate. Evidence was also found that low-SES students were as a group more sensitive to the mean SES of the class than high-SES students, and performed much more heterogeneously in low-SES classes compared to high-SES classes and in comparison with high-SES students.

Van de gaer et al. (2006) also found differential effects of climate variables in secondary education related to gender. However, in their study, the differential effect indicated that boys seemed to be, on average, more sensitive than girls. Lietaert et al. (2015) also found that boys seemed to be more sensitive than girls in secondary education, at least with regard to autonomy support in relation to academic engagement. Autonomy support seemed to be relevant only for boys in relation to academic

engagement in their study. However, they did not find differential effects of teacher involvement in relation to gender.

Research of Vanlaar et al. (2014) revealed differential effects of class process (instruction practice: well-organized and attractive instruction, differentiated instruction, discovery learning, cooperative learning) in primary education related to low- versus high-risk students. The majority of these differential effects were related to the level of reading comprehension (not to growth, although the effect of well-organized and attractive instruction seemed to have a bit less of impact on high-risk students, and differentiated instruction had no effect for high-risk students). It seems that more constructivist practices (discovery learning, cooperative learning) have rather negative effects on the reading comprehension of high-risk students, in contrast to more traditional, structured approaches, which mostly have positive effects on all students, but a bit less so on high-risk students. In general, high-risk students are sensitive to negative effects of constructivist approaches (after controlling for student gender and the percentage of high-risk students). In contrast, Vanlaar et al. (2014) found generic positive effects for teaching learning strategies and effective classroom management (level and growth). This seems to be in line with research of Droop et al. (2016). Droop et al. (2016) did not find evidence for differential effects of sustained strategic reading instruction in primary education in relation to a diversity of student background characteristics.

Opdenakker and Minnaert (2014) investigated differential effects of learning environment experiences (the fulfillment of the psychological basic needs of competence, autonomy and relatedness by the teacher) on academic engagement at the end of primary education in the Netherlands related to the initial engagement of students, gender and ethnic-cultural background of the students. They found that the effect of the fulfillment of the need of competence was more important for initially high academic engaged students, indicating that for students with a higher prior engagement, the effect of the competence need satisfaction by the teacher is stronger and their engagement at the end of primary education is higher in line with their prior engagement when they experience a high level of competence induced by their teacher. In addition, the overall basic need satisfaction seemed to be more important for Dutch-speaking students. Furthermore, den Brok et al. (2010) found evidence for differential effects of teacher interpersonal relationships on students' attitudes related to students' ethnicity.

Research of Peetsma et al. (2006) also revealed differential effects of class composition: the effect of a high percentage of low-SES students was negative for high-SES students and positive for low-SES students.

10.5 Conclusion, Discussion and Future Directions

Three decades of EER in Belgium and the Netherlands has provided a field that is substantial in scope, with well-validated findings in several areas. In general, findings are quite in line with the international EER literature, and the EER studies

clearly demonstrate that schools, teachers and classes matter with regard to students' cognitive and non-cognitive functioning and their educational positions in school from their first experiences with school and educational arrangements onwards. In particular, the quality of instructional support and guidance, and the instructional and learning experiences students have in their learning environments, have effects on their level of functioning, growth and development in relation to cognitive and non-cognitive functioning. Teachers play an important role in creating effective, structured and supportive learning environments and can be seen as primary actors in helping to fulfill students' needs to feel competent, eager to learn and autonomous as well. The learning and relational climate in classes and in schools (relationships with teacher and classmates) is also of particular importance. Good and warm relationships with teachers, and a high amount of teacher trust and involvement regarding their students, is of utmost importance not only to motivate students for school, but also for their cognitive achievement and growth. According to Dewulf et al. (2017), teachers, and in particular teacher trust, play a crucial role in fostering equity and quality in education. Others (e.g. Goddard, Tschannen-Moran, & Hoy, 2001) refer to teachers as the key figures in overcoming the disadvantaged positions of minority students. In addition, a learning environment in which a learning- and study-oriented climate is shaped, is of importance to motivate students for learning, and a well-integrated class is important for students' wellbeing at school. Furthermore, clear evidence is found for effects of composition characteristics of the class (and school) group the students belong to. Several EER studies referred to in this chapter have clearly demonstrated (and tested) that group composition is not an artefact of measurement error, as some researchers have suggested in the past. In addition, research has clearly demonstrated that group composition and the quality of instructional arrangements and climate (learning and relational) are interrelated, often indicating that low-risk compositions are at an advantage with regard to the quality of learning environments. However, research has also demonstrated that when educational policy makes it financial possible to provide compensational arrangements at schools (e.g. more financial resources to schools with a lot of at-risk students to reduce class size, to make it possible to have more teachers per number of students or to have more possibilities to meet and accommodate students' educational needs), effects of group composition diminish. In addition, teachers themselves can also play a role in overcoming negative effects of group composition by creating learning opportunities in correspondence to students' educational needs, not only for majority students or students with low risk, but also for ethnic minority students and students at risk. Giving trust and developing good social relationships with students in general, and in particular with minority students and students at risk, is quite important towards overcoming negative group composition effects.

At the school level, there is some evidence that the cooperation of teachers matters and that it has an influence on the quality of learning environments (instruction support, climate) in classes. This suggests that school characteristics such as teacher cooperation can play a role in enhancing educational outcomes at least in an indirect way, which is in agreement with ideas on the working of school level factors of Creemers and Reezigt already formulated in 1996 (Creemers & Reezigt, 1996). It

is striking that during the past decades of EER in Belgium and the Netherlands only a few studies paid attention to cross-level relations between school characteristics and characteristics/quality indicators of learning environments at the level of the class, with no study investigating this in a longitudinal way. The same holds for cross-level relations between teacher characteristics, such as preferred teaching style, and characteristics of classes and learning environments within classes. Evidence for the importance of such cross-level relations can be found, e.g. in Opdenakker and Van Damme (2006a, 2006b, 2007), Opdenakker (2004), Creemers and Reezigt (1996). In addition, the context-specificity of effectiveness-enhancing characteristics have seldom been studied. In general, these findings are in line with international EER (see, e.g., Scheerens, 2007 and the state-of-the-art articles in School Effectiveness and School Improvement; Reynolds et al., 2014). I agree with the request of Reynolds et al. (2014) in their state-of-the-art article for more studies on cross-level interactions between schools and their teachers/classes, and on the context-specificity of effectiveness-enhancing characteristics of schools, teachers and classes. This request is also important for EER in Belgium and the Netherlands, although attempts to address these topics seem to have been made already somewhat more in Belgium. Moreover, there is a case to be made that longitudinal studies are needed in which not only the development of students' functioning and school career is explored in relation to characteristics of their schools and educational arrangements within their schools and classes, but also that this relationship is studied in a dynamic way, paying attention to changing characteristics of schools and educational arrangements and to the dynamic interplay between characteristics of schools (organization, context, processes, composition) and educational arrangements, teacher behaviour, and processes in classes. Evidence for the importance of this can be found in the work of Opdenakker, Maulana and colleagues (see e.g. Maulana et al., 2013; Maulana et al., 2016; Opdenakker, 2014; Opdenakker & Maulana, 2010; Opdenakker et al., 2012; Stroet et al., 2015), in which evidence is found for changes in the quality of learning environments during the school year as well as for links between changes in motivational aspects and quality of learning environments. In addition, research should continue to pay attention to differential effects of educational arrangements and to study the effectiveness of teaching and educational arrangements from the viewpoint of students' educational and basic psychological needs. In addition, it is time to pay more attention to new learning skills such as metacognitive and self-regulation skills, thinking skills and other twenty-first century skills in relation to effective educational arrangements at school and class level.

Furthermore, more work is needed at a theoretical and conceptual level, including conceptualizing schools and learning environments in a holistic way instead of focusing on separate factors and characteristics, e.g. by conceptualizing them in (dynamic) configurations or constellations. In addition, attention should be paid to school and class composition in relation to educational arrangements in future research. Although EER in Belgium and the Netherlands is more and more grounded in (some) theoretical underpinnings, and some attempts have been made to ground studies in theoretical models of educational effectiveness (e.g. work of Opdenakker and of Creemers and Scheerens), an even more theory-driven approach is needed as

well as more theoretical work on explanations of effects, explanations of mechanisms, (dynamic) constellations, and models including mediating and moderating effects to interpret findings and to guide future EER.

References

Aelterman, A., Engels, N., Van Petegem, K., & Verhaeghe, J. (2007). The well-being of teachers in Flanders: The importance of a supportive school culture. *Educational Studies, 33*, 285–298.

Agirdag, O., Driessen, G., & Merry, M. S. (2017). The Catholic school advantage and common's hoop effect examined: A comparison between Muslim immigrant and native pupil in Flanders. *School Effectiveness and School Improvement, 28*, 123–135.

Agirdag, O., Merry, M. S., & Van Houtte, M. (2016). Teachers' understanding of multicultural education and the correlates of multicultural content integration in Flanders. *Education and Urban Society, 48*, 556–582.

Agirdag, O., Van Houtte, M., & Van Avermaet, P. (2012). Why does the ethnic and socioeconomic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review, 28*, 366–378.

Avram, S., & Dronkers, J. (2011). School sector variation on non-cognitive dimensions: Are non-public but publicly supported schools more effective? *Educational Research and Evaluation, 17*, 115–139.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*, 725–739.

Belfi, B., Gielen, S., De Fraine, B., Verschueren, K., & Meredith, C. (2015). School-based social capital: The missing link between schools' socioeconomic composition and collective teacher efficacy. *Teaching and Teacher Education, 45*, 33–44.

Belfi, B., Goos, M., Pinxten, M., Verhaeghe, J. P., Gielen, S., De Fraine, B., & Van Damme, J. (2014). Inequality in language achievement growth? An investigation into the impact of pupil socio-ethnic background and school socio-ethnic composition. *British Educational Research Journal, 40*, 820–846.

Belfi, B., Haelermans, C., & De Fraine, B. (2016). The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach. *British Journal of Educational Psychology, 86*, 501–525.

Boonen, T., Pinxten, M., Van Damme, J., & Onghena, P. (2014). Should schools be optimistic? An investigation of the association between academic optimism of schools and student achievement in primary education. *Educational Research and Evaluation, 20*, 3–24.

Boonen, T., Speybroeck, S., de Bilde, J., Lamote, C., Van Damme, J., & Onghena, P. (2014). Does it matter who your schoolmates are? An investigation of the association between school composition, school processes and mathematics achievement in the early years of primary education. *British Educational Research Journal, 40*, 441–466.

Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: Which aspects matter most? *School Effectiveness and School Improvement, 25*, 126–152.

Brandsma, H. P., & Knuver, J. W. M. (1989). Effects of school and classroom characteristics on pupil progress in language and arithmetic. *International Journal of Educational Research, 13*, 777–788.

Brandsma, H. P., & Stoel, W. G. R. (1987). Schoolleiders in het voortgezet onderwijs. Een exploratieve studie naar de determinanten en effecten van schoolleiderskenmerken [School leaders in secondary education. An explorative study into the determinants and effects of school leader characteristics]. In J. Scheerens & W. G. R. Stoel (Eds.), *Effectiviteit van onderwijsorganisaties* (pp. 99–113). Lisse, The Netherlands: Swets & Zeitlinger.

Bressoux, P., & Bianco, M. (2004). Long-term teacher effects on pupils' learning gains. *Oxford Review of Education, 30*, 327–345. https://doi.org/10.1080/0305498042000260476

Brutsaert, H. (1998). Home and school influences on academic performance: State and Catholic elementary schools in Belgium compared. *Educational Review, 50*, 37–43.

Brutsaert, H. (2006). Gender-role identity and perceived peer group acceptance among early adolescents in Belgian mixed and single-sex schools. *Gender and Education, 18*, 635–649.

Brutsaert, H., & Van Houtte, M. (2002). Girls' and boys' sense of belonging in single-sex versus co-educational schools. *Research in Education, 68*, 48–56.

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.

Cottaar, A. (2012). Low (linear) teacher effect on student achievement in pre-academic physics education. *Journal of Research in Science Teaching, 49*, 465–488.

Creemers, B. P. M. (1974). *Evaluatie van onderwijsstijlen binnen het aanvankelijk lezen. Een onderwijskundig onderzoek naar de relatie tussen methodegebruik van de leerkracht en de prestaties van leerlingen* [Evaluation of teaching styles for initial reading. An educational study into the relation between methods used by teachers and student achievement]. PhD thesis, Utrecht University. Vaassen, The Netherlands: van Walraven.

Creemers, B. P. M. (1983). De bijdrage van onderwijsonderzoek aan de verhoging van de kwaliteit van het onderwijs [The contribution of educational research to the enhancement of the quality of education]. In B. Creemers, W. Hoeben, & K. Koops (Eds.), *De Kwaliteit van het onderwijs* [The quality of education] (pp. 215–232). Haren/Groningen, The Netherlands: RION/Wolters-Noordhoff.

Creemers, B. P. M. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B. P. M. (2007). Educational effectiveness and improvement: The development of the field in mainland Europe. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 223–242). Dordrecht, The Netherlands: Springer.

Creemers, B. P. M., & de Jong, R. (2002). Explaining differences in student outcomes. *Journal of Classroom Interaction, 37*(2), 16–26.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK: Routledge.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London, UK: Routledge.

Creemers, B. P. M., & Lugthart, E. (1989). School effectiveness and improvement in the Netherlands. In D. Reynolds, B. P. M. Creemers, & T. Peters (Eds.), *School effectiveness and improvement. Proceedings of the First International Congress, 1988* (pp. 89–103). London, UK/Cardiff, UK/Groningen, The Netherlands: School of Education, University of Wales College of Cardiff/RION, Instituut voor Onderwijsonderzoek.

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*, 197–228.

Creemers, B. P. M., & Schaveling, J. (1985). *Verhoging van onderwijseffectiviteit* [Improving educational effectiveness]. The Hague, The Netherlands: WRR.

de Bilde, J., Boonen, T., Speybroeck, S., De Fraine, B., Van Damme, J., & Goos, M. (2015). Experiential education in kindergarten: Association with school adjustment and the moderating role of initial achievement. *The Elementary School Journal, 116*, 173–197.

de Bilde, J., Van Damme, J., Lamote, C., & De Fraine, B. (2013). Can alternative education increased children's early school engagement? A longitudinal study from kindergarten to third grade. *School Effectiveness and School Improvement, 24*, 212–233.

De Fraine, B., Van Damme, J., & Onghena, P. (2002). Accountability of schools and teachers: What should be taken into account? *European Educational Research Journal, 1*, 403–428.

De Fraine, B., Van Damme, J., & Onghena, P. (2007). A longitudinal analyses of gender differences in academic self-concept and language achievement: A multivariate multilevel latent growth approach. *Contemporary Educational Psychology, 32*, 132–150.

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29*, 841–859.

de Haan, A., Elbers, E., Hoofs, H., & Leseman, P. (2013). Targeted versus mixed preschools and kindergartens: Effects of class composition and teacher managed activities on disadvantaged children's emergent academic skills. *School Effectiveness and School Improvement, 24*, 177–194.

de Jager, B., Jansen, M., & Reezigt, G. (2005). The development of metacognition in primary school learning environments. *School Effectiveness and School Improvement, 16*, 179–196.

de Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*, 3–31.

De Maeyer, S., Rymenans, R., Van Petegem, P., van den Bergh, H., & Rijlaarsdam, G. (2007). Educational leadership and pupil achievement: The choice of a valid conceptual model to test effects in school effectiveness research. *School Effectiveness and School Improvement, 18*, 125–145.

De Smedt, F., & Van Keer, H. (2018). Fostering writing in upper primary grades: A study into the distinct and combined impact of explicit instruction and peer assistance. *Reading and Writing, 31*(2), 325–354.

De Smedt, F., Van Keer, H., & Merchie, E. (2016). Student, teacher and class-level correlates of Flemish late elementaire school children's writing performance. *Reading and Writing: An Interdisciplinary Journal, 29*, 833–868.

den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness School Improvement, 15*, 407–442.

den Brok, P., Levy, J., Brekelmans, M., & Wubbels, T. (2005). The effect of teacher interpersonal behaviour on student subject-specific motivation. *Journal of Classroom Interaction, 40*, 20–33.

den Brok, P., van Tartwijk, J., Wubbels, T., & Veldman, I. (2010). The differential effect of the teacher-student interpersonal relationship on student outcomes for students with different ethnic backgrounds. *British Journal of Educational Psychology, 80*, 199–221.

Denies, K., Magnus, I., Desmet, P., Gielen, S., Heyvaert, L., & Janssen, R. (2016). Motivation in a French L2 context: Teacher motivational practices and student attitudes in relation to proficiency. *Vigo International Journal of Applied Linguistics, 13*, 93–126.

Dewulf, L., van Braak, J., & Van Houtte, M. (2017). The role of teacher trust in segregated elementary schools: A multilevel repeated measures examination. *School Effectiveness and School Improvement, 28*, 259–275.

Dijkstra, A. B., Geijsel, F., Ledoux, G., van der Veen, I., & ten Dam, G. (2015). Effects of school quality, school citizenship policy, and student body composition on the acquisition of citizenship competences in the final year of primary education. *School Effectiveness and School Improvement, 26*, 524–553.

Driessen, G. (2002). School composition and achievement in primary education: A large scale multilevel approach. *Studies in Educational Evaluation, 28*, 347–368.

Driessen, G., Agirdag, O., & Merry, M. S. (2016). The gross and net effects of primary school denomination on pupil performance. *Educational Review, 68*, 466–480.

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement, 11*, 57–79.

Driessen, G., & van der Slik, F. (2001). Religion, denomination, and education in the Netherlands: Cognitive and noncognitive outcomes after an era of secularization. *Journal of the Scientific Study of Religion, 40*, 561–572.

Dronkers, J. (2004). Do public and religious schools really differ? Assessing the European evidence. In P. J. Wolf & S. Macedo (Eds.), *Educating citizens. International perspectives on civic values and school choice* (pp. 287–314). Washington, DC: Brookings Institution Press.

Dronkers, J., & Robert, P. (2003). *The effectiveness of public and private schools from a comparative perspective* (Working Paper SPS 2003/12). SAN Domenico di Fiesole, Italy: European University Institute. Retrieved December 1, 2005, from http://www.iue.it/PUB/sps2003-13.pdf

Droop, M., van Elsäcker, W., Voeten, M. J. M., & Verhoeven, L. (2016). Long-term effects of strategic reading instruction in the intermediate elementary grades. *Journal of Research on Educational Effectiveness, 9*, 77–102.

Dumay, X. (2009). Origins and consequences of schools' organizational culture for student achievement. *Educational Administration Quarterly, 45*, 523–555.

Dumay, X., Boonen, T., & Van Damme, J. (2013). Principal leadership long-term indirect effects on learning growth in mathematics. *The Elementary School Journal, 114*, 225–251.

Dumay, X., & Dupriez, V. (2007). Accounting for class effect using the TIMSS 2003 eighth-grade database: Net effect of group composition, net effect of class process, and joint effect. *School Effectiveness and School Improvement, 18*, 383–408.

Dumay, X., & Dupriez, V. (2008). Does the school composition effect matter? Evidence from Belgian data. *British Journal of Educational Studies, 56*, 440–477.

Dumay, X., Dupriez, V., & Maroy, C. (2010). Ségrégation entre écoles, effets de la composition scolaire et inégalités de résultats. *Revue Française de Sociologie, 51*, 461–480.

Dumay, X., & Galand, B. (2012). The multilevel impact of transformational leadership on teacher commitment: Cognitive and motivational pathways. *British Educational Research Journal, 38*, 703–729.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–27.

Elchardus, M., & Kavadias, D. (2000, July). *The socializing effect of educational networks. The relevance of the distinction between public and private schools with the relation to the non-cognitive outcomes of the last-year pupils in Flanders (Belgium)*. Paper presented at the International Sociological Association Midterm Conference RC04 Education: Outcomes and Governance of Schooling, Groningen, The Netherlands.

Gadeyne, E., Ghesquière, P., & Onghena, P. (2006). Psychosocial educational effectiveness criteria and their relation to teaching in primary education. *School Effectiveness and School Improvement, 17*, 63–85.

Goddard, R. D., Tschannen-Moran, M., & Hoy, W. K. (2001). A multilevel examination of the distribution and effects of teacher trust in students and parents in urban elementary schools. *The Elementary School Journal, 102*, 3–17.

Goldstein, H., & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: A cross-classified multilevel analysis. *School Effectiveness and School Improvement, 8*, 219–230.

Guldemond, H., & Bosker, R. J. (2009). School effects on students' progress – A dynamic perspective. *School Effectiveness and School Improvement, 20*, 255–268.

Harter, S. (2012). *The construction of the self: Developmental and sociocultural foundations* (2nd ed.). New York, NY: Guilford Press.

Heers, M., Ghysels, J., Groot, W., & Maassen van den Brink, H. (2015). Differentiated effects of community schooling on cognitive and social-emotional learning outcomes. *School Effectiveness and School Improvement, 26*, 354–381.

Herweijer, L. (2009). *Making up the gap: Migrant education in the Netherlands*. The Hague, The Netherlands: SGP.

Hofman, R. H., de Boom, J., Meeuwisse, M., & Hofman, W. H. A. (2013). Educational innovation, quality and effects: An exploration of innovations and their effects is in secondary education. *Education Policy, 27*, 843–866.

Hofman, R. H., Hofman, W. H. A., & Gray, J. M. (2015). Three conjectures about school effectiveness: An exploratory study. *Cogent Education, 2*(2), 1–13.

Hofman, R. H., Hofman, W. H. A., & Guldemond, H. (1999). Social and cognitive outcomes: A comparison of contexts of learning. *School Effectiveness and School Improvement, 10*, 352–366.

Hofman, R. H., Hofman, W. H. A., & Guldemond, H. (2001). The effectiveness of cohesive schools. *International Journal of Leadership in Education, 4*, 115–135.

Hofman, R. H., Hofman, W. H. A., & Guldemond, H. (2002). School governance, culture, and achievement. *International Journal of Leadership in Education, 5*, 249–272.

Hofman, R. H., Hofman, W. H. A., & Guldemond, H. (2003). Effective families, peers, and schools. A configuration approach. *Educational Research and Evaluation, 9*, 213–237.

Hofman, W. H. A. (1994). School effects on performance of minority pupils. *School Effectiveness and School Improvement, 5*, 26–44.

Hofman, W. H. A. (1995). Cross-level relationships within effective schools. *School Effectiveness and School Improvement, 6*, 146–174.

Hofman, W. H. A., & Hofman, R. H. (2011). Smart management in effective schools: Effective management configurations in general and vocational education in the Netherlands. *Educational Administration Quarterly, 47*, 620–645.

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2015a). Does classroom composition make a difference: Effects on developments in motivation, sense of classroom belonging, and achievement in upper primary school. *School Effectiveness and School Improvement, 26*, 125–152.

Hornstra, L., van der Veen, I., Peetsma, T., & Volman, M. (2015b). Innovative learning and developments in motivation and achievement in upper primary school. *Educational Psychology, 35*, 598–633.

Hulpia, H., Devos, G., & Rosseel, Y. (2009). The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organisational commitment. *School Effectiveness and School Improvement, 20*, 291–317.

Karssen, M., van der Veen, I., & Volman, M. (2016). Ethnic diversity in schools and bi-ethnic Dutch students' educational outcomes and social functioning. *School Effectiveness and School Improvement, 27*, 613–628.

Keuning, T., van Geel, M., Visscher, A., & Fox, J.-P. (2016, March). *The effects of a schoolwide data-based decision making intervention on elementary schools' student achievement growth for mathematics and spelling*. Paper presented at the Society for Research on Educational Effectiveness (SREE) Conference, Washington, DC. (ED567600)

Knuver, A. W. M., & Brandsma, H. P. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement, 4*, 189–204.

Kruger, M. L., Witziers, B., & Sleegers, P. (2007). The impact of school leadership on school level factors: Validation of a causal model. *School Effectiveness and School Improvement, 18*, 1–20.

Kyriakides, L., & Creemers, B. P. M. (2008). A longitudinal study on the stability over time of school and teacher effects on student learning outcomes. *Oxford Review of Education, 34*, 521–545.

Lietaert, S., Roorda, D., Laevers, F., Verschueren, K., & De Fraine, B. (2015). The gender gap in student engagement: The role of teachers' autonomy support, structure, and involvement. *British Journal of Educational Psychology, 85*, 498–518.

Lomos, C., Hofman, R. H., & Bosker, R. J. (2011). The relationship between departments as professional communities and student achievement in secondary education. *Teaching and Teacher Education, 27*, 722–731.

Luyten, H. (1994). Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research, 21*, 197–216.

Luyten, H. (2004). Succes in het voortgezet onderwijs: Capaciteiten, inzet of achtergrond? [Success in secondary education: Abilities, effort or background?]. *Pedagogische Studiën, 81*, 155–166.

Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: Regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32*, 397–429.

Luyten, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20*, 145–165.

Luyten, H., & Van der Hoeven-Van Doornum, A. (1995). Classroom composition and individual achievement. Effects of classroom composition and teacher goals in Dutch elementary education. *Tijdschrift voor Onderwijsresearch, 20*, 42–62.

Maslowski, R., Scheerens, J., & Luyten, H. (2007). The effect of school autonomy and school internal decentralization on students' reading literacy. *School Effectiveness and School Improvement, 18*, 303–334.

Maulana, R., Opdenakker, M.-C., & Bosker, R. (2016). Teachers' instructional behaviors as important predictors of academic motivation: Changes and links across the school year. *Learning and Individual Differences, 50*, 147–156.

Maulana, R., Opdenakker, M.-C., Stroet, K., & Bosker, R. (2013). Changes in teachers' involvement versus rejection and links with academic motivation during the first year of secondary education: A multilevel approach. *Journal of Youth and Adolescence, 42*, 1348–1371.

Meelissen, M. R. M., & Luyten, H. (2011). Schooleffectiviteit en prestatieniveau natuuronderwijs in groep 6: Secundaire analyses op TIMSS-2007 data [School effectiveness and science achievement in grade 4: Secondary analyses on TIMSS-2007 data]. *Pedagogische Studiën, 88*, 309–322.

Meelissen, M. R. M., & Luyten, J. W. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation, 34*, 82–93.

Meijnen, G. W., Lagerweij, N. W., & de Jong, P. F. (2003). Instruction characteristics and cognitive achievement of young children in elementary schools. *School Effectiveness and School Improvement, 14*, 159–187.

Minnaert, A., Boekaerts, M., de Brabander, C., & Opdenakker, M.-C. (2011). Students' experiences of autonomy, competence, social relatedness and interest within a CSCL environment in vocational education: The case of commerce and business administration. *Vocations and Learning, 4*, 175–190.

Opdenakker, M.-C. (2004). *Leerling in Wonderland? Een onderzoek naar het effect van leerling-, lesgroep-, leerkracht- en schoolkenmerken op prestaties voor wiskunde in het secundair onderwijs* [Student in Wonderland? A study on the effect of student, class, teacher, and school characteristics on mathematics achievement in secondary education]. Leuven, Belgium: Acco.

Opdenakker, M.-C. (2013, April). *Effects on students' motivation: Does the pedagogical approach of traditional, social-constructivist and combined-approach schools matter?* Paper presented at the annual meeting of the AERA, San Francisco, CA.

Opdenakker, M.-C. (2015, August). *Teacher support versus thwart: Their importance to motivation, self-regulation and need satisfaction*. Paper presented at the EARLI conference, Limassol, Cyprus.

Opdenakker, M.-C., & Maulana, R. (2010, April). *Teacher-student relationships and academic engagement: How do they develop and link?* Paper presented at the International Conference on Interpersonal Relationships in Education (ICIRE), Boulder, CO.

Opdenakker, M.-C., Maulana, R., & den Brok, P. (2012). Teacher-student interpersonal relationships and academic motivation within one school year: Developmental changes and linkage. *School Effectiveness and School Improvement, 23*, 95–119.

Opdenakker, M.-C., & Minnaert, A. (2011). Relationship between learning environment characteristics and academic engagement. *Psychological Reports, 109*, 259–284.

Opdenakker, M.-C., & Minnaert, A. (2014). Learning environment experiences in primary education: Their importance to academic engagement. In D. Zandvliet, P. den Brok, T. Mainhard, & J. van Tartwijk (Eds.), *Interpersonal relationships in education: From theory to practice* (pp. 183–194). Rotterdam, The Netherlands: Sense.

Opdenakker, M.-C., & Van Damme, J. (2000a). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement, 11*, 165–196.

Opdenakker, M.-C., & Van Damme, J. (2000b). The importance of identifying levels in multilevel analysis: An illustration of the effects of ignoring the top or intermediate levels in school effectiveness research. *School Effectiveness and School Improvement, 11*, 103–130.

Opdenakker, M.-C., & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal, 27*, 407–432.

Opdenakker, M.-C., & Van Damme, J. (2006a). Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools. *School Effectiveness and School Improvement, 17*, 87–117.

Opdenakker, M.-C., & Van Damme, J. (2006b). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and Teacher Education, 22*, 1–21.

Opdenakker, M.-C., & Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal, 33*, 179–206.

Opdenakker, M.-C., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematics achievement. *School Effectiveness and School Improvement, 13*, 399–427.

Opdenakker, M.-C., Van Damme, J., & Minnaert, A. (2002a, January). *Are there equal opportunities in our secondary schools? An investigation of the relationship between class composition, indicators of instructional support, effort, and academic achievement*. Paper presented at the International Congress for School Effectiveness and Improvement, Kopenhagen, Denmark.

Opdenakker, M.-C., Van Damme, J., & Minnaert, A. (2002b). *Groepssamenstelling en de klaspraktijk: een onderzoek naar de relatie tussen de beide, ook in hun effect op inzet en schoolse prestaties* (LOA-rapport nr. 9) [Group composition and class practice: An investigation into their relation and their effect on effort and academic achievement]. Leuven, Belgium: Steunpunt Loopbanen doorheen Onderwijs naar Arbeidsmarkt, Cel voor Analyse van Onderwijsloopbanen.

Opdenakker, M.-C., Van Damme, J., & Minnaert, A. (2005). *Are there equal opportunities in our classes and schools? An investigation into the relationship between class composition, indicators of the learning environment and the class climate, effort, and achievement of classes*. Leuven, Belgium: University of Leuven.

Opdenakker, M. C. J. L. (2014). Leerkracht-leerlingrelaties vanuit een motivationeel perspectief: het belang van betrokken en ondersteunende docenten [Teacher-student relations from a motivational perspective: The importance of involved and supportive teachers]. *Pedagogische Studien, 91*, 332–351.

Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement, 17*, 275–302.

Pinxten, M., De Fraine, B., Van Den Noortgate, W., Van Damme, J., & Anumendem, D. (2012). Educational choice in secondary school in Flanders: The relative impact of occupational interests on option choice. *Educational Research and Evaluation, 18*, 541–569.

Pustjens, H., Van de gaer, E., Van Damme, J., & Onghena, P. (2004). Effect of secondary schools on academic choices and on success in higher education. *School Effectiveness and School Improvement, 15*, 281–311.

Pustjens, H., Van de gaer, E., Van Damme, J., & Onghena, P. (2008). Curriculum choice and success in the first two grades of secondary education: Students, classes, or schools? *School Effectiveness and School Improvement, 19*, 155–182.

Pustjens, H., Van de gaer, E., Van Damme, J., Onghena, P., & Van Landeghem, G. (2007). The short- and long-term effects of primary schools and classes on mathematics and language achievement scores. *British Educational Research Journal, 33*, 419–440.

Reezigt, G.J., & Weide, M.G. (1990, April). *The effects of group-based mastery learning on language and arithmetic achievement and attitudes in primary education in The Netherlands*. Paper presented at the annual meeting of the AERA. Boston, MA. (ERIC ED317584).

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness (EER): A state-of-the-art- review. *School Effectiveness and School Improvement, 25*, 197–230.

Roeleveld, J., & Dronkers, J. (1994). Bijzondere scholen of buitengewone scholen? Verschillen in effectiviteit van openbare en confessionele scholen in regio's waarin hun richting een meerderheids- of minderheidspositie inneemt [Denominational schools or exceptional schools?

Differences in effectiveness according to the majority or minority position of the denomination in the region]. *Mens en Maatschappij, 69*(1), 85–108.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist, 55*, 68–78.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York, NY: Guilford Press.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*, 61–80.

Scheerens, J. (1992). *Effective schooling: Research theory and practice*. London, UK/New York, NY: Cassell.

Scheerens, J. (with the cooperation of Luyten, H., Steen, R., & Luyten-de Thouars, Y.). (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede, The Netherlands: Department of Educational Organization and Management, University of Twente.

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: A critical review of the knowledge base*. Dordrecht, The Netherlands: Springer.

Scheerens, J. (2017). The perspective of "limited malleability" in educational effectiveness: Treatment effects in schooling. *Educational Research and Evaluation, 23*, 247–266.

Scheerens, J., & Creemers, B. P. M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research, 13*, 691–706.

Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2015). What motivates early adolescents for school? A longitudinal analysis of associations between observed teaching and motivation. *Contemporary Educational Psychology, 42*, 129–140.

Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2016). Fostering early adolescents' motivation: A longitudinal study into the effectiveness of social constructivist, traditional and combined schools for prevocational education. *Educational Psychology, 36*, 1–25.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2012). *Effective pre-school, primary and secondary education 3–14 Project (EPPSE 3–14). Final report from the Key Stage 3 phase: Influences on students' development from age 11–14*. London, UK: DFE/Institute of Education.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation, 22*, 561–582.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London, UK: Falmer Press.

ten Bruggencate, G., Luyten, H., Scheerens, J., & Sleegers, P. (2012). Modeling the influence of school leaders on student achievement: How can school leaders make a difference? *Educational Administration Quarterly, 48*, 699–732.

Timmermans, A. C., Bosker, R. J., de Wolf, I. F., Doolaard, S., & van der Werf, M. P. C. (2014). Value added based on educational positions in Dutch secondary education. *British Educational Research Journal, 40*, 1057–1082.

van Batenburg, T. A., & Lokman, I. H. (1991, January). *Effectiveness in senior secondary vocational education*. Paper presented at the annual meeting of the ICSEI. Cardiff, Wales, UK. (ERIC ED358364).

Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: An introduction. *School Effectiveness and School Improvement, 13*, 383–397.

Van Damme, J., Opdenakker, M.-C., & Van den Broeck, A. (2004, May). *Do classes and schools have an effect on attitudes towards mathematics?* Paper presented at the first IEA international research conference, Lefkosia, Cyprus.

Van Damme, J., Van Landeghem, G., De Fraine, B., Opdenakker, M.-C., & Onghena, P. (2000). *Maakt de school het verschil? Effectiviteit van scholen, leraren en klassen in de eerste graad van het middelbaar onderwijs* [Does the school make the difference? Effectiveness of schools, teachers, and classes in the first cycle of secondary education]. Leuven, Belgium: K.U.Leuven, Department Pedagogische Wetenschappen, LIVO.

Van de gaer, E., De Fraine, B., Pustjens, H., Van Damme, J., De Munter, A., & Onghena, P. (2009). School effects on the development of motivation toward learning tasks and the development of academic self-concept in secondary education: A multivariate latent growth curve approach. *School Effectiveness and School Improvement, 20*, 235–253.

Van de gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2006). The gender gap in language achievement: The role of school-related attitudes of class groups. *Sex Roles, 55*, 397–408.

Van de gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2007). Impact of attitudes of peers on language achievement: Gender differences. *Journal of Educational Research, 101*, 78–92.

van de Grift, W. (1987). Zelfpercepties van onderwijskundig leiderschap en gemiddelde leerlingprestaties [Self-perceptions of educational leadership and mean student achievement]. In J. Scheerens & W. G. R. Stoel (Eds.), *Effectiviteit van onderwijsorganisaties* (pp. 83–98). Lisse, The Netherlands: Swets & Zeitlinger.

van de Grift, W., & Houtveen, A. A. M. (1999). Educational leadership and pupil achievement in primary education. *School Effectiveness and School Improvement, 10*, 373–389.

Van den Broeck, A., Opdenakker, M.-C., & Van Damme, J. (2005). The effects of student characteristics on mathematics achievement in Flemish TIMSS 1999 data. *Educational Research and Evaluation, 11*, 107–121.

Van den Noortgate, W., Opdenakker, M.-C., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement, 16*, 281–303.

van der Slik, F. W. P., Driessen, G., & De Bot, K. L. J. (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudinal multilevel examination in Dutch primary schools. *European Sociological Review, 22*, 293–308.

van der Werf, G. (1997). Differences in school and instruction characteristics between high-, average-, and low-effective schools. *School Effectiveness and School Improvement, 8*, 430–448.

van der Werf, G., Opdenakker, M.-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel growth curve approach. *School Effectiveness and School Improvement, 19*, 447–462.

Van Gasse, R., Vanlommel, K., Vanhoof, J., & Van Petegem, P. (2017). The impact of collaboration on teachers' individual data use. *School Effectiveness and School Improvement, 28*, 489–504.

Van Houtte, M. (2004). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies, 30*, 159–173.

Van Landeghem, G., Van Damme, J., Opdenakker, M.-C., De Fraine, B., & Onghena, P. (2002). The effect of schools and classes on noncognitive outcomes. *School Effectiveness and School Improvement, 13*, 429–451.

Van Petegem, K., Aelterman, A., Rosseel, Y., & Creemers, B. P. M. (2007). Student perception as moderator for student wellbeing. *Social Indicators Research, 83*, 447–463.

Van Petegem, K., Aelterman, A., Van Keer, H., & Rosseel, Y. (2008). The influence of student characteristics and interpersonal teacher behaviour in the classroom on student's wellbeing. *Social Indicators Research, 85*, 279–291.

Van Petegem, K., Creemers, B., Aelterman, A., & Rosseel, Y. (2008). The importance of pre-measurements of wellbeing and achievement for students' current wellbeing. *South African Journal of Education, 28*, 451–468.

Vandecandelaere, M., Vanlaar, G., De Fraine, B., Van Damme, J., & Verhaeghe, J.-P. (2011). *Invloed van de etnische en sociale samenstelling en diversiteit van de schoolpopulatie op de prestaties en leerwinst voor wiskunde in het eerste leerjaar* [The influence of the ethnic and social composition and the diversity of the school population on mathematic achievement and learning gain in the first grade of primary education] (No. SSL/OD1/2010.31). Leuven, Belgium: Steunpunt Studie- en Schoolloopbanen.

Vanlaar, G., Denies, K., Vandecandelaere, M., Van Damme, J., Verhaeghe, J. P., Pinxten, M., & Fraine, D. (2014). How to improve reading comprehension in high-risk students: Effects of class practices in grade 5. *School Effectiveness and School Improvement, 25*, 408–432.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal-contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*, 19–31.

Vansteenkiste, M., Soenens, B., Sierens, E., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology, 101*, 671–688.

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017a). Long-term effects of primary schools on non-cognitive outcomes of students at age 17. *Studies in Educational Evaluation, 55*, 83–93.

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017b). Long-term effects of primary schools on educational positions of students 2 and 4 years after the start of secondary education. *School Effectiveness and School Improvement, 28*, 167–190.

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017c). Long-term effects of primary schools on mathematics achievement of students at age 17. *British Educational Research Journal, 43*, 1131–1148.

Vause, A., Dupriez, V., & Dumay, X. (2010). L'efficacité différence des pratique pédagogiques sur les performances en mathématiques des élèves de sixième primaire en belgique francophone [The difference in effectiveness of the teaching practices on the performances in mathematics of raised of sixth primary education in French-speaking Belgium]. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportment, 42*, 234–246.

Veenman, G.-J. M., Van de Werfhorst, H. G., & Dronkers, J. (2013). Ethnic composition of the class and educational performance in elementary education in the Netherlands. *Educational Research and Evaluation, 19*, 370–401.

Verachtert, P., Van Damme, J., Onghena, P., & Ghesqiere, P. (2009). A seasonal perspective on school effectiveness: Evidence from a Flemish longitudinal study in kindergarten and first grade. *School Effectiveness and School Improvement, 20*, 215–233.

Vermeulen, C. J. (1987). De effectiviteit van onderwijs bij zeventien Rotterdamse stimuleringsscholen [Educational effectiveness in seventeen educational priority schools in Rotterdam]. *Pedagogische Studiën, 64*, 49–58.

Vervaet, R., Van Houtte, M., & Stevens, P. A. J. (2018). Multicultural teaching in Flemish secondary schools: The role of ethnic school composition, track and teachers' ethnic prejudice. *Education and Urban Society, 50*, 274–299.

Weide, M. G. (1993). Effective provisions for ethnic minorities in primary schools. *Tijdschrift voor Onderwijsresearch, 18*(2), 74–84.

Witziers, B., & Bosker, R. (1997, January). *A meta-analysis on the effects of presumed school effectiveness enhancing factors.* Paper presented at the international congress for school effectiveness and improvement, Memphis, TN.

Witziers, B., Bosker, R., & Kruger, M. (2003). Educational leadership and student achievement; the elusive search for an association. *Educational Administration Quarterly, 39*, 398–425.

Chapter 11
The Impact of Socioeconomic Segregation in U.S. High Schools on Achievement, Behavior, and Attainment and the Mediating Effects of Peers and School Practices

Gregory J. Palardy

11.1 Introduction

The socioeconomic composition (SEC) of a school is the average socioeconomic status (SES) of the enrolled students. It is believed to be a proximal measure of the normative and cultural environment of the school. As such, it may impact students' attitudes about the value of schooling, appropriate behaviors, motivation levels, aspirations, etc., and it may also influence school personnel's beliefs and practices (Coleman et al., 1966; Hallinan & Williams, 1990; McDonough, 1997; Palardy, 2015b). It is often conceptualized as a proxy measure for student socioeconomic segregation in that low- and high-SEC schools have concentrated poverty and concentrated affluence, respectively. To the degree that socioeconomic segregation in schools creates inequitable educational contexts, it is an important policy issue. A somewhat extensive body of research exists indicating SEC has a significant positive association with student achievement and learning (for a recent meta-analysis see Van Ewijk & Sleegers, 2010). However, there is much less research examining the associated between SEC and other important educational outcomes, such as student engagement, high school graduation, and college enrollment, and there is no research comparing the relative effects of SEC across a range of high school and post-secondary outcomes.

The knowledge gap resulting from this overwhelming focus on achievement outcomes limits the scope our understanding of the impact SEC has on students. Indeed, the limitations of achievement outcomes has been noted to the larger educational research community and has resulted in a greater use of multiple and alternative outcomes to better capture the breadth of school effects and educational interventions. While school segregation is a concern in many nations, it is particularly concerning in the United States due to its long history and because, among developed nations, it currently has one of the highest levels of income inequity and

G. J. Palardy (✉)
Graduate School of Education, University of California, Riverside, CA, USA
e-mail: gjpalardy@gmail.com

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_11

largest poverty gaps after adjusting for tax redistribution (OECD, 2019). These nation-wide measures of income inequality suggest that any negative effects SEC are expected to be more prevalent in the U.S. compared with other OECD nations.

The chapter begins with a historical overview of segregation in U.S. schools, which is followed by a synopsis of a recent trend in U.S. educational research towards using alternative and multiple outcomes in addition to achievement test scores. That is followed by the empirical study of the effects of SEC on multiple outcomes and a discussion of the implications of the findings to policy and practice.

11.1.1 Historical Background on School Segregation in the United States

School segregation and its negative consequences on equality of educational opportunity has been an enduring issue throughout the history of the United States. This section provides an overview of the history of segregation in the United States and its effects on educational equality. That history began with the segregation of Blacks, especially in the South. However, over the past half century several factors led to a shift from a focus on racial integration to socioeconomic integration.

Historical repression of Blacks. Segregation in the United States is grounded in its deplorable history of Black slavery. By 1860, just prior to the Civil War, slaves made up approximately a third of the population of Southern United States. Slaves, of course, did not have basic citizenship rights. After the Civil War, the 14th Amendment to the U.S. Constitution provided citizenship rights and equal protection under the law to former slaves. However, the spirit of the 14th Amendment was largely circumvented from the beginning as Southern states quickly adopted what became known as the "separate, but equal" doctrine, whereas states provide separate facilities for Blacks in order to comply with the 14th Amendment, while also maintaining segregation in almost all social, educational, and government institutions.

Plessy v. Ferguson (1896), a carefully orchestrated challenge of the constitutionality of the separate but equal doctrine, was rejected by the U.S. Supreme Court. This federal decision meant that racial segregation was legal in all facets of society, including schools, as long as equal facilities were provided. While *Plessy v. Ferguson* is often cited as one of the worst Supreme Court rulings in U.S. history, it was never overturned. However, the historic *Brown v. Board of Education* (1954) U.S. Supreme Court ruling redressed *Plessy* in the educational context, declaring, "In the field of public education, the doctrine of 'separate but equal' has no place," (Brown v. Board of Education, 1954).

While *Brown vs. Board of Education* abolished racial segregation of schools under the law, de facto racial segregation continued to exist due in large part to actual structural barriers such as neighborhood segregation. However, artificial structural barriers, such as where school catchment boundaries were placed and where new schools were built, also contributed to school segregation. It was not until the Civil

Rights Act of 1964 that federal policies requiring schools to desegregate led to a substantial reduction in school racial segregation, particularly in the South, where it was most prevalent.

As a result of the Civil Rights Act, racial segregation in U.S. schools was pronouncedly reduced over the next 25 years. For example, from the mid-1960s to the late 1980s, the percentage of African American students attending majority-White schools in the South increased by a factor of 22 (from 2% to 44%), resulting in Southern schools being the most integrated of any region in the United States (Orfield, 2005).

11.1.1.1 Shift in School Integration Efforts from Race to SES

Since the Civil Rights Act of 1964, there has been a gradual shift in school integration efforts from a focus on race to a focus on SES. There are a number of reasons for this. The shift began with the publication of the Coleman Report (Coleman et al., 1966), a massive study commissioned by the United States Congress at the height of the civil rights era to uncover school factors that undermined educational inequality. Coleman and his colleagues found that SEC was the school factor with the strongest association with student achievement. This finding was validated by numerous subsequent studies (Borman & Dowling, 2010; Jencks & Mayer, 1990; McDonough, 1997; Palardy, 2008; Perry & McConney, 2010; Rumberger & Palardy, 2005a; Willms, 1986, 2010). Furthermore, several of the subsequent works found that the racial composition of the school is not associated with student achievement once SEC is statistically controlled. That is, on average, students attending low-SEC schools tend to underperform regardless of their ethnicity or race. This suggests that SEC is the primary student body compositional effect driving student underperformance, rather than racial composition.

The shift from racial to SES integration is also partially the result of changes in racial demographics since the 1960s, especially rapid growth in the Latino population over the past two decades. By 2015 the national K-12 public school population was 26% Latino, 15% African American children (down slightly from the past), and 49% White (down considerably from the past). This shifting demographics led to fewer majority-White schools. Moreover, the Civil Rights Act's focus on racial desegregation was largely due to historical racism towards African Americans in the United States. While Latinos face some of the same educational challenges as African Americans (e.g., fewer educational resources, lower quality schools, low achievement and graduation rates), they mostly immigrated in recent generations and did not experience a comparable level of multigenerational repression. However, low socioeconomic status is a condition that both Latinos and African Americans are overrepresented. For example, poverty rates in 2015 among African American and Hispanic children were 3.2 and 2.7 times higher than Whites (38%, 32%, and 12%, respectively) (Snyder, de Brey, & Dillow, 2016).

The shift towards a focus on school SEC is also due in part perceptions by some members of the African American community that racial integration is not in their best interests. Perhaps at the heart of the issue is a lack of trust in schools among

some African American families, particularly White-majority schools, and perceptions that such schools do not sufficiently serve their children's best interests or value their talents (Bryk & Schneider, 2002; Ladson-Billings, 2006). A related issue is why it is assumed that African American children need to be integrated into White schools in order to maximize their academic performance, as that notion can be taken to imply the offending supposition that it is to overcome a cultural deficit (Ladson-Billings, 2006).

Another reason for the shifting focus towards SEC is because in recent years neighborhoods in U.S. cities have become increasingly segregated based on socioeconomics rather than race (Owens, 2018). While this change began 40 years ago, it has accelerated over the past 15 years. This is an issue because approximately 75% of public school students attend their neighborhood school (Grady & Bielick, 2010) and because, as many scholars have conceptualized, neighborhoods serve as an educational resource that can influence a range of educational and wellbeing outcomes (Ainsworth, 2002; Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993; Garner & Raudenbush, 1991; Jencks & Mayer, 1990; Mendenhall, DeLuca, & Duncan, 2006; Owens, 2018; Sampson, Morenoff, & Gannon-Rowley, 2002). Furthermore, affluent families of all races have the economic capital to move to the more education-supportive locations, and they generally do (Owens, 2018).

It is also worth noting that while racial composition and socioeconomic composition of schools are correlated, that correlation has gradually weakened over the past 30 years. The correlation between SEC and the percentage of the students attending a school who are African American or Hispanic has declined from −0.52 in 1990, to −0.29 in 2002, to −0.14 in 2009, based on national samples of high schools (Palardy, 2013).[1] This means the degree of racial integration in schools has become much more similar across the range of school SECs, as opposed to the previous clear pattern of relatively lower racial integration in high-SEC schools and relatively higher racial integration in low-SEC schools.

Another reason for the shift towards a focus on SEC integration is research suggests that socioeconomic achievement gaps are now larger than racial achievement gaps (Reardon and Bischoff, 2011). In addition, socioeconomic segregation has a stronger detrimental impact on student learning and attainment (e.g., high school graduation and college enrollment) than does racial segregation (Palardy, 2013; Rumberger & Palardy, 2005a). Both of these factors suggest that SEC rather than racial segregation, is the primary form of school segregation contributing to students' educational challenges and underperformance.

In summary, for a number of well-founded reasons, the problem of school segregation in U.S. schools has gradually shifted over the past 50 years from being mostly a racial issue to mostly a socioeconomic issue.

[1]These figures were derived from three consecutive national education surveys of U.S. high schools, including NELS:88, ELS: 2002, and HSLS: 2009. The first two figures are cited in Palardy (2013). The 2009 figure was estimated for the current paper.

11.1.1.2 Towards the Use of Multiple and Alternative Outcomes and Not Just Achievement

Over the past 15 years, there has been a trend in U.S. educational effectiveness research (EER) towards the use of multiple outcome measures and a reduction in the use of achievement outcomes. Some outcomes that have become more common are attainment (e.g., high school graduation, college enrollment, etc.), social and emotional development, student engagement, and other non-cognitive measures. This trend in EER outcomes is apparent in the empirical studies published in academic journals. For example, Table 11.1 shows the outcomes of studies published in two prominent educational research journals based in the United States, *Educational Evaluation and Policy Analysis* and *American Education Research Journal*, in 2003 and 2017. Over this period, the use of achievement outcomes decreased 16%, whereas the use of attainment and school behavior outcomes (e.g., attendance, misconduct, attitudes, and social and emotional skills) increased 24.4%. There was also a substantial increase (20.6%) in the percentage of empirical studies that used multiple outcomes rather than a single outcome.

11.1.1.3 Linking Research Trends to Policy and Practice

The changes in research outcomes summarized in Table 11.1 are in part a response to educational policy in the U.S. The dominant federal education policy in the United States for the past two generation has been the Elementary and Secondary Education Act (ESEA) which was initially passed by the U.S. Congress in 1965 as part of Lyndon Johnson's Great Society legislation. ESEA is updated and reauthorized periodically, with the new version often being the current presidential administration's signature educational policy and typically receiving enormous publicity in the media and attention in the research community. The two most recent updates are the No Child Left Behind Act (NCLB, 2001) and the Every Student Succeeds Act (ESSA, 2015). The NCLB act focused on raising student achievement, reducing achievement gaps, and increasing school accountability for those outcomes. It led to

Table 11.1 Changes in outcomes in studies published in AERJ and EEPA, 2003–2017

Year	Achievement	Attainment/behavior/attitude/ social-emotional	Other	Multiple outcomes
2003 (post NCLB)	16/26 (61.5%)	2/26 (7.7%)	11/26 (43.3%)	3/26 (11.5%)
2017 (post ESSA)	24/53 (45.3%)	17/53 (32.1%)	27/53 (50.9%)	17/53 (32.1%)
Change (2003–17)	−16.2%	+24.4%	+7.6%	+20.6%

Note: Only empirical studies using quantitative outcomes were tallied (i.e., review, conceptual, and qualitative studies are not included). Special issues were also omitted, due to the typical requirement that all articles focus on the same topic

an increase in research focused on achievement, achievement gap outcomes, and quantitative models for estimating school effectiveness based on gains in achievement test scores. This research contributed to our understanding of the limitations of achievement test scores for informing educational policy and practice and for gauging teacher and school performance (Darling-Hammond, 2004; Hanushek & Raymond, 2005; Linn, 2003). In comparison, the latest version of the ESEA, titled the *Every Student Succeeds* Act (ESSA, 2015), altered the accountability provision to place a greater emphasis on the use of outcomes other than achievement test scores, such as absenteeism rates, graduation rates, college readiness, and college attendance. Predictably, ESSA has led to a shift in the focus of educational research to reflect the new emphasis of ESSA, which is apparent in Table 11.1.[2]

11.1.1.4 The Importance of Diversity in Research Outcomes

Public schools are civic institutions that serve the function of developing children into productive members of society. While developing academic skills is a primary component of that general function, most people agree that there are other important components. For example, Rothstein and Jacobsen (2006) came up with a list of eight functions of public schools and asked school district superintendents, school board members, state legislators, and other adult members of society to consider the value of each as a percentage of the total function of public schools. The authors' list of functions included core academic skills, critical thinking, social skills, civics, employment skills, physical health, mental health, and arts. There was a surprisingly high level of parity in responses across the eight functions, with no function having below 9% of the total and only core academic skills (22%) and critical thinking (18%) having above 12%. This parity suggests that well-informed members of society believe that schools serve a range of civic functions beyond academics. Research is needed to assess the effective educational policies and practices for promoting each of these functions.

Another reason for using multiple outcomes is that schools that are effective based on one outcome are not necessarily effective based on others. For example, Rumberger and Palardy (2005b) examined school performance based on three different outcomes (achievement growth, graduation rates, and student mobility rates) and found the correlations between were all below 0.10. An extension of this concept is that school practices that impact achievement do not necessarily impact other pertinent outcomes. Both of these notions show that a single outcome presents an incomplete picture of a school's effectiveness, particularly if one agrees that schools serve multiple functions.

[2]Table was constructed to test the hypothesis of whether research outcomes tend to change to reflect the federal policy emphases. The year 2003 was selected for the first year because it is 2 years after the *NCLB* act was enacted and 2017 was selected for the second period because it is 2 years after the enactment of the *ESSA*. The two-year lag is used because research typically takes at least a year to conduct and publish.

Another reason for using a diverse set of outcomes, especially when evaluating effectiveness, is that a single outcome is more readily manipulated, which can create a perverse incentive. As Don Campbell (1979) observed, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (p. 85). In the context of school effectiveness, these distortions are not typically due to outright cheating, but rather subtle manipulations such as "teaching to the test" or more generally limiting the curricular foci of the school to the evaluation criteria. Another example of this manipulation is schools may pressure low-performing students to transfer or dropout in an effort to raise the school's average achievement. An addendum to "Campbell's Law" may be, the more diverse the set of indicators of overall performance, the more difficult it is to manipulate and distort estimates of overall performance.

11.1.1.5 Brown vs. Board of Education Revisited

It is important to note that the *Brown* decision was not based solely on the notion that segregated schools compromised equality of opportunity pertaining to academics, but also on the idea that isolating children based on personal traits like skin color has lasting negative social and emotional consequences. The *Brown* decision states:

> To separate [African American children] from others of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hearts and minds in a way unlikely ever to be undone. (Brown v. Board of Education, 1954)

Research on this type of potential effect of school segregation requires measures of students' attitudes, aspirations about education, and their level of engagement, and other school behaviors rather than just academic achievement. This is yet another reason the use of multiple outcomes is important for understanding the impact of school segregation.

11.2 Empirical Study

This section presents the results of an empirical study on the effects of socioeconomic segregation in schools on several outcomes and the degree to which peer influence and school practices mediate those effects. An emphasis is placed on comparing effect size and mediating mechanisms across outcomes. The section begins with an overview of the research and theory on socioeconomic segregation in schools. This is followed by the research questions, descriptions of the data source and methods, and the results of the analyses.

11.2.1 The Effect of SEC on Achievement and Other Outcomes

The effect of socioeconomic segregation in schools on student achievement has been studied extensively over the past 50 years (Borman & Dowling, 2010; Jencks & Mayer, 1990; Kahlenberg, 2001; Palardy, 2008; Perry & McConney, 2010; Rumberger & Palardy, 2005a; Van Ewijk & Sleegers, 2010; Willms, 1986, 2010). Dissimilarity indices are most commonly used to measure degree of segregation among groups, such as racial groups. However, because SES is a continuous variable and there are no inherent groupings, socioeconomic segregation is typically based of the composition of the student body, measured using the mean SES of students attending the school. This measure is referred to as the socioeconomic composition (SEC) of the school in this study. Schools with high or low values on SEC are segregated in that they have high concentrations of high- or low-SES students. Like other measures of school composition used to study schools, its effect on student outcomes is conceptualized as being above and beyond the corresponding student characteristics—in this case, family SES (Raudenbush & Bryk, 2002). In other words, family SES must be controlled for in the statistical models to obtain accurate estimates of SEC. If SES is not controlled in the model, its effect is subsumed by the SEC effect estimate and the SEC effect is inflated. Research has consistently found positively associations between SEC and student achievement and learning in reading, math, and other content areas (Van Ewijk & Sleegers, 2010). While other measures of student body composition have been studied (Bryk & Thum, 1989; McNeal, 1997; Rumberger & Palardy, 2005a), SEC has been studied the most and is generally considered to have the most robust associations with student achievement. For example, research suggests that SEC accounts for most of the effects of racial composition on achievement (Coleman et al., 1966; Jencks & Mayer, 1990; Rumberger & Palardy, 2005a).

Although a large number of studies have examined the effect of SEC on student achievement, relatively few have studied its effect on non-achievement outcomes. For example, only a few studies examined the association between SEC and high school graduation (Jencks & Mayer, 1990; McNeal, 1997; Murnane, 1981; Palardy, 2013), and even fewer have focused on the effect of SEC on college enrollment (Coleman & Hoffer, 1987; Engberg & Wolniak, 2010; Hill, 2008; Perna & Titus, 2005) or college choice (McDonough, 1997; Palardy, 2015b). Furthermore, of the studies that estimate the effect of SEC attainment outcomes, most do not consider the mechanisms through which SEC impacts the outcome.

Another student outcome that is seemingly relevant to socioeconomic segregation but has received little attention in the research literature is behavioral engagement, which is commonly based on attendance and truancy rate, suspensions, disciplinary problems, etc. Understanding the role that schools play in the development and incidence of problem school behaviors is important because research shows behavioral engagement is predictive of a number of future educational, social, and employment outcomes, such as achievement, high school dropout, college

attendance, receiving future public assistance, criminal conduct, and employment earnings (Bowles & Gintis, 1976; Christle, Jolivette, & Nelson, 2007; Farkas, 2003; Jackson, 2013; Ou & Reynolds, 2010; Palardy & Rumberger, 2019; Palardy, Rumberger, & Butler, 2015).

11.2.2 Mediating Mechanisms: Peer Influences vs. School Practices

SEC is believed to capture the normative and cultural environment of the school. This may impact student attitudes about the value of schooling, appropriate behaviors, motivation levels, aspirations, etc., and it may also impact school personnel's attitudes and practices (Coleman et al., 1966; Hallinan & Williams, 1990; McDonough, 1997; Palardy, 2015b). But what are the specific school mechanisms through which SEC impacts student outcomes? Two leading theories have emerged: socioeconomic-based peer influences and socioeconomic-based school practices.

Peers are known to influence one another in a variety of ways and children are particularly susceptible to such influences in late adolescence (Cook, Deng, & Morgano, 2007; Hallinan & Williams, 1990; Jang, 2002; Mounts & Steinberg, 1995). Research has reported that the more enriched learning environment more commonly found at high-SEC schools is partially due having more highly motivated and academically advanced peers (Hanushek, Kain, Markman, & Rivkin, 2003; Jencks & Mayer, 1990; Legewie & DiPrete, 2012). These peer influences can lead to improved achievement, higher educational aspirations, fewer behavioral infractions, and more refined non-cognitive skills (Hallinan & Williams, 1990; Jencks & Mayer, 1990; Kahlenberg, 2001). These peer influences are theorized to mediate the effects of SEC on student outcomes. Socioeconomic-based peer influences tend to undermine attitudes, skills development, behaviors, and ultimately, educational performance in low-SEC schools, where peers tend to have lower levels of the educational attributes that enhance those outcomes (Palardy, 2013).

A second theory is that socioeconomic composition influences school personnel to adapt instructional practices, curricula, and academic orientation to what they perceive to be appropriate for the student body of the school (Coleman et al., 1966; McDonough, 1997; Thrupp, 1999). As a result, low-SEC high schools tend to have less academically demanding curricula, and students tend to be less academically engaged and spend less time on homework (Lippman, Burns, & McArthur, 1996). Low-SEC schools also tend to place a greater focus on discipline and order, typically at the expense of an emphasis on academics (Thrupp, 1999). Besides impacting academic skills and school behaviors, these differences between high- and low-SEC school practices can impact whether students have acquired the necessary coursework and academic skills for admission to a selective college and whether students have been socialized to believe that attending a selective college or any college is a desirable postsecondary option (Adelman, 1999; McDonough, 1997; Palardy, 2015b; Perna, 2004).

11.2.3 Implications of Theory to Policy Interventions

Determining the mechanism(s) through which SEC impacts student outcomes not only addresses theory, but also informs which policy interventions are likely to be effective in alleviating the consequences of socioeconomic segregation. School practices are considered to be largely alterable aspects of schools. Hence, if school practices are the primary mediating mechanism, then altering practices at low-SEC schools may sufficiently address the matter. However, peer influences are considered to be far less malleable by school personnel, as they are imbued in the traits and backgrounds of the students. Therefore, addressing peer influences may require altering the peer composition of the school. That is, it may require redistributing students among schools so that schools are more similar on SEC. Note that the SEC effects may not be fully mediated by either school practices or peer influences. In that case, the SEC effect may have its own genuine direct effect or there may be one or more other (unknown) mediating mechanisms. While this unexplained part of the SEC effect cannot be addressed by peer influences or school practices, it is safe to assume that it can be addressed by making schools more equal in terms of SEC. This means that redistributing students to make schools more equitable on SEC is the most asured method of addressing the negative consequences of socioeconomic segregation among schools.

11.2.4 Research Questions

This study addresses the following research questions.

1. Is SEC associated with each of the six high school and college-going outcomes considered in this study, and if so, what is the effect size for each? This question addresses which outcomes are impacted most by socioeconomic segregation.
2. To what degree are the effects of SEC mediated by peer influences and/or school practices? This question addresses the theory of SEC effects and informs policy and practice interventions.

11.3 Methods

11.3.1 Data

This study used data from the Education Longitudinal Study (ELS), a survey of approximately 15,000 tenth graders who attended 750 public, Catholic, and other private high schools in 2002 (Ingels, Pratt, Rogers, Siegel, & Stutts, 2004). Data were collected on an extensive number of variables from students, their parents, teachers, and school principals in the Spring of 2002 when students were completing

tenth grade and again in the Springs of 2004, 2006, and 2012. Education Longitudinal Study is an outstanding data source for modeling the effects that high schools have on student outcomes.[3] Moreover, because ELS is relatively new, results based on ELS are more relevant for addressing current educational issues than results are from older databases. This is important because a substantial proportion of the research literature on school segregation in U.S. high schools is based on data that are a generation old and may be obsolete given the substantial social and educational change that occurred during that period. The sample used in this study is limited to students who attended public schools because public schools have greater significance to educational policy and because students attending private schools are more susceptible to selection bias that can confound estimated effects. The sample used in this analysis includes 10,151 students who attended 581 public high schools.

11.3.1.1 Outcome Variables

This study uses six different outcomes: Academic Performance, Behavioral Engagement, High School Graduation, College Choice, 2-year College Enrollment, and 4-year College Enrollment. These outcomes were selected to reflect a range of functions schools serve, such as the development of academic skills and study skills, the promotion of behavioral engagement, and the completion of high school and enrollment in post-secondary education. One challenge is that these outcomes include continuous, ordinal, and categorical variables, each requiring a specialized modeling approach, as described in the Statistical Modeling section below. The descriptive statistics for each outcome are provided in Table 11.2.

11.3.1.2 Independent Variables

Independent variables were selected based on theory and the research literature on school segregation and compositional effects (McDonough, 1997; Palardy, 2015b; Palardy & Rumberger, 2019; Palardy, Rumberger, & Butler, 2015; Perry & McConney, 2010: Van Ewijk & Sleegers, 2010). Four broad categories of independent variables are used to address the research questions: Student Inputs, School

[3] A stratified two-stage sampling design was used that involved selecting a sample of schools that enroll tenth graders based on probabilities proportional to school enrollment. Adolescents of Asian, Pacific Islander, and Hispanic ethnicity were oversampled to ensure sufficiently large samples of minority groups. As a result of these sampling strategies, neither the student nor the school sample can be considered representative of the population of 2002 tenth graders or schools that enrolled 2002 tenth graders. However, NCES provides student and school sample weights to transform the data into nationally representative samples of tenth graders and high schools. The present study uses the ELS:2002 first follow-up, base year panel weight (F1PNLWT), and the base year school sample weight (BYSCHWT). For additional information on ELS:2002, see http://nces.ed.gov/surveys/els2002

Table 11.2 Descriptive statistics

Variable name	Mean (SD)	Descriptions
Student-level variables		
Outcomes		
Academic performance	0.00 (1.00)	FS of GPA, math and reading achievement tests
Behavioral engagement	0.00 (1.00)	FS of attendance rate, suspensions, and retentions.
High school graduation	0.86	Graduated on time
College choice (selectivity)		Ordinal measure of selectivity of the college attended
0 = no college**	1.23	Did not attend college
1 = non-selective 2-year**	0.32	Attended non-selective 2-year
2 = less than highly selective 4-year**	0.26	Attended less than highly selective 4-year
3 = selective 4-year**	0.30	Attended highly selective 4-year college
Enrollment Didnot enroll	0.31	Enrolled in a 2- or 4-year college vs. otherwise
2-year college	0.26	
4-year college	0.43	
Student inputs		
Demographics and family background		
SES	0.00 (1.00)	Socioeconomic status composite
Traditional family structure	0.53	Live with both birth parents
Asian/Pacific islander	0.04	
Black	0.14	
Hispanic	0.15	
American Indian	0.01	
Parental engagement	0.00 (1.00)	Factor score
Academic background and aspirations		
Cumulative academic GPA (tenth grade)	2.50 (0.91)	10th grade academic GPA
Math/reading achievement (tenth grade)	0.00 (1.00)	Reading and math test composite
Student's attainment expectation	0.79	Student plans to attend college
Student engagement		
Participant varsity sports	0.50	Participated in at least one varsity or JV sport
Participant school club	0.50	Participated in at least one school club
Transfer	0.19	Transferred schools during H.S.

(continued)

Table 11.2 (continued)

Variable name	Mean (SD)	Descriptions
Financial aid needs and availability[b] (used in post-secondary outcome models only)		
Financial aid offered	0.62	Grant, loan, work study, or waiver offered
College expenses high importance	0.35	
College expenses moderate importance	0.48	
School-level Variables		
School inputs		
Student body composition		
SEC (mean SES)	0.00 (1.00)	Standardized mean 10th & 12th grade SES
Percent minority	27.68 (31.39)	Percent African American and Hispanic
School resources		
Student/teacher ratio	15.55 (5.10)	Student/teacher ratio
Learning hindered by facilities	0.00 (1.00)	Factor score
Learning hindered by equipment	0.00 (1.00)	Factor score
Percent full teacher certification	90.59 (15.72)	
Teacher salary	40,123 (7638)	Mean teacher salary
School structures		
Urban	0.15	School located in urban setting
Rural	0.44	School located in rural setting
Small school	0.56	Enrollment less than 600
Large school	0.13	Enrollment = 1201–1800
Extra-large school	0.09	Enrollment greater than 1800
Peer influences		
Friend effects		
Mean friend dropout	0.51 (0.20)	Proportion have friend who dropped out
Closest friend college expectations	0.45 (0.16)	Proportion friend desires student to attend college
Peer achievement and peer Parental capital		
Mean parental engagement	0.03 (0.28)	Mean factor score
Mother college expectations	0.59 (0.17)	Proportion mother desires child to attend college
Prior achievement	−0.02 (0.49)	Mean reading and math composite

(continued)

Table 11.2 (continued)

Variable name	Mean (SD)	Descriptions
School practices and contexts		
Academic		
Academic press	0.00 (1.00)	Factor score
Homework time	0.00 (1.00)	Mean hours of homework per week
Carnegie units	0.00 (1.00)	Mean Carnegie units earned in H.S.
Math pipeline	0.00 (1.00)	Mean level of mathematics coursework completed
Teaching		
Teacher quality	0.00 (1.00)	Factor score based on student ratings
Teacher efficacy	0.00 (1.00)	Factor score
Teacher recognition/support	0.49	Proportion principal recognizes good teachers
Teacher morale	0.00 (1.00)	Factor score
Disciplinary		
Discipline fair	0.00 (1.00)	Factor score
Unsafe school environment	0.00 (1.00)	Factor score
Classroom disruption	0.00 (1.00)	Factor score
Disorder	0.00 (1.00)	Factor score

Note: Student variables are weighted by F1PNLWT (normalized), and school variables are weighted by BYSCHWT (normalized). These sample weights are designed to produce nationally representative samples of students and high schools

Inputs, Peer Influences, and School Practices. Each category has two to four subcategories. The primary independent variable is SEC. It is the mean SES of sampled students in each school. Socioeconomic status was measured in tenth grade and was constructed by NCES as equally-weighted composites of five measures: family income and each parent's education level and occupational status. To differentiate the effects of SEC from racial/ethnic segregation, the ethnic/racial compoistion measure, Percent Underserved Minority, is also used, which is the percentage of the students attending each school that are African American or Hispanic. Table 11.2 provides a list of all variables used in this study and their descriptive statistics. The independent variables on the table are organized into the four categories described next.

Student and School Inputs

One of the controversies in the school composition literature is regarding which factors should be statistically controlled (Harker & Tymms, 2004). While most scholars agree that it is essential to at least control for the corresponding student measure (i.e., family SES when modeling SEC), some argue that the alone may be insufficient and can result in the overestimation of compositional effects (Harker & Tymms, 2004). Student and school inputs are used in this study to statistically control for factors that may introduce selection biases into estimates of the effects of SEC. These inputs include a range of student demographic, family, and academic background measures that are generally associated with SEC and the outcomes of this study and therefore may be driving the associations between SEC and the outcomes. Student inputs also include forms of student engagement and, for the models with post-secondary outcomes, measures of college financial aid need, which can impact decisions on whether to attend college. School inputs are aspects of schools that are largely beyond the control of school personnel and include variables measuring school resources and structures, many of which tend to be correlated with SEC. These extensive controls for student and school inputs are included to avoiding overestimation of the SEC effects and instead produce conservative estimates.

Peer Influences

As described earlier, peer influences are one of two leading explanations, along with school practices, of mediating mechanisms through which school segregation impacts student outcomes. Two types of peer influences are used in this study. The first type is the attitudes and behaviors of close friends. This type is considered most influential because students interact more with close friends than with other students at their school and therefore have more opportunity to be influenced, and also because the friendship creates emotional connections that result in greater perceptibility to and conformity with their friend's attitudes and behaviors (Palardy, 2013). The second type of influence is peer achievement and peers' parental social capital. This type is a less direct influence, but it can affect the social milieu of the school, particularly in terms of academic and behavioral expectations and the value placed on educational attainment (McDonough, 1997; Palardy, 2015b).

School Practices and Contexts

As was the case with peer influences, school practices and contexts tend to be associated with the outcomes of this study and with SEC and are a leading theoretical explanation for mechanisms that mediate the effects of SEC. They are conceptualized as being under the control of school personnel to a substantial degree. School practices and contexts are divided into three subcategories: academic, teaching, and disciplinary. The academic subcategory includes measures of the degree to which the academics are emphasized at the school and include measures of the strength of academic focus of the curriculum and instruction (Academic Press), the mean number of hours per week that students report spending on homework (Homework Time), and how far the average student advances in the mathematics curriculum (Math Pipeline). Teaching includes measures of the average student ratings of teacher quality, self-ratings of teachers' efficacy, and principal ratings of teacher

morale. The disciplinary subcategory includes average student ratings of the fairness of the discipline policy, how safe they feel at school, the level of classroom disturbances, and the level of school disorder.

11.3.2 Statistical Models

Multilevel models were used because students (level 1) are nested in schools (level 2). Nested data violate the assumption of statistical independence when analyzed using traditional statistical approaches, which can bias estimates of effects and their standard errors. Multilevel models alleviate these problems and provide the opportunity to model the dependencies in the data, which are school effects on student outcomes in the present study. This section describes the mediation models, the model-building strategy that was used to address the research question of the study, and the statistical equations.

11.3.2.1 Mediation

A central aspect of this study is estimating the degree to which two different factors—peer influences and school practices—mediate the SEC effect on each of the six outcomes. Multilevel mediation is used because the antecedent, SEC, is a school-level measure, whereas all outcomes are measured at the student-level (Bauer, Preacher, & Gil, 2006; Krull & MacKinnon, 2001). Figures 11.1a and 11.1b depict multilevel models for estimating the total effect, the direct effect, and the mediation (indirect) effects of SEC on a student outcome. In Fig. 11.1a, c represents the total effect. Note that there is only one other predictor in the total effects model, student SES. Student SES is included because the correct estimation of school compositional effects requires controlling for the corresponding individual-level variable, in this case, SES (Raudenbush & Bryk, 2002). Figure 11.1b depicts the final mediation model, which includes control variables

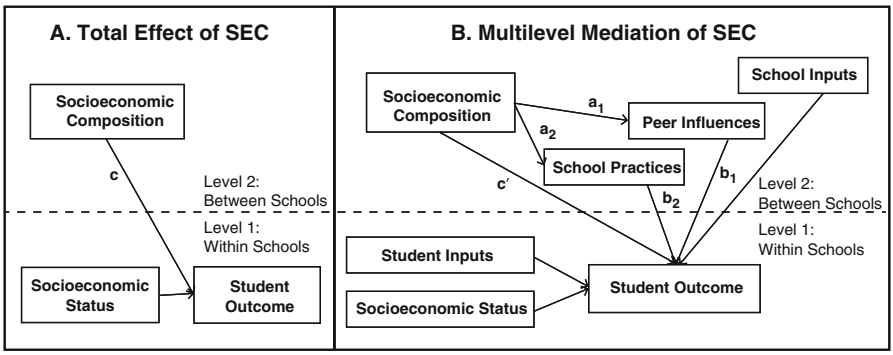

Fig. 11.1 (**a**) and (**b**) Total effects of mediation effects of SEC on student outcomes

for Student Inputs and School Inputs in addition to Peer Influences and School Practices mediators. In the mediation models, c' represents the direct effect of SEC on the outcome and $a_1 * b_1$ and $a_2 * b_2$ represents the mediation effect or indirect effect of SEC on the outcome through or Peer Influences and School Practices, respectively.

Note that Fig. 11.1b is simplification of the conditions of this study in that there are actually multiple measures of each of the two types of mediators. For example, Peer Influences includes the five measures listed in Table 11.2. The mediation effects of each of those individual measures of Peer Influences can be estimated and summed to obtain the combined Peer Influence mediation effect as shown by Palardy (2015).

However, there is a shortcut for estimating the combined mediated effect. Note that the total effect can be partitioned into the direct and indirect (mediation) effects as follows: $c = c' + a*b$ (MacKinnon, 2008). This equation can be solved for the mediation effect: $a*b = c - c'$. Furthermore, the percentage of the total effect that is mediated can be computed using $(1 - c'/c)^*100$. The usefulness of this shortcut is that estimating c and c' does not require estimating the "a" parameters in Fig. 11.1b or summing the multiple mediation effects within each type. This shortcut is particularly efficient when there are multiple measures of each type of mediator being tested and the interests is in the combined effect of each type, as is the case for the present study.

One potential limitation of the c-c′ approach when there are multiple types of mediation effects is the effects of each type cannot be estimated simultaneously. However, that is not a limitation when the types of mediation effects are conceptualized as hierarchical, in which case the objective is typically to determine the mediation effect of each subsequent type above and beyond the preceding types using a stepwise model building strategy. In the present study the mediation effects are conceptualized as hierarchical. Peer Influences are precursors to School Practices in that they are aspects of the students that are given to schools and that school personnel have little control over but are correlated with SEC, School Practice, and the outcomes. Therefore, the mediation effects of Peer Influences will be estimated first. Then the mediation effects of School Practices will be estimated in the model retaining the Peer Influences measures.

The mediation effect derived from the two approaches are equal for OLS estimates and highly similar for multilevel mediation. One limitation is that for categorical outcomes the c-c' method can produce low biased estimated because the log-odds transformation fixes the scale on the outcome and therefore the scale is not equal across models that have different predictors. MacKinnon (2008, p. 306, Fig. 11.2) shows that the degree to which mediation estimates from the c-c' method deviate from the a∗b method depends on the magnitude of the effects of the additional covariates on the outcome in the model. The larger the effects of the additional covariates, the greater the downward bias in the c-c' estimates compared with the a∗b estimates. The workaround to this potential bias is to standardize the mediation effects (MacKinnon, 2008). This study used the Chinn (2000) method for standardizing the mediation effects for the categorical outcomes. More research is needed on whether this standardization approach is sufficient for addressing the bias. As we will see later, the mediation effects are substantial for each of the of the categorical outcomes, yet they may still be underestimates.

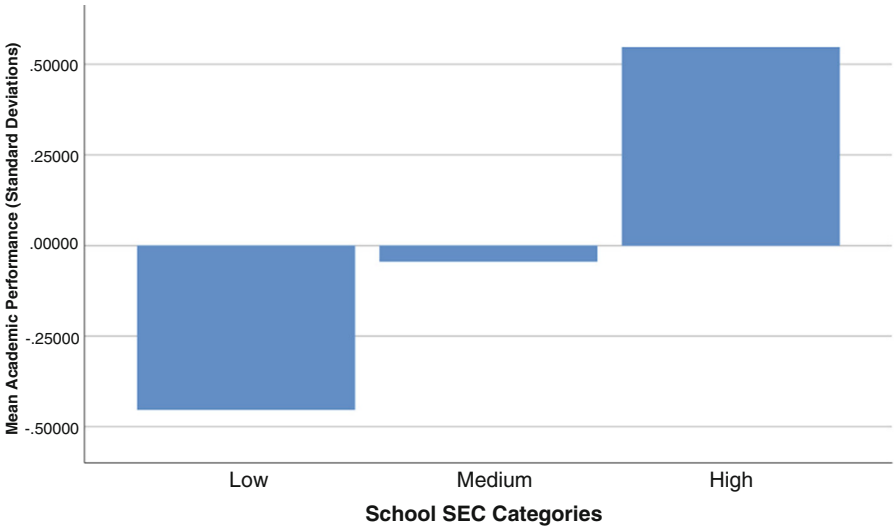

Fig. 11.2 Mean academic performance in low-, medium-, and high-SEC schools

11.3.2.2 Model Building

A series of four models were fit to each outcome: Baseline, Inputs, Peer Influences, and Schools Effects. This section provides a description of each model and a rationale for the hierarchical order of model building.

The Baseline model includes only student SES and school SEC. SES is grand mean centered, which adjusts the outcomes for differences across schools on student SES. This allows the effect of SEC on the outcome to be interpreted as above and beyond the effect of students' own SES backgrounds. However, even with the SES control, this estimate of SEC may be inflated due to school selection biases, as students are not randomly assigned to schools, because their parents often chose to live in specific neighborhoods and attend specific schools for a reason.

The Input model controls for student and school characteristics for the propose of reducing selection biases. While middle- and upper-middle class families often choose to live in specific neighborhoods because of the reputations of the schools there, low-SES families are often limited to schools located in affordable neighborhoods (Owens, 2018). This suggests that SES is the primary selection variable that should be statistically controlled to obtain unbiased estimates of the SEC effects. However, other students, family, and school factors may also contribute to the school choice decision and therefore should be controlled to minimize the chance that the SEC effect is confounded with the selection effects. The Input model is designed to control for those other factors and provide somewhat conservative estimates of the total SEC effects. The controls include a range of students' demographic, academic, engagement, and financial aid measures (for post-secondary

outcome models only) in addition to school inputs measuring student composition, school resources, and school structures.

Peer Influences is the third model. This model builds on the Input model by adding several measures of the influences of close friends, peer achievement, and peer parental capital. The c-c′ method is used to provide an estimate of the percent of the total SEC effect from the Input model that was mediated by Peer Influence, whereas c is the total effect estimate from the Input model and c′ is the direct effect from the Peer Influences model.

The fourth and final model is the School Practices model. It builds on the Peer Influence model by adding measures of school practices related to academics, teaching, and discipline. The purpose of this model is to provide an estimate of the degree to which School Practices mediate the SEC effect after controlling for student and school inputs and peer influences. Peer influences are controlled in the School Practices model because school they are conceptualized a school condition that is largely beyond the control of school personnel and therefore will impact the SEC effect independent of school practices. The c-c′ method is also used here, whereas c is the (direct) SEC estimate from the Peer Influence model and c′ is the direct effect from the School Practice model.

11.3.2.3 Statistical Equations

This study uses three types of outcomes: continuous (Academic Performance and Behavioral Engagement), ordinal (College Choice), and categorical (High School Graduation and College Enrollment). The equations below are in the form applicable to the continuous outcomes. Within the framework of hierarchical generalized linear models, link functions (i.e., logit) are used to fit otherwise equivalent models to ordinal and categorical outcomes (Raudenbush & Bryk, 2002). Furthermore, the SEC effect for each outcome is converted to an effect size to promote comparability. The nonlinear model effects (i.e., ordinal and categorical outcomes) were converted to effects sizes using the methods described by Chinn (2000).

The following multilevel equations are for the School Practice model. The School Practice model equations are presented as an example because it is the final and most fully-specified model in this study. The other models (Baseline, Inputs, and Peer Influences) are each simplifications of this model where sets of predictors are omitted. The level 1, or student level, can be represented by the following equation:

$$Y_{ij} = \beta_{0j} + \beta_{10} SES_{ij} + \sum_{p=2}^{P} \beta_{pj} X_{pij} + r_{ij} \qquad r_{ij} \sim N\left(0, \sigma^2\right),$$

where Y_{ij} is the outcome for individual i in school j and β_{0j} is the expected value on the outcome for students attending school j conditioned on student covariates X_1 through $X_{p.}$. β_{1j} through β_{qj} represent the expected change in the outcome per unit change in the respective covariate, X_1 through X_p, and r_{ij} is the level 1 residuals,

which capture the deviation in the estimated values from the observed values for each student. The residuals are assumed to have a mean of zero, be normally distributed, and have a variance (σ^2) that is the within-school variance in the outcome not accounted for by the predictors in the model.

The level 2 or school-level model can be represented by the following set of equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{10} SEC_{1j} + \sum_{q=2}^{I} \gamma_{q0} Inputs_{qj} + \sum_{q=I+1}^{P} \gamma_{q0} Peers_{qj} + \sum_{q=P+1}^{Q} \gamma_{q0} Schools_{qj} + \mu_{0j} \quad \mu_{0j} \sim N(0, \tau_u)$$

$$\beta_{1j} = \gamma_{10}$$

$$\vdots \qquad \vdots$$

$$\beta_{pj} = \gamma_{p0},$$

where γ_{00} is the intercept estimate (mean on outcome adjusted for covariates), SEC_{1j}, $Inputs_{qj}$, $Peers_{qj}$, and $Schools_{qj}$ are types of school-level covariates, and γ_{10} through γ_{q0} are school-level slope coefficients corresponding with the school covariates, which describe the expected change in the outcome per unit change in the covariate variable. I, P, and Q represent the number of Inputs, Peer Influence, and School Practice variables, respectively. As described in the Model Building section, the types of covariates were added sequentially during model building to address the research questions of the study. u_{0j} represents the school residuals, which describe the deviation in the model-estimated effects from the observed values for each school. The residuals are assumed to be normally distributed with a mean of 0 and variance of τ_u. Note that the equations indicate that only the level 1 intercept coefficient (β_{0j}) randomly varies across schools and that each of the level 1 slope coefficients (β_{1j} through β_{qj}) is fixed to be equal for all schools.

All variables used in this study had moderate (less than 30%) to low (less than 5%) rates of missing values. Missing values on student and school variables were imputed via the EM method using SPSS.

11.4 Results

This section presents the results for each outcome separately beginning with Academic Performance and followed by Behavioral Engagement, High School Graduation, College Choice, and 2- and 4-year College Enrollment. The focus is on describing the effects size of SEC on each outcome and the degree to which Peer Influences, and School Practices mediate those effects. Note that throughout the results low-SES schools are defined as having a mean SES 1.0 or more standard deviations below that of the average school (i.e., the bottom 16% of the schools on mean SES), whereas high-SES schools are defined as having a mean SES 1.0 or

more standard deviations above average (i.e., top 16%). The intra-class correlation coefficient (ICC) for each model, which is proportion of the total variance for each outcome that is between schools or at level 2 of the multilevel model, is not provided because several of the outcomes are categorical and there is controversy about the appropriate method of estimating the ICC for categorical outcomes. For more information on the estimation of ICCs for not linear outcomes, see Browne (2011).

11.4.1 Academic Performance

Recall that academic performance is a composite measure of reading and math achievement and GPA in academic coursework that was constructed using factor analysis (see Appendix Table 11.8 for details on the factor structure). Figure 11.2 shows that the mean academic performance at low-SEC schools is nearly a full standard deviation lower than at high-SEC schools (i.e., approximately 30 vs. 70 percentile), whereas students attending medium-SEC schools are in the middle. Table 11.3 shows that the total effect of SEC from the Baseline model is 0.62 standard deviations. However, that effect is likely overestimated due to the lack of controls for selection biases of students into schools. The Input model results show an SEC effect size of 0.55, a slight reduction compared with the Baseline model. This model controls for several demographic and academic background measures, student engagement measures, and school resources and structures (see Table 11.2 for a full list of the control variables used in this and other models) and therefore is considered a conservative estimate of the total SEC effect size before examining mediation by peer influences and school practices.

The 0.55 effect size for SEC is larger than the 0.32 cited in a recent meta-analyses (Van Ewijk & Sleegers, 2010). That is likely because the meta-analyses estimate is based on achievement test score outcomes, whereas the academic performance measure used in this study is a factor score of reading achievement, math achievement, and GPA for academic coursework. The addition of academic GPA results in a broader measure of academic performance than achievement test scores alone and

Table 11.3 Academic performance

Variable name	Baseline SEC	Inputs	Peer effects	School practice
SEC	0.62^{**}	0.55^{**}	0.49^{**} (−11%)	0.17^{**} (−65%)
SES	0.29^{**}	0.24^{**}	0.24^{**}	0.24^{**}
Student controls entered?	No	Yes	Yes	Yes
School input controls entered?	No	Yes	Yes	Yes
Peer influence controls entered?	No	No	Yes	Yes
School practice controls entered?	No	No	No	Yes

† = significant at α = 10%; * = significant at α = 5%; ** = significant at α = 1%. The percentage values in parentheses are the reduction in the SEC effect size compared with the previous model

one that is arguably more relevant in that it is similar to the criteria used for admission to most 4-year colleges in the U.S.

The addition of the Peer Influences mediators reduces the direct SEC effect to 0.49 standard deviations. This indicates Peer Influences mediation accounts for a relatively small percentage of the SEC effect from the Input model [(1-c′/c)*100 = 1–.49/.55*100 = 11%)]. However, School Practices mediate a far larger percentage of the SEC effect, reducing the direct effect of SEC to 0.17 standard deviations ($p < 0.01$). Compared with the direct effect of SEC from the Peer Influence model, School Practices mediates an additional 65% the SEC effect [(1-c'/c)*100 = 1–.17/.49*100 = 65%)]. These findings suggest that the primary mechanism through which SEC impacts Academic Performance is through school practices rather than peer influences.

11.4.2 Behavioral Engagement

The Behavioral Engagement outcome is also a factor score. It was derived from variables measuring student absenteeism, suspension, and grade retention rates (see Appendix Table 11.8 for details on the factor structure). Figure 11.3 shows that the magnitude of the difference between-low and high-SEC schools is approximately 0.28 standard deviations, which is about a fourth of the differences for Academic Performance. This suggests that in the Behavioral Engagement is less strongly linked with SEC, which helps explain why it is the only outcome in this study for which SEC does not have a statistically significant association in the Baseline model

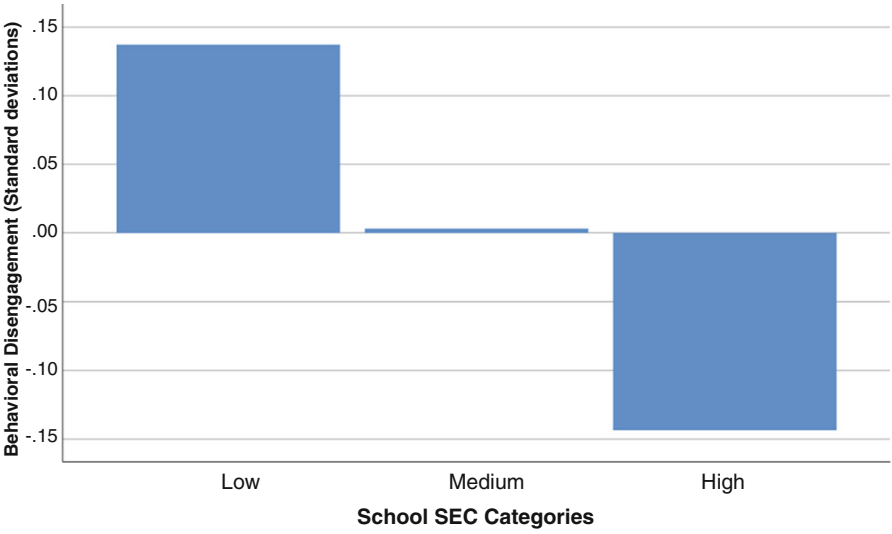

Fig. 11.3 Mean behavioral engagement in low-, medium-, and high-SEC schools

Table 11.4 Behavioral engagement

Variable name	Baseline SEC	Inputs	Peer effects	School practice
SEC	0.03	0.02	0.06	−0.09
SES	0.30**	0.26**	0.26**	0.26**
Student controls entered?	No	Yes	Yes	Yes
School input controls entered?	No	Yes	Yes	Yes
Peer influence controls entered?	No	No	Yes	Yes
School practice controls entered?	No	No	No	Yes

† = significant at α = 10%; * = significant at α = 5%; ** = significant at α = 1%

(see Table 11.4). Also unique for the Behavioral Engagement outcome is that controlling for Peer Influences *increases* the SEC effect size slightly (from 0.02 to 0.06), although it is still non-significant ($p > 0.10$). However, adding the School Practices to the model reduces the SEC effect well into the negative territory. While the new effect (ES = −0.09) remains non-significant, the change in the ES was −0.15 (from 0.06 to −0.09) after adding school practices is statistically significant ($p < 0.05$). These results indicate that while SEC is not predictive of Behavioral Engagement in high school, school practices still mediate the effect.

11.4.3 High School Graduation

High School Graduation is a binary outcome. Hence, the multilevel logistic regression was used, and the initial coefficient estimates were in the logit metric. For comparison across outcomes, the SEC effect was converted to an effect size using Chinn's (2000) method. Figure 11.4 shows that graduation rates differ linearly from low- to medium- to high-SEC schools, with rates of 80%, 86%, and 92%, respectively. Note that the standard deviation for graduation rates is 12%, and therefore the difference between low- and high-SEC schools is quite large at a full standard deviation in magnitude and equivalent to the difference for the Academic Performance outcome discussed previously.

Table 11.5 shows that the total SEC effect from Baseline model is 0.19 standard deviations ($p < 0.01$). Controlling for student and school inputs reduces the effect by more than half to a total effect of 0.08, although it remains highly statistically significant ($p < 0.01$). Peer Influences mediates half of the total effect. The remaining direct effect of SEC is a non-significant (ES = 0.04, $p > 0.10$). School Practices mediate an additional 75% of the SEC effect, reducing the direct SEC effect to almost zero (ES = 0.01). These results suggest that Peer Influences play a prominent role in the association between SEC and high school graduation. After accounting for peer influences, the SEC effect is not statistically significant.

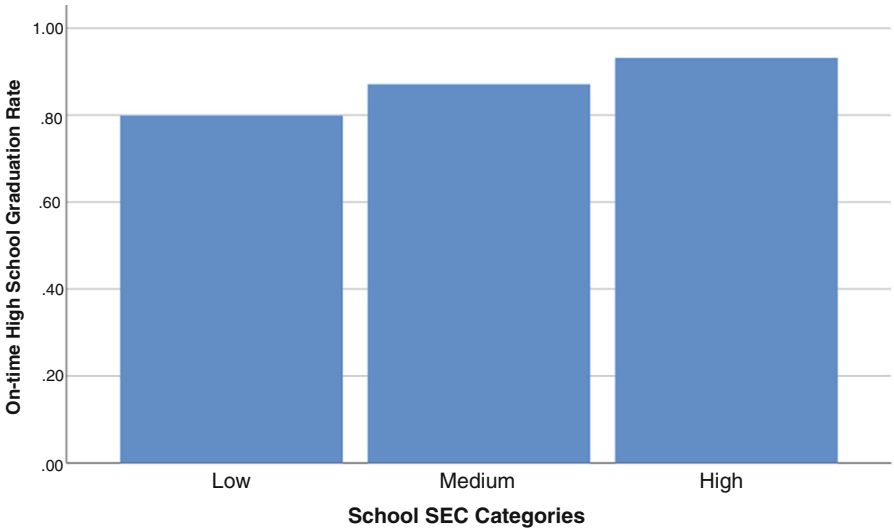

Fig. 11.4 High school graduation rate in low-, medium-, and high-SEC schools

Table 11.5 High school graduation

Variable name	Baseline SEC	Inputs	Peer effects	School practice
SEC	0.19**	0.08**	0.04 (−50%)	0.01 (−75%)
SES	0.22**	0.04	0.04	0.04
Student controls entered?	No	Yes	Yes	Yes
School input controls entered?	No	Yes	Yes	Yes
Peer influence controls entered?	No	No	Yes	Yes
School practice controls entered?	No	No	No	Yes

† = significant at α = 10%; * = significant at α = 5%; ** = significant at α = 1%. The percentage value in parentheses is the reduction in the SEC effect size compared with the previous model

11.4.4 College Choice

College Choice is an ordinal outcome for which the initial estimates of the effect of SEC are in the logit metric. Like the estimates for High School Graduation, they were converted to effect sizes using the methods described by Chinn (2000). Figure 11.5 indicates that the College Choice distributions for low- and high-SEC schools are close to mirror images: Directly after high school, most students from low-SEC schools do not attend college or attend a non-selective 2-year college, whereas most students from high-SEC schools attend a moderately selective or selective 4-year college. As was the case with the three outcomes discussed previously, the distribution of the College Choice outcome for students attending medium-SEC schools is approximately the average of the distributions for low- and high-SEC schools. However, there is an exception to that generalization: students from medium-SEC schools have a notably lower rate of attending a

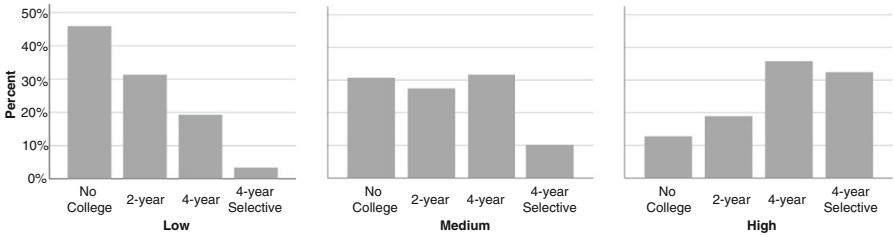

Fig. 11.5 College choice in low-, medium, and high-SEC schools

Table 11.6 College choice

Variable name	Baseline SEC	Inputs	Peer effects	School practice
SEC	0.34^{**}	0.17^{**}	0.08^{**} (−53%)	0.07^{**} (−13%)
SES	0.42^{**}	0.33^{**}	0.33^{**}	0.33^{**}
Student controls entered?	No	Yes	Yes	Yes
School input controls entered?	No	Yes	Yes	Yes
Peer influence controls entered?	No	No	Yes	Yes
School practice controls entered?	No	No	No	Yes

† = significant at α = 10%; * = significant at α = 5%; ** = significant at α = 1%. The percentage values in parentheses are the reduction in the SEC effect size compared with the previous model

selective college than do the average of the low- and high-SEC groups. While the overall rate of selective college attendance is 12%, the rates of students from low-, medium-, and high-SEC high schools is 4%, 9%, and 29%. That means if the percentage at medium-SEC schools were the average of low- and high-SEC schools, about 16.5% would attend a selective college, but the actual percentage is about half that (9%). Also, recall that the data used in this study are from public high schools, whereas most elite high schools in the United States are private. Hence, these figures do not likely fully reflect the disparity in selective college attendance rates across all U.S. high schools—public and private. Selective colleges in the United States remain largely the domain of the privileged class of students attending high-SEC schools. One noteworthy exception is the University of California, which includes selective campuses and enroll a substantial percentage of students from low SEC high schools (Heller, 2004).

Table 11.6 shows that the total SEC effect from the Baseline model is 0.34 standard deviations. Controlling for student and school inputs reduced the SEC effect by 50% to 0.17 ($p < 0.01$). Peer Influences mediate 53% of that, reducing the direct SEC to 0.08, which is statistically significant ($p < 0.01$). School Practices mediate the remaining direct effect by 13%, and the direct effect remains highly significant ($p < 0.01$). These findings suggest that the SEC effect on College Choice is exerted mostly through peer influences and that school practices play a relatively minor role.

11.4.5 College Enrollment

Figure 11.6 shows that, similar to the College Choice outcome, the College Enrollment distributions for students from low- and high-SEC schools are almost opposite, and medium-SEC is approximately the average of the two. Most students from low-SEC high schools do not attend college or attend a 2-year college (76% combined), whereas two-thirds of the students from high-SEC schools attend a 4-year college directly after high school.

Table 11.7 shows that the SEC effect on 4-year College Enrollment is several times larger than for 2-year College Enrollment. That is because a 2-year college is the only accessible post-secondary option for some students from low-SEC schools due to financial limitations and poor access to information about financial aid. However, students from high-SEC schools, and to a lesser extent, medium-SEC schools, tend not to have those limitations for enrolling in a 4-year college. Hence,

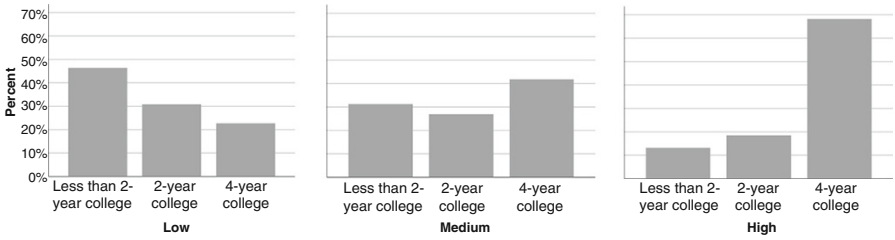

Fig. 11.6 College enrollment in low-, medium-, and high-SEC schools

Table 11.7 College enrollment

Variable name	Baseline SEC		Inputs		Peer effects		School practice	
	2-year	4-year	2-year	4-year	2-year	4-year	2-year	4-year
SEC	0.08**	0.24**	0.06**	0.21**	0.03 (−50%)	0.17** (−19%)	0.03	0.11** (−35%)
SES	0.14**	0.29**	0.10**	0.24**	0.10**	0.24**	0.10**	0.24**
Student controls entered?	No		Yes		Yes		Yes	
School input controls entered?	No		Yes		Yes		Yes	
Peer influence controls entered?	No		No		Yes		Yes	
School practice controls entered?	No		No		No		Yes	

† = significant at α = 10%; * = significant at α = 5%; ** = significant at α = 1%. The percentage values in parentheses are the reduction in the SEC effect size compared with the previous model

the association between SEC and 2-year College Enrollment is weak. For this reason, the following discussion focuses on the results for 4-year College Enrollment.

Table 11.7 shows that the Baseline model effect size for SEC on 4-year College Enrollment is a highly significant 0.24 ($p < 0.01$). Controlling for student and school inputs reduced that only marginally (by 12.5%) to an effect size of 0.21 ($p < 0.01$). Peer Influences mediate about 19% of the total direct effect from the Input model, reducing it to 0.17 ($p < 0.01$). School Practice mediate the SEC effect an additional 35% to 0.11, which is still highly statistically significant ($p < 0.01$). These findings suggest that the effect of SEC on 4-year College Enrollment is primarily mediated by School Practices and, to a lesser extent, by Peer Influences.

11.5 Discussion

This study focuses on estimating the effects of high school SEC on a range of outcomes measured during, and immediately after completing, high school and estimating the magnitude of two mediating mechanism (peer effects or school practices) through which SEC impacts the outcomes. This section begins with a comparison of the total effect of SEC with the mediation effects of Peer Influences and School Practices across the six outcomes. That is followed by a comparison of the effect size of SES and SEC across outcomes. Lastly, the implications of the results for policy and practice, the limitations of the study, and future directions of this line of research are discussed.

11.5.1 Multiple Outcomes – Multiple Effects

Figure 11.7 breaks down the SEC effect for each of the six outcomes into three components: the total effect from the Input model and the mediated effects of Peer Influences and School Practices. The figure shows that the total effect sizes vary substantially across the outcomes. Note that three of the outcomes have total effects of 0.10 standard deviations or greater, which might be considered the lower boundary for a substantive effect.[4] At 0.55 standard deviations, Academic Performance has by far the largest total SEC effect, and College Choice and 4-year College Enrollment also have substantial total SEC effects of 0.17 and 0.21 standard deviations. In

[4]Cohen's (1988) widely-used effect size categories states that a correlational effect of 0.10 or less is considered a small effect size and equivalent to a Cohen's d effect size for mean comparisons of 0.20 or smaller. However, these guidelines are context-specific, and some may argue that a school effect 0.10 is substantive. For example, the average annual effect on achievement of reducing class sizes from 15 to 24 students in the early grade levels is approximately 0.06 standard deviations (Finn & Achilles, 1999). While small, some argue it is meaningful.

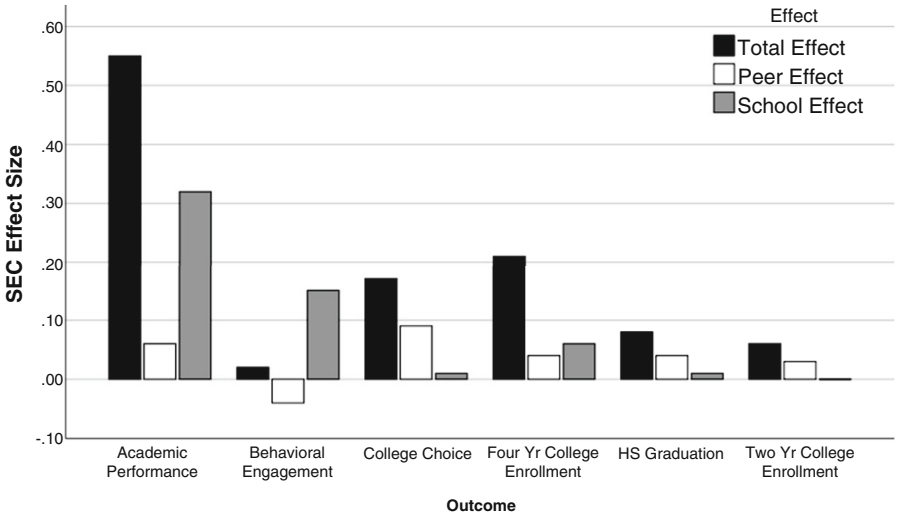

Fig. 11.7 Total, peer, and school effect sizes of high school SEC for each outcome

comparison, the SEC effects on High School Graduation and 2-year College Enrollment are much smaller, and the SEC effect on Behavioral Engagement is essentially zero.

11.5.1.1 SEC Effect Larger for Academic Outcomes

This pattern of results suggests that the greater the academic focus of the outcome, the more influential SEC tends to be. For example, scoring high on the Academic Performance outcome and enrolling at a selective 4-year college both requires high grades in academic coursework and high scores of standardized achievement tests. However, scoring high on behavioral engagement, graduating high school, or enrolling at a 2-year college requires only regular attendance, staying out of trouble at school, and completing high school graduation requirements, all of which necessitate minimal academic foci. These results suggest that socioeconomic segregation in schools is most detrimental to academic outcomes.

11.5.1.2 Peer Influence Mediation Strongest for College Choice, but School Practices are for Academic Performance

A high percentage of students enrolled in some type of postsecondary educational institution directly after high school (69%); however, selective college enrollment was much less common (12%). Figure 11.7 shows that Peer Influences are the primary mechanism mediating the SEC effect on College Choice, accounting for

the majority of the SEC effect for College Choice (an ES reduction of 0.09 or 53%). These findings suggest that peers may more readily influence one another's attitudes and choices than each other's academic skills and performance. For example, peers can readily influence one another's attitudes regarding the type of college that is most desirable to attend, but it is more difficult to impact their math problem-solving skills. Unfortunately, positive peer influences for pursuing admissions at a selective 4-year college are far less likely to be present at low-SEC high schools than affluent ones. While Peer Influences are the predominent mediator of the effects of SEC on College Choice, School Pracices are for Academic Performance. That suggest the academic skills and knowledge learned at school are much more a function of the teaching quality, curriular rigor, and disciplinary climate than the influences of one's classmates.

11.5.1.3 SEC vs. SES

Another way of interpreting the effect sizes for SEC is to compare them with the effect sizes for family SES, which is widely considered to be one of the most robust student predictors of educational outcomes. This comparison can also be used to gauge the degree to which SES functions as an individual factor impacting educational outcomes vs. a school compositional effect. Figure 11.8 shows that SES effect size tends to be a little larger than the SEC effect. Of the three outcomes with substantive total SEC effects (i.e., ES > 0.10 for the Input model), the effect size for SEC on Academic Performance is more than double than for SES, whereas for College Choice it is about half the size and for 4-year College Enrollment, it is slightly smaller. These findings suggest that the relative effect size of SEC and SES

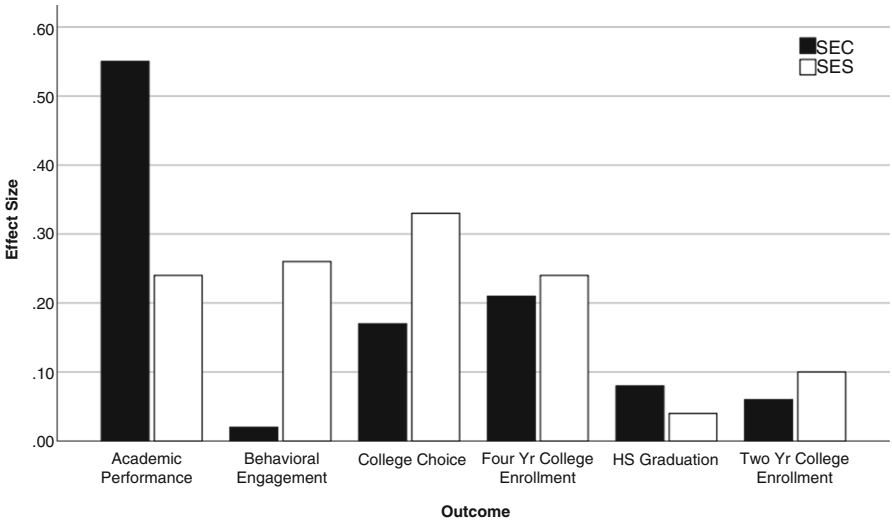

Fig. 11.8 Input model effect sizes for SEC compared with SES

varies depending on the outcome, but on average tends to be similar, a conclusion that is consistent with previous research (Palardy, 2013; Rumberger & Palardy, 2005a). The one clear exception to this generalization is Behavioral Engagement, for which SES has an effect size of 0.26 and SEC is essentially zero. This indicates that misbehavior is more closely linked to one's own socioeconomic background than to the school socioeconomic context.

11.5.2 Implications for Policy and Practice

The statistically significant total effect of SEC for five of the six outcomes indicates that socioeconomic segregation in high school is robustly associated student outcomes and that attending a low-SEC school has consistent detrimental effects. However, the degree to which the SEC effect is mediated by Peer Influences and School Practices varies considerably across the outcomes. Whether Peer Influences or School Practices are the primary mediator has implications to policy for addressing the school segregation. In general, school practices are highly malleable in that they are largely within the control of school personnel. However, peer influences tend not to be and may be difficult to alter without changing the peer composition of the school. That is because peer influences are a function of peer educational resources, attitudes, and behaviors that are shared through normal school interactions, and students attending schools with concentrated poverty tend to have less of those educationally-valuable resources and attributes. Hence, if School Practices account for a substantial part of the SEC effect, adopting effective practices should be the first course of action. That is because altering practices is typically less intrusive, costly, and disruptive compared with altering peer composition. However, if Peer Influences are the primary mediator, addressing its detrimental effects in low-SEC schools will likely require redistributing students among schools to reduce concentrated poverty.

Figure 11.7 and Tables 11.3, 11.4, 11.5, 11.6, and 11.7 show that three of the outcomes have substantial SEC effects that are not accounted for by School Practices: Academic Performance (peer effect mediated = 0.06; unexplained effect = 0.17; net non-school effect = 0.06 + 0.17 = 0.23), College Choice (peer effect mediated = 0.09; unexplained effect = 0.07; net non-school effect = 0.09 + 0.07 = 0.16), and 4-year College Enrollment (peer effect mediated = 0.04; unexplained effect = 0.11; net non-school effect = 0.04 + 0.11 = 0.15). That is half of the outcomes considered in this study and arguably the three with the greatest academic orientation. Hence, socioeconomic segregation in schools is primarily a concern to academic development. These findings suggest that altering school practices will not be sufficient for addressing the negative consequences of socioeconomic segregation for these outcomes and that redistribution of students among schools to reduce socioeconomic segregation may be necessarily.

Fortunately, reducing differences in SEC among schools is feasible. In fact, a substantial number of school districts nationwide currently assign students to schools based in part on income or SES to limit the consequences of concentrated poverty in schools (Kahlenberg, 2012). However, because desegregation plans are typically limited to schools within a given district, for such student assignment strategies to work, schools within districts must initially differ considerably on SEC. The reality of the matter is that SEC varies far more between districts than among schools within the same district. This limits the effectiveness of current intra-district redistribution methods. Inter-district redistribution of students to reduce SES segregation is uncommon and when it does happen it is typically voluntary and generally between adjacent districts with similar student populations and therefore has limited impact on SEC. Inter-district redistribution between low- and high-SEC districts would likely require state intervention. However, past similar "forced" desegregation based on race was very unpopular in the U.S. and had serious unintended consequences. Forced inter-district SES integration could also have unintended consequences such as affluent families moving their children to private schools. Fortunately, however, besides student assignment practices, there are other strategies for promoting socioeconomic integration of schools (see the following for recent reviews: Kahlenberg, 2001; Mantil, Perkins, & Aberger, 2012; Palardy, 2013; Rothwell, 2012). Several of these strategies involve long-term efforts to facilitate neighborhood integration in search of a more permanent solution.

11.5.3 Limitations

A limitation of this study is related to the nature of modeling school compositional effects. Almost all studies of school compositional effects are based on non-experimental data sources and correlational models. That is because large scale random assignment of children to schools, which is the strongest condition for estimating causal school effects, is quite challenging and therefore extremely rare. Estimating causal mediation effects is even more challenging. Beyond random assignment to schools, it generally requires random assignment to a mediation condition or the use of an instrumental variable to tease-out endogeneity in the effect (Imai, Keele, & Tingley, 2010; Park & Palardy, 2020). Hence, true causal mediation is extremely rare in school effectiveness research. For this reason, there is some controversy regarding the true or causal effect sizes of compositional effects such as SEC. The methodological literature suggests that lack of sufficient controls for confounding factors can lead to overestimation of the compositional effects or even to "phantom effects" that disappear once sufficient controls are added (Harker & Tymms, 2004). The substantial number of student and school control measures in the Input model are for the purpose of minimizes this problem and may have results in biased estimates on the conservative side.

Another potential limitation, which is common in studies of compositional effects, is the reliability of the corresponding student measure, in this study, SES.

When the reliability of the student measure is low, its effect on the outcome is underestimated, which typically results in an overestimation of the compositional effect (Harker & Tymms, 2004). That is because the compositional effect is supposed to be above and beyond the student effect, and when the student controls are insufficient, the unaccounted for part of the student effect is largely subsumed by the compositional effect. Unfortunately, estimates of the reliability of the SES measure for ELS: 2002 were not provided in the data documentation. However, given the method by which SES was constructed by the National Center for Educational Statistics, it is expected to be sufficiently reliable (Ingels et al., 2004).

11.5.4 Future Work

This study highlights the importance of examining a range of relevant outcomes when studying school effects. Future work is needed to examine the effects of SEC on other outcomes. Most of the research literature on SEC is at the high school level. Research is needed to understand whether the effects of SEC at the pre-school, elementary school grades (kindergarten through 5th grade), middle school (grades 6–8), and college levels differ from at the high school level. Another issue is, most of the research on SEC assumes its effects are approximately linear. Very little work has explored non-linear effects, which would have important implications for policy for setting targets for minimum school SEC.

11.6 Summary and Conclusions

Throughout much of U.S. history, schools have been segregated by race or class, which has created inequality of educational opportunity. Since the Supreme Court decision on *Brown vs. the Board of Education* (1954) and the civil rights reforms of the mid-1960s, the key concern has shifted from racial segregation by law to de facto socioeconomic segregation due mostly to concentrated poverty in certain neighborhoods. The research literature on socioeconomic segregation focuses mainly on its impact on academic performance. Yet, there are other educational outcomes that are arguably of similar importance for which little research has examined the effects of SEC, particularly attainment and behavioral outcomes, as they are as or more strongly predictive of future employment success and wellbeing than achievement (College Board, 2004). This study is designed to begin addressing that gap in the research literature by examining the effects of SEC on a range of high school and college-going outcomes and the degree to which peer influences and school practices mediate them. Beyond their theoretical implications, differentiating the mediating

roles of peer influences and school practices has implications for educational policy for addressing the negative consequences of socioeconomic segregation.

The results show that SEC is broadly predictive of high school and college-going outcomes, with statistically significant associations with five of the six measures considered in this study. However, the total effect of SEC and the degree to which peer influences and school practices mediate the SEC effects vary substantially across the outcomes. These findings accentuate the importance of examining a range of relevant outcomes when studying school practices, as both the effects of SEC and the mediating mechanisms depend on the outcome.

As summarized in Fig. 11.7, the total effect of SEC is substantial (ES > 0.10 standard deviations) for three of the six outcomes. Of those, the effect size is by far the largest for Academic Performance at 0.55 standard deviations. SEC also has substantial effects on 4-year College Enrollment (ES = 0.21) and College Choice (ES = 0.17), the latter of which measures choice regarding the selectivity of the college enrolled. In addition, SEC has small but statistically significant associations with 2-year College Enrollment and High School Graduation but was not associated with Behavioral Engagement. This pattern of results suggests that the greater the academic focus of the outcome, the larger the SEC effect tends to be. The results also show that School Practices only substantially mediate one outcome—Academic Performance. Peer Influences also only strongly mediate the SEC effect for one outcome—College Choice. Moreover, the part of the total effect not mediated by Peer Influence or School Practices remains substantial for three of these outcomes (Academic Performance = 0.23; College Choice = 0.16; 4-year College Enrollment = 0.15).

These findings have implications for educational policies to address the negative consequences of socioeconomic segregation. Adjusting school practices to emphasize academics, quality teaching, fair discipline, and low disorder will likely reduce the negative consequences of attending a low-SEC school on academic performance by about half. However, relying solely on school practices will likely leave a substantial SEC effect on academic performance and college-going outcomes. Addressing the SEC effect more fully will likely require redistributing students so schools are more similar in terms of SEC. Fortunately, there are currently numerous examples of public school systems in the United States that have implemented such socioeconomic-based school integration policies that can serve as examples (for summaries see Kahlenberg, 2012, Mantil, Perkins, & Aberger, 2012; Rothwell, 2012). However, the underlying problem is neighborhood socioeconomic segregation. Therefore, a more permanent solution may require greater integration of neighborhoods and society (Ainsworth, 2002; Garner & Raudenbush, 1991; Owens, 2018).

Appendix

Table 11.8 Factor score measurement models

Academic performance outcome		
Reading	Reading achievement test score, spring 10th grade	.891
Math	Math achievement test score, spring 10th grade	.907
GPA	Grade point average, 10th grade (transcript data)	.773
Percent of variance explained		77.9
Behaviors engagement outcome		
Attendance	Number of absences the previous term, spring 10th grade	.715
Retained	Retained in 9th or 10th grade	.571
Suspended	Suspended from school during the previous term	.725
Percent of variance explained		58.4
Parental engagement		
BYP57I	Went shopping with 10th grader	.704
BYP57J	Went to restaurants with 10th grader	.696
BYP57K	Spent time talking with 10th grader	.767
BYP57L	Did something else fun with 10th grader	.804
Percent of variance explained		55.4
Facilities hinder learning		
BYA50A	Learning hindered by poor condition of buildings	.665
BYA50B	Learning hindered by poor heating/air/light	.658
BYA50C	Learning hindered by poor science labs	.687
BYA50D	Learning hindered by poor fine arts facilities	.594
BYA50E	Learning hindered by lack of space	.613
BYA50F	Learning hindered by poor library	.582
Percent of variance explained		55.4
Equipment hinders learning		
BYA50G	Learning hindered by lack of texts/supplies	.777
BYA50H	Learning hindered by too few computers	.798
BYA50I	Learning hindered by lack of multimedia	.858
BYA50K	Learning hindered by poor voc/tech equipment/facilities	.735
Percent of variance explained		63.7
Academic Press		
F1A38B	Teachers press students to achieve	.800
F1A38D	Learning is high priority for students	.717
F1A38E	Students expected to do homework	.667
F1A38G	Classroom activities are highly structured	.714
F1A38L	Counselors encourage student enrollment in academic courses	.613
Percent of variance explained		50.3
Teacher morale		
F1A38C	Teacher morale is high	−.754
F1A38H	Many teachers are negative about students	.747
F1A38M	There is often conflict between teachers and administrators	.810

(continued)

Table 11.8 (continued)

Percent of variance explained		59.4
Teacher quality/support		
BYS20E	The teaching is good	.661
BYS20F	Teachers are interested in students	.813
BYS20G	Teachers praise effort	.553
Percent of variance explained		46.8
Teacher efficacy		
BYS20ED	Importance of teacher's attention to student success	.665
BYS20EE	Importance of teaching methods to student success	.836
BYS20EF	Importance of teacher's enthusiasm to student success	.668
Percent of variance explained		52.9
Classroom disruptions		
BYS20D	Other students often disrupt class	.708
BYS20K	Disruptions get in way of learning	.746
BYS20L	Misbehaving students often get away with it	.765
Percent of variance explained		54.8
Disorder		
F1A40K	How often student bullying a problem at school	.647
F1A40L	How often verbal abuse of teachers a problem at school	.878
F1A40M	How often disorder in classrooms a problem at school	.649
F1A40N	How often student disrespect for teachers a problem at school	.840
Percent of variance explained		57.9
Unsafe school environment		
BYS20J	Does not feel safe at this school	.705
BYS20M	There are gangs in school	.820
BYS20N	Racial/ethnic groups often fight	.805
Percent of variance explained		60.6
Discipline fair		
BYS21A	Everyone knows what school rules are	.640
BYS21B	School rules are fair	.701
BYS21C	Punishment same no matter who you are	.742
BYS21D	School rules are strictly enforced	.589
Percent of variance explained		45.0

Note: Items are on 4- or 5-point Likert-type scale

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces, 81*(1), 117–152.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11*, 142–163.

Borman, G. D., & Dowling, M. (2010). Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data. *Teachers College Record, 112*(5), 1201–1246.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America*. New York, NY: Basic Books.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology, 99*(2), 353–395.

Browne, W. J. (2011). *Partitioning of variance in multilevel models*. Retrieved from: https://seis.bristol.ac.uk/~frwjb/materials/wbvpc.pdf

Bryk, A., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.

Bryk, A. S., & Thum, Y. M. (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal, 26*, 353–383.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67–90.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19*, 3127–3131.

Christle, C. A., Jolivette, K., & Nelson, C. M. (2007). School characteristics related to high school dropout rates. *Remedial and Special Education, 28*(6), 325–339.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York, NY: Basic Books.

College Board. (2004). *Education pays 2004: The benefits of higher education for individuals and society*. Washington, DC: College Board.

Cook, T. D., Deng, Y., & Morgano, E. (2007). Friendship influences during early adolescence: The special role of friends' grade point average. *Journal of Research on Adolescence, 17*(2), 325–356.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*(6), 1047–1085.

Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education, 51*, 132–153.

Every Student Succeeds Act, Pub. L. No. 114-195 (2015).

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*, 97–109.

Farkas, G. (2003). Cognitive skills and noncognitive traits and behaviors in stratification processes. *Annual Review of Sociology, 29*, 541–562.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of Education*, 251–262.

Grady, S., & Bielick, S. (2010). *Trends in the use of school choice: 1993 to 2007 (NCES 2010-004)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Hallinan, M. T., & Williams, R. A. (1990). Students' characteristics and the peer-influence process. *Sociology of Education, 63*, 122–132.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics, 18*, 527–544.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 24*(2), 297–327.

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*, 177–200.

Heller, D. E. (2004). Pell Grant recipients in selective colleges and universities. In R. D. Kahlenberg (Ed.), *America's untapped resource: Low-income students in higher education* (pp. 157–166). Washington, DC: Century Foundation Press.

Hill, L. D. (2008). School strategies and the "college-linking" process: Reconsidering the effects of high schools on college enrollment. *Sociology of Education, 81*, 53–76.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309–334.

Ingels, S. J., Pratt, D. J., Rogers, J. E., Siegel, P. H., & Stutts, E. S. (2004). Education longitudinal study of 2002: Base year data file User's manual. NCES 2004-405. *National Center for Education Statistics.*

Jackson, C. K. (2013). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina*. NBER working paper no. 18624. Washington, DC: National Bureau of economic research. Retrieved from https://www.newyorkfed.org/medialibrary/media/research/education_seminar_series/fulltext.Pdf

Jang, S. J. (2002). The effects of family, school, peers, and attitudes on adolescents' drug use: Do they vary with age? *Justice Quarterly, 19*, 97–126.

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In L. Lynn Jr. & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* (pp. 111–186). Washington, DC: National Academy Press.

Kahlenberg, R. D. (2001). Learning from James Coleman. *The Public Interest, 144*, 54–72.

Kahlenberg, R. D. (2012). Turnaround schools and charter schools that work: Moving beyond separate but equal. In R. D. Kahlenberg (Ed.), *The future of school integration: Socioeconomic diversity as an education reform strategy*. Washington, DC: The Century Foundation.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249–277.

Ladson-Billings, G. J. (2006). Yes, but how do we do it? Practicing culturally relevant pedagogy. In J. Landsman & C. W. Lewis (Eds.), *White teachers/diverse classrooms: A guide to building inclusive schools, promoting high expectations, and eliminating racism* (pp. 29–42). Sterling, VA: Stylus.

Legewie, J., & DiPrete, T. A. (2012). School context and the gender gap in educational achievement. *American Sociological Review, 77*(3), 463–485.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3–13.

Lippman, L., Burns, S., & McArthur, E. (1996). *Urban schools: The challenge of location and poverty*. Washington, DC: National Center for Education Statistics.

Mantil, A., Perkins, A. G., & Aberger, S. (2012). The challenge of high-poverty schools: How feasible is socioeconomic integration? In R. D. Kahlenberg (Ed.), *The future of school integration: Socioeconomic diversity as an education reform strategy* (pp. 155–222). Washington, DC: The Century Foundation.

McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany, NY: State University of New York Press.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Taylor & Francis.

McNeal, R. B. (1997). High school dropouts: A closer examination of school effects. *Social Science Quarterly, 78*, 209–222.

Mendenhall, R., DeLuca, S., & Duncan, G. (2006). Neighborhood resources, racial segregation, and economic mobility: Results from the Gautreaux program. *Social Science Research, 35*(4), 892–923.

Mounts, N. S., & Steinberg, L. (1995). An ecological analysis of peer influence on adolescent grade point average and drug use. *Developmental Psychology, 31*(6), 915–922.
Murnane, R. J. (1981). Interpreting the evidence on school effectiveness. *Teachers College Record, 83*, 19–35.
No Child Left Behind Act, Pub. L. No. 107–110 (2001).
OECD. (2019). Income Inequality and Poverty gap (indicators). https://doi.org/10.1787/459aa7f1-en and https://doi.org/10.1787/349eb41b-en. Accessed 23 Mar 2019.
Orfield, G. (2005). The southern dilemma: Losing Brown, fearing Plessy. In J. C. Boger & G. Orfield (Eds.), (pp. 1–25). Chapel Hill, NC: University of North Carolina Press.
Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. *Sociology of Education, 91*(1), 1–27.
Ou, S., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis, 32*(1), 118–139.
Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multilevel, multiple group latent growth curve analysis. *School Effectiveness and School Improvement, 19*, 21–49.
Palardy, G. J. (2013). High school socioeconomic segregation and student attainment. *American Educational Research Journal, 50*, 714–754.
Palardy, G. (2015a). Classroom-based inequalities and achievement gaps in first grade: The role of classroom context and access to qualified and effective teachers. *Teachers College Record, 117*(2), 1–48.
Palardy, G. J. (2015b). High school socioeconomic composition and college choice: Multilevel mediation via organizational habitus, school practices, peer and staff attitudes. *School Effectiveness and School Improvement, 26*(3), 329–353.
Palardy, G. J., & Rumberger, R. W. (2019). The Effects of 21st Century Skills on Behavioral Disengagement in Sacramento High Schools. In Resisting Education: A Cross-National Study on Systems and School Effects (pp. 53–80). Springer, Cham.
Palardy, G. J., Rumberger, R. W., & Butler, T. (2015). The effect of high school socioeconomic, racial, and linguistic segregation on academic performance and school Behaviors. *Teachers College Record, 117*(12), 1–52.
Park, S., & Palardy, G. J. (2020). Sensitivity Evaluation of Methods for Estimating Complier Average Causal Mediation Effects to Assumptions. *Journal of Educational and Behavioral Statistics, 45*(4), 454–486.
Perna, L. W. (2004). The key to college access: A college preparatory curriculum. In W. G. Tierney, Z. B. Corwin, & J. E. Colyar (Eds.), *Preparing for college: Nine elements of effective outreach* (pp. 113–134). Albany, NY: State University of New York Press.
Perna, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education, 76*, 486–518.
Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teacher College Record, 112*, 1137–1162.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Reardon, S. F., & Bischoff, K. (2011). *Growth in the residential segregation of families by income, 1970–2009*. US2010 project, Brown University. Retrieved August 8, 2018, from: https://s4.ad.brown.edu/Projects/Diversity/Data/Report/report111111.pdf
Rothstein, R., & Jacobsen, R. (2006). The goals of education. *Phi Delta Kappan, 88*(4), 264–272.
Rothwell, J. (2012). *Housing costs, zoning, and access to high-scoring schools*. Washington, DC: Brookings Institution, Metropolitan Policy Program.
Rumberger, R. W., & Palardy, G. J. (2005a). Does segregation still matter? The impact of social composition on academic achievement in high school. *Teachers College Record, 107*, 1999–2045.

Rumberger, R. W., & Palardy, G. J. (2005b). Test scores, dropout rates, and transfer rates as alternative measures of school performance. *American Education Research Journal, 42*, 1–42.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*(1), 443–478.

Snyder, T. D., de Brey, C., & Dillow, S. A. (2016). *Digest of education statistics 2015 (NCES 2016–014)* (National Center for education statistics, Institute of Education Sciences). Washington, DC: U.S. Department of Education.

Thrupp, M. (1999). *Schools making a difference: Let's be realistic!* Buckingham, UK: Open University Press.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134–150.

Willms, J. D. (1986). Socioeconomic segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review, 51*, 224–241.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112*, 1008–1037.

Chapter 12
Leadership for Learning in Diverse Settings: School Leaders Setting the Agenda in Australia and New Zealand

Tony Townsend, Mere Berryman, David Gurr, and Lawrie Drysdale

12.1 Introduction

After 40 years of research, policy and practice changes, educational effectiveness and school improvement research has become an influential, though contested, set of understandings about schools and how they might impact on the lives of their students. The initial 'black box' model, which quantified inputs (resources and prior achievements) and outputs (achievement attained) without really considering what happened within the black box of school processes, has become increasingly more sophisticated, with value adding, multi-level modelling, growth curve modelling and quasi experimental research attempting to account for the complexity of student learning. One thing that has been consistent, from the early work on school effectiveness by Weber (1971), Edmonds (1978, 1979), Rutter, Maughan, Mortimore, and Ouston (1979), Reynolds (1982) and Mortimore, Sammons, Stoll, Lewis, and Ecob (1988) and on school leadership by Hallinger and Murphy (1985) and Sergiovanni (1987) to the more recent work of Marzano, Waters, and McNulty (2005), MacBeath and Dempster (2009), MacBeath (2010), Robinson, Hohepa, and Lloyd (2009), Day et al. (2010) and Townsend and MacBeath (2011), is that the important inter-relationships between leadership and student learning have been confirmed over and over again.

T. Townsend (✉)
Griffith Institute for Educational Research, Griffith University, Queensland, Australia

University of Tasmania, Hobart, Australia
e-mail: t.townsend@griffith.edu.au

M. Berryman
Waikato University, Hamilton, New Zealand
e-mail: mere.berryman@waikato.ac.nz

D. Gurr · L. Drysdale
Melbourne Graduate School of Education, University of Melbourne, Melbourne, Australia
e-mail: d.gurr@unimelb.edu.au; drysdale@unimelb.edu.au

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_12

While leadership has widely been treated as synonymous with headteachers or principals (e.g. Townsend & MacBeath, 2011), researchers have in recent years, widened the compass to examine student leadership (Mitra, 2007), teacher leadership (Barth, 1999; Little, 1990; Wenner & Campbell, 2017), district or local authority leadership (Waters & Marzano, 2007) and government or system leadership (Hopkins, 2010; Southworth, 2005), each being identified as helping to increase the level of student achievement within schools. There has also been a plethora of leadership types, hierarchical or transactional (Silins, 1994), transformational (Leithwood & Janzti, 2000), shared (Lambert, 2002), distributed (Gronn, 2000, 2002; Spillane, Halverson, & Diamond, 2001, 2004), moral (Sergiovanni, 1992), ethical (Starratt, 2004), and democratic (Møller, 2002; Starratt, 2001), among others, that have found their way into the educational literature.

12.2 Leadership for Learning

It could be argued that the research has identified several shifts in the way leadership is applied in schools: from the concept of a principal being *the* school leader (that is, in a *position* of leadership) to one where leadership is seen as an *activity*, one that is not just the domain of the principal, but others as well; from one of leadership being the responsibility of a single person who oversees everything to one where people work together to ensure that requirements are covered; and from one of leadership as requiring generic skills (and so principals can be equally successful regardless of the school they work in) to one of leadership being context and purpose specific. The term instructional leadership is still seen as being a dominant form of school leadership spanning many countries, but in recent times a new terminology has been used for the leadership of schools, one called *Leadership for Learning*, which has been used for more than a decade in the United Kingdom (see, for example MacBeath & Dempster, 2009; Townsend & MacBeath, 2011). However, Hallinger and Murphy, who began using the term instructional leadership in the mid-1980s, have both recently, and separately, used the term leadership for learning in their writing (see Hallinger & Heck, 2010; Murphy, Elliott, Goldring, & Porter, 2007). Hallinger & Heck (2010, p. 657) argued

> In recent years, the phrase 'leadership for learning' has gained international currency (MacBeath et al., 2008; Robinson et al., 2008). In our view, this approach to school leadership represents a blend of two earlier leadership conceptualizations: instructional leadership and transformational leadership.

Murphy too, has recently used the term 'leadership for learning' (Murphy et al., 2007, p. 179):

> This type of leadership can best be labeled 'leadership for learning', 'instructionally focused leadership' or 'leadership for school improvement.' The touchstones for this type of leadership include the ability of leaders (a) to stay consistently focused on the right stuff - the core technology of schooling, or learning, teaching, curriculum and assessment and (b) to

> make all the other dimensions of schooling (e.g. administration, organization, finance) work in the service of a more robust core technology and improved student learning.

However, MacBeath & Townsend (2011, pp. 1249–1250) argue that terminology IS important and that the difference between instructional leadership and leadership for learning, is that the focus of instructional leadership is on leadership and the focus of leadership for learning is on learning. They argue:

> Whereas much of the instructional leadership literature reduces learning to 'outcomes', leadership for learning embraces a much wider, developmental view of learning. Nor is its focus exclusively on student achievement. It sees things through a wide-angle lens, embracing professional, organisational and leadership learning. It understands the vitality of their interconnections and the climate they create for exploration, inquiry and creativity. Its concern is for of all those who are part of a learning community.

This is not the place to argue the case for one or the other in detail, but we wish to simply point out that such is the complexity of school leadership that even those that have been researching it for decades have not yet come to full agreement about what it means. Perhaps the most challenging task of leadership is to address continuing social inequalities common to all countries but manifested in different ways and with differing consequences. The historic triumvirate of gender, class and race remain to a greater or lesser extent, playing out in distinctive forms in relation to the socio-economic factors and the role and potency of schools in addressing them. As Enomoto (1997) describes them, schools are 'nested communities', not only internally but within local neighbourhoods, within local administrations/authorities or districts, states and countries, but also increasingly within the global competitive policy environment (also see Hallinger's discussion of the multiple contexts that influence school leadership; Hallinger, 2018).

With the nature of education changing, as more responsibility is placed at the school level, the role of the school leader has expanded and become more complex to the point where Townsend (2016) argued that leading a school in the twenty-first Century is akin to having to drive a car in the fast lane (to keep up) but needing to do it carefully to ensure that all of the passengers arrive safely. Part of the rapidly changing environment that school leaders are now facing is increasing diversity within and across schools. As Hopkins, Harris, and Jackson (2010) argue, schools at different stages of growth require different types of leadership and recognizing the context in which the school works is one of the key issues for any new leader of a school. This chapter explores some of the issues related to leading in diverse contexts, but also tries to establish some common ground for leaders, regardless of the school in which they might work. The chapter starts by considering what successful leadership looks like in Australia and New Zealand and then considers two case studies, one from each country, where successful leadership practices have been fostered and developed.

12.3 Successful School Leadership in Australasia

A focus on notions of leadership and school success is a relatively recent phenomenon in Australia and New Zealand. For the Australia context, reviews such as Gurr, (2009) and Gurr and Drysdale (2016) describe this history. In the 1960–1970s the focus was on supervision; a good school had good staff, which the principal helped to create (Bassett, Crane, & Walker, 1967). In the 1980s and 1990s there was a focus on school effectiveness and instructional leadership, in an effort to improve teaching and learning, and indirectly, student outcomes (e.g. Beare, Caldwell, & Millikan, 1989; Duignan et al., 1985). From the 2000s, whilst there has been a focus on leadership and school effectiveness in terms of improving student learning outcomes, there has been research on the broader concept of school success, and the complexity of leading schools in a rapidly changing environment (Duignan & Gurr, 2007). There is also evidence that successful middle-level leadership can foster a range of positive outcomes that go beyond improving teaching and learning (e.g. Dinham, 2007; Gurr & Drysdale, 2013). It has become clear that principal leadership, and leadership more broadly, contribute to success and effectiveness. In this section we report on continuing work in this area through consideration of findings from Australian and New Zealand research as part of the International Successful School Principalship Project (ISSPP).

The ISSPP has been actively researching the work of successful principals since its construction in 2001 (www.uv.uio.no/ils/english/research/projects/isspp/). Stimulated by the success of an earlier study (Day, Harris, Hadfield, Tolley, & Beresford, 2000), Day wanted to explore on a large scale the characteristics and practices of principals leading successful schools, and so assembled a group comprising researchers from seven countries: Australia, Canada, China, Denmark, England, Norway, and Sweden. This group agreed to conduct multiple perspective case studies focused on the leadership of principals in successful schools. Principals were selected using evidence of student achievement beyond expectations on state or national tests, principals' exemplary reputations in the community and/or school system, and other indicators of success that were country and site-specific (such as school review/inspection reports); once in the schools, the evidence of success indicated a broad array of student and school outcomes (see Gurr, 2015). Each case involved individual interviews with the principal, senior staff and other teachers (6–8), school council/board members (2), and group interviews (two parent and two student groups), as well as an analysis of relevant documents. Later cases also included observation.

The project began because at the time what was known about principal leadership relied too much on studies that only used principals as the data source, and too much of the literature was derived from studies in North America and the United Kingdom. Gathering the opinions of others in the schools (school board members, teachers, parents and students), and doing this across several countries, was a way to extend and enhance knowledge of the contribution of principals to school success. The project continues today with active research groups in 23 countries, producing more

than 100 case studies, and nearly as many papers, book chapters and books published, with four project books, and seven special journal issues in English, as well as books and journals in Spanish, Norwegian, Danish, and Swedish.

In Australia, fourteen initial case studies were conducted in the states of Victoria and Tasmania between 2003 and 2005 (Gurr, Drysdale, & Mulford, 2005, 2006). The five case studies in Tasmania were conducted under the leadership of Bill Mulford, and nine cases from Victoria were conducted under the supervision of David Gurr and Lawrie Drysdale. Schools included government, Catholic and independent schools, from the primary, secondary and special school sectors. Three of the schools in Victoria were subsequently revisited to explore the sustainability of success (Goode, 2017). In recent times, the focus was on exploring leadership in schools that have histories of underperformance, but which are on an improvement journey. Two cases were completed on schools in communities with high educational advantage, and one case was in a school with low educational advantage (Gurr, Drysdale, Longmuir, & McCrohan, 2018a, 2018b; Longmuir, 2017).

Led by Ross Notman, New Zealand has contributed thirteen cases to the ISSPP including one specialist school, one early childhood centre, one intermediate school, six primary schools and four secondary schools. The early childhood example is the only one in all of the ISSPP cases. Findings from an initial five cases were published in Notman and Henry (2009) and Notman (2012), and a further seven cases were published in a ten-case edited book Notman (2011a), and one additional case can be found in Notman (2014).

From the Australian and New Zealand cases it seems that principals were contextually aware of their environment and shared a set of values and beliefs, personal qualities and had a range of practices or interventions that contributed to success. We can identify common features of successful school principals centred on: values and beliefs; personal qualities and skills; interventions/practices that lead to success; and, capacity building. The first three features (values, qualities and skills) are to do with principal identity. These are personal qualities and characteristics attributed by participants to the principal that shaped their perception of positive leadership which enabled the principals to influence and have impact. The other aspects are more to do with what principals do – their practices and interventions. In the end, their ability to be successful (and effective) is a combination of their personal factors and their behaviours within the school and broader context in which they are operating. Successful school leaders interact within a particular school context to deliver strategic interventions aimed at improving student outcomes.

12.3.1 Principal Contribution

Principals made a difference and contributed to success by being a positive influence on the quality of education in the school (Gurr et al., 2006). The contribution was manifest in aspects such as improving the image of the school, setting new direction

through a common vision, establishing high expectations, building school capacity (especially in regard to staff development), re-organising the school, and focusing on improving teaching and learning. Notman (2011b) described how New Zealand principals articulated an overarching vision and communicated this clearly to the school/centre community. They employed strategies that focussed on cultural change such as being culturally responsive to demographic changes in the school community and using an ethic of care to promote a positive culture. An acute contextual awareness (both internal and external), resulted in a strong sense of advocacy for students and the school community. In most cases across the two countries, school communities identified the principal as the 'engine room' of school improvement and change. School communities were able to identify milestones and achievements attributable to the principal.

12.3.2 Values

Sergiovanni (1991) noted that style itself is less important than what the principal stands for, believes in and communicates to others. This was clearly one of the key findings from the Australian case studies. Initially we categorised this as the principal's personal philosophy (Gurr & Drysdale, 2007). Subsequently we have defined them as values and beliefs (Drysdale & Gurr, 2011). The successful leaders were able to clearly articulate their values and were observed to act in accordance with their values. The values were perceived on multiple levels. For example, they expressed core values, such as respect for others, fairness, trustworthiness and responsibility. But they also had universal values, such as social justice, dignity and freedom, empathy for the less well off, compassion and tolerance. Other levels included professional values and beliefs (service to staff, acceptance of diversity, accepting constructive feedback from others, maintaining confidentiality) and social and political values (respect for life and the environment; respect for minority rights; respect for the law).

12.3.3 Qualities and Skills

Gurr et al. (2006) found that particular personal qualities and characteristics seemed important for the success of the leadership of principals, and Belchetz and Leithwood (2007) noted these features were important, not so much for what leaders do, but for how they do it. Gurr et al. (2006) highlighted traits such as passion, optimism, enthusiasm, persistence, determination and assertiveness. The leaders were people-centred, good at developing relationships, modeling appropriate behavior, and establishing relational trust. They could articulate their core beliefs and values and demonstrate these through their actions. Critical self-reflection, and personal resiliency were important elements of their successful practice (see

Notman, 2012, for a detailed discussion of the interpersonal factors that contributed to leadership success).

12.3.4 Interventions/Practices

Values, qualities and skills are only part of the equation. Who they are is important, but significantly what they do and how they do it also determines their success. The principals acted purposefully and strategically. They engaged in a series of interventions that reflected the contexts and the needs of their schools. Gurr et al. (2006) identified personal, professional, organisational and community capacity building to be common interventions in all the Australian case studies, and these have been included in all the major models describing the ISSPP research (e.g. Drysdale & Gurr, 2011; Gurr, 2015). Notman (2011b) described how principals promoted teacher quality through recruiting, inducting, developing and motivating teaching staff; and built individual capacity among staff through professional development and use of distributed leadership practices. Notman (2011b) used pedagogical leadership to describe the core focus of the work of the New Zealand principals. Aspects included in this were a vision for teaching and learning that aims to increase student achievement, an orientation to the possibilities and opportunities rather than limitations of government curriculum mandates, fostering staff collaboration through stimulating learning conversations amongst staff, encouraging explicit sharing of pedagogical strategies and the use of assessment data to guide student learning programs. In the early childhood and primary settings, building school-parent partnerships to support children's learning was also considered.

While the major focus of these case studies was the principal, an important finding was that leadership was cast more broadly. Many of the principals empowered staff and embraced shared decision making and shared leadership in ways that supported distributed leadership approaches. In many cases there was a close and positive relationship between the principal and the assistant principal/s.

12.3.5 A Model of Successful School Leadership

Figure 12.1 shows a model that has been constructed from consideration of various models produced by groups within the ISSPP and description of the various models is provided in Gurr (2015).

The model shown in Fig. 12.1 has a division between the why, how and what of successful schools articulated by Mulford and Johns (2004), and the use of impact levels from Gurr et al. (2003), moving from the least direct on learning outcomes (level 3, wider context), to level 2 (leadership and management), and then level 1 (teaching and learning) which, of course, directly impacts on student outcomes. The use of impact levels is helpful in locating the mainly indirect impact on student

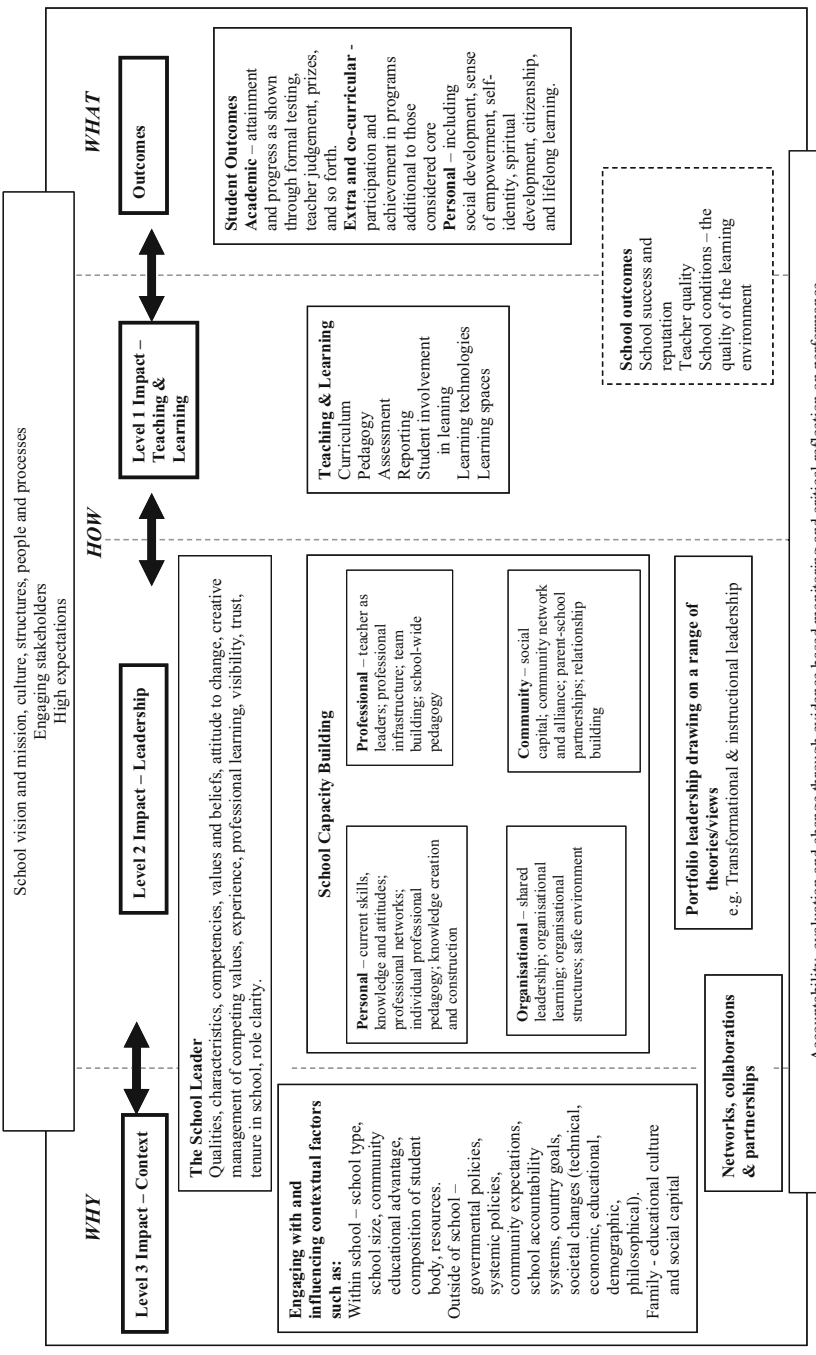

Fig. 12.1 The International Successful School Principals Project (ISSPP) model of successful school leadership from Gurr (2015)

learning of principal leadership across levels three and two, and the more direct impact of middle level leaders across levels two and one. The context description in level three describes the school, family and external contexts that leaders need to respond to and influence. Networks, collaborations and partnerships are located across levels three and two. Level two has an emphasis on the capacity building of teachers and other adults in the school, has detail about the characteristics and practices of school leaders, and encourages a portfolio approach to using leadership styles. At level 1 there is clarity about teaching and learning, with this including student involvement in learning, and the nature and quality of the spaces and technologies that support teaching and learning. Outcomes include broad descriptions of student learning outcomes, and also includes school outcomes. Use of evidence based monitoring and critical reflection remains across all levels. Across the levels there are elements describing the use of evidence based monitoring and critical reflection, the general nature of the school (the shared vision and mission, and culture of the school and the structure, people and processes that make this), the engagement of stakeholders within and outside the school, and the promotion of high expectations for all.

12.4 Australia: Principals as Literacy Leaders (PALL)

The Principals as Literacy Leaders (PALL) project emerged in 2009 as a response to the Australian federal government's concern about a gap in reading achievement demonstrated in successive results from the Program for Student Assessment (PISA) studies. These results prompted the then Labor government to establish a series of projects under the "Closing the Gap" program. The Principals as Literacy Leaders (PALL) Pilot Project was an idea developed by the Australian Primary Principals Association (APPA) (Dempster et al., 2017) and established a partnership that included the Department of Education and Children's Services, the Australian Primary Principals Association, three universities, and the state, Catholic and Independent school authorities from four states, Queensland, South Australia, Western Australia and the Northern Territory (Dempster et al., 2012). Sixty principals, from those four states completed the program of professional learning in 2010. Since the pilot program, the PALL professional learning activity has been offered in every state of Australia.

The program was based on the research evidence from a range of studies showing that principals' leadership can impact student achievement (e.g., Leithwood, Day, Sammons, Harris, & Hopkins, 2006) together with a body of research indicating that the quality of school leadership (particularly distributed leadership and leadership for learning; e.g. MacBeath & Dempster, 2009; Seashore-Louis, Leithwood, Wahlstrom, & Anderson, 2010), and well-designed professional learning and targeted support programs (Hord, 1997; Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009) are vital in progressing students' reading improvement. Linked to this view of leadership were government reviews of reading and literacy

(e.g. National Institute of Child Health and Human Development (NICHD), 2000; Rose, 2006; Rowe, 2005) that argued that secure knowledge and skills in five components of phonological awareness, letter-sound knowledge (alphabet and phonics), vocabulary, comprehension, and fluency, were pivotal for reading acquisition, and that how teachers design and structure their reading programs can have a significant impact. Routman (2014, p. 1) made the point that "teachers must be leaders, and principals must know literacy [because] without a synergy between literacy and leadership and a committed, joint effort by teachers and principals, fragile achievement gains do not hold." An acceptance of these principles suggests that to improve reading performance in schools, school leaders need to have not only leadership knowledge and skills but also an understanding of how students learn to read and knowledge of ways in which teachers might be supported in this key learning area.

From the research identified above five positions emerged that were adopted for the development and implementation of the PALL program, namely:

- That the role of the principal (and leadership team) is absolutely critical for the improvement of student learning by establishing a clear moral purpose within the school.
- That high levels of learning improvements will occur if leadership is shared with others in the school.
- That there are certain specific factors (the BIG 6) that must be given explicit attention by teachers if students are to learn to read well.
- That for improvement to occur, interventions must be planned and implemented. A school cannot expect improvement by continuing to do what it has always done.
- That for school leaders to undertake learning improvements successfully, they will need support to improve their own learning and to implement appropriate strategies.

The PALL program is essentially a 2-year program of learning and implementation. The first year consists of the learning and planning phase during five one day modules over the course of a school year. During this year a specific plan for reading intervention is developed, to be implemented in the second year. The professional learning was designed to provide school leaders with the needed content knowledge about the six key areas of reading (the BIG 6) – oral language, vocabulary, phonological awareness, letter-and-sound knowledge, comprehension and fluency - together with an understanding of strategies of distributed leadership that would see teachers, as well as leaders, leading interventions in reading. The five PALL positions were used to create a Leadership for Learning Blueprint (LfLB) that became a central feature of the PALL program, one that could be used by school leaders to implement change back in schools. The modules were as follows:

Module 1: Leadership for learning – What does this mean?
Module 2: What leaders need to know about learning to read
Module 3: Leading reading data gathering and analysis

Module 4: Designing, implementing and monitoring literacy interventions
Module 5: Evaluating and reporting on reading interventions

Module 1 introduced one of the key elements of the program, the Leadership for Learning Blueprint (LfLB), one synthesised to connect leadership to learning. The LfLB is illustrated in Fig. 12.2 below.

At the centre of leaders' work is their commitment to the *moral purpose* of improving the lives of students through learning. To do this rests on a commitment to focused professional conversations or "*disciplined dialogue*", always stimulated by *strong evidence* of what students can or cannot do, so that what they need to learn next is well informed. Surrounding this central core is a commitment to *active professional learning* by school leaders and members of staff, an understanding that *shared leadership* is essential in schools and that all structures and processes should be organized accordingly. When this is undertaken, a clear commitment to *a well-planned reading program with teaching and learning carefully coordinated and monitored* is essential, as is a concentration on creating helpful and *supportive conditions* for students' learning, by developing the physical, cultural, social and emotional learning environment through strategic resourcing. The last of the dimensions refers to the importance of making *connections beyond the school*: with families, their communities, and with other agencies which may make different but necessary contributions to improvements in learning. All of these dimensions combine to make up a complex agenda for leaders and teachers who want to make a difference to the lives of learners in the contexts in which they work.

Subsequent modules provided the evidence-base in support of the BIG 6 by highlighting the research endorsing the need to include ongoing quality teaching of the six elements of reading (module 2), consideration of the use of formal and

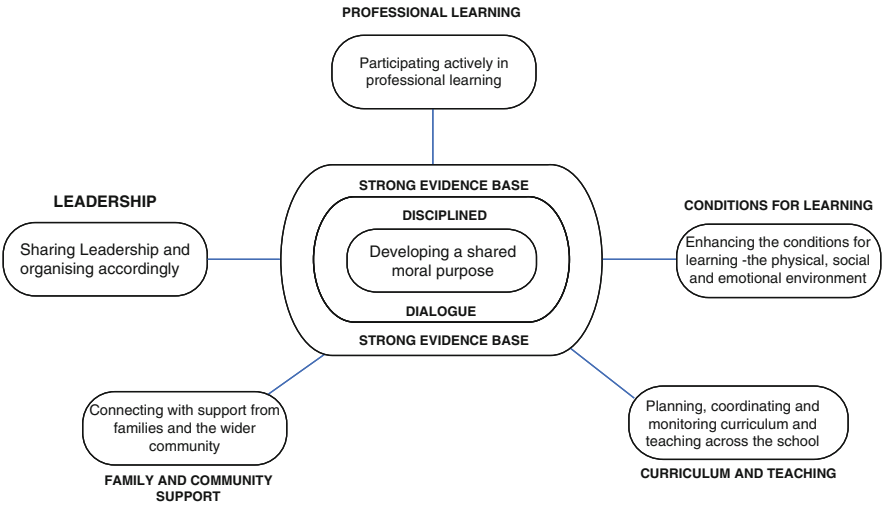

Fig. 12.2 The Leadership for Learning Blueprint (LfLB)

informal assessment processes and the logic of having whole-of-school agreements about what was assessed and when and how this diagnostic information could be shared with students, families, and colleagues (module 3), the need for planning interventions to ensure that specific concerns identified through these processes were addressed (module 4) and school-based evaluation of the interventions as a means for reporting progress and making decisions about future steps (module 5). Between each of the modules, identified literacy advisers would work with school leaders to provide support and encouragement to use the skills, tools and knowledge from the modules, adapted in ways to suit the particular context of their school. PALL has expanded over time to include different types of schools and school contexts, and a wider variety of school leaders. Since 2010, more than 2000 school leaders from around Australia have been involved in PALL.

12.4.1 PALL Research

Alongside the professional learning activity, there have been numerous research studies related to the implementation of the PALL program, to identify its efficacy and to look at its effect on leader and teacher behaviours as they relate to the teaching of reading (Australian Primary Principals Association (APPA), 2013; Dempster et al., 2012; Dempster, Johnson, & Stevens, 2014; Johnson et al., 2014; Townsend, Dempster, Johnson, Bayetto, & Stevens, 2015a; Townsend, Wilkinson, & Stevens, 2015b; Townsend, 2017). The results of the first six of these studies were compiled into the book Leadership and Literacy: Principals, Partnerships and Pathways to Improvement (Dempster et al., 2017). The data collected over the years has been drawn from both quantitative and qualitative sources. The most consistently used piece of quantitative data collection has been the Personal Leadership Profile (PLP). It contains 36 statements related to the elements identified in the Leadership for Learning Blueprint. Participants in the PALL program were asked to complete the PLP in module 1, and in a number of different studies, were also asked to complete it again in module 5. In each instance, where the data has been collected twice, they indicate there has been a growth (near or above 20%) in the school leaders' perceptions of their ability to lead their schools in ways that will support improvements in reading (for details see Figure 8.2 in Dempster et al., 2012, and Table 1, Townsend, 2018). Qualitative data was collected in various ways, such as interviews with participants or literacy advisers, collection and analysis of schools' intervention plans and through case studies of schools that were implementing an intervention during the second year of the program. Data collected from the case study schools included interviews with school leaders, teachers and parents, student surveys of their attitudes towards reading, reading achievement data and samples of changes in student work over the course of the year.

12.4.2 Results from the Data

Student Learning. As the 2 years of the PALL program was not sufficient time to establish whole school trends in student achievement, determinations about the impact of the PALL program on educational effectiveness looked at some broader parameters, such as student engagement, changes in pedagogical and assessment practices, how the school encouraged a positive climate, and so on, as it applied to reading. Dempster et al. (2017, p. 169) reported:

> ...increases in student achievement in reading were recorded; in every report, changed teaching practices (particularly increased knowledge about diagnostic data sources and increased ability to interrogate data for individual students as well as for whole classes and whole schools).

They also argued there was:

> ...evidence of shifts in teaching practices in many schools in remarkably diverse contexts across six states and one territory... [and] also demonstrated how these changed teaching and assessment practices have improved the conditions for learning for thousands of students across Australia.

There was substantial data collected over the seven studies to indicate that the five PALL positions described above could be ratified by the research data. What follows is a brief overview of the findings from the collective research on each.

Moral Purpose. Data collected over the course of the PALL research program of seven studies included participants responses to the Personal Leadership Profile taken at the beginning and at the end of the professional learning, conversations with Literacy Advisers and with PALL participants and others in the case study activities. It was clear that a great deal of attention was paid to the moral purpose of leadership as described by a focus on the purposes, goals and expectations related to reading. In the Pilot program, more than 1500 conversations addressed these issues which resulted in a clear improvement over time for principals' perceptions of their ability to promote factors associated with moral purpose, setting high expectations, collaboratively building vision and setting directions, seeing that goals are embedded in school and classroom routines, and ensuring consensus on goals.

In addition to the quantitative data related to moral purpose, numerous responses from the case study research also suggested a strong focus on the development of moral purpose. Dempster et al. (2017, pp. 178–79) summed up the implications of these findings.

> First, reconnecting with the moral purpose of schools becomes tangible when it is linked with a learning priority as foundational as reading... Second, finding ways to embed the goal of reading improvement in classroom practice at every year level should be a constant demand to which all members of staff are able to respond. Third ... Principals need to develop a deep understanding of the capabilities of their teachers in order to ensure that all are able to play their part in the pursuit of reading improvement.

Of particular interest for this element of the Blueprint is the Principals as Literacy Leaders in Indigenous Communities (PALLIC) study (Johnson et al., 2014) which

included seven case studies, as these studies helped to identify the important role of context within diverse school communities. In the case of PALLIC, the *PALLIC professional development has created awareness and has raised the profile of Indigenous staff at the school and increased their confidence in their work*, and the term "both ways leadership" was used to recognise the explicit recognition that a school needed Indigenous Leadership Partners (ILPs), where the leadership partner was both embraced by the principal and accepted by the indigenous community. One of the contextual complexities came from each of the PALLIC schools having its own terminology for Indigenous leadership partners. Indigenous Leaders were referred to as Partners (ILPs), Aboriginal Education Workers (AEWs) and Teaching Assistants (TAs). In one PALLIC school an ILP had her own desk in the principal's office and the principal even said *. . .She [the ILP] is the Aboriginal version of me*, and another principal reported *the Indigenous leaders are just as important as the principal.* Some schools had up to three ILPs and others could not get any. This was further complicated by the fact that in many indigenous schools, staff turnover, and leader turnover, was much higher than in other schools, but people believed that the "both ways leadership" could overcome this problem. In one case, as reported in Dempster et al. (2017, p. 87),

> . . .the principal indicated how the ILPs had embraced the PALLIC Program, recommended it to the community and worked alongside the teachers to implement improvement strategies. The principal was unclear how the two ILPs had been able to sell PALLIC's idea to their community, but was nevertheless somewhat confident that if he, as principal, were to leave, the two ILPs would promote the continuation of the PALLIC Program to a future leadership team.

This particular study helped to show that the context in which leaders work to improve literacy, the students, the families and the community, needs to be considered, respected and involved if success is to be achieved.

Shared Leadership. It is clear from the data collected that there were many examples of shared leadership being undertaken and that those involved in the PALL program had higher levels of confidence in doing this after completing the PALL program. The position taken on shared leadership implies the acceptance of the need for leadership depth and breadth – depth within the school and breadth beyond its boundaries. PALL research interviews are replete with terms such as "same page", "common language", "team work", "team planning", "professional conversations", "agreed strategies and solutions", "trust in each other", "collective responsibility" and so on. All of these terms point indisputably to leadership as a shared activity within the school when the goal is reading improvement. In contrast, the collective research uncovered only a small number of concerted efforts to move outside the gates of a single school. However, when and where this occurred, with pre-schools, other child agencies and with pre-school parents directly, the value to principals and teachers was reported enthusiastically, which suggests that more schools might benefit from attempts to encourage families to be more involved. The PALL Learning in Families Together (LIFT) program offered in Tasmania from 2017, and its associated research, might assist us to better understand how schools might encourage this partnership further.

Dempster et al. (2017, p. 181–2) sum up by saying

> ...we can say with certainty that most principals and teachers expanded their capacity to share leadership within their schools, thus seeing leadership as activity, not position. In many schools, leadership depth was increased amongst teachers from the early childhood years to those in the upper school, with structural arrangements in place to ensure that communities of practice took responsibility for the implementation of agreed strategies and planned interventions in reading.

Learning to read. The PALL studies overall revealed an unevenness of principals' and teachers' knowledge about reading, which reemphasized the important role that principals need to take but also the need to share leadership of this important foundational activity. However, the case study data showed numerous instances of school leaders' and teachers' willingness to learn and also to take responsibility for leading student learning to higher levels.

Dempster et al. (2017, pp. 183–4) identify five messages from the data collected for this position:

1. school principals need to have a high level of understanding about what it means to teach reading if they are to lead staff in their schools to strengthen student satisfaction and achievement;
2. teachers require targeted professional development in specific methodologies for teaching the fundamentals of reading, dependent on capabilities, over an extended period of time;
3. interventions in the teaching of reading should be based on robust evidence about students' capabilities in learning to read and teachers' knowledge about the explicit teaching of known areas of student needs;
4. knowledge about the cultural and social context for the teaching of reading should take account of student data and conditions for learning in particular schools and communities; and
5. a strengths-based approach to engaging parents and community members in the teaching and support of reading is likely to result in a more productive take up than the more familiar deficit alternative.

A major implication is that there is a clear need for national and state administrators to reconsider a national strategy to support professional learning for the teaching of reading. It is also clear that still more must be done to engage parents and members of the community in supporting children's reading.

Taking an intervention approach. Taking an intervention approach brought together a number of elements of the LfLB, using strong evidence to make decisions, then focusing on how to resource, and change, the conditions for learning, the curriculum, teaching practices and family engagement. A growing confidence in each of these areas was demonstrated by the increase in positive responses to the Personal Leadership Profiles of PALL participants, especially in relation to the element seeking to improve parent and community support.

The PALL research showed that that successful interventions in reading have been based on sound qualitative and quantitative evidence. While some schools had previously been using evidence to guide their teaching and learning programs, they

had not necessarily used it consistently or strategically. PALL engendered sharper and more purposeful data collection and analysis, and the ability to then use disciplined dialogue in ways that enabled planning for improvement and future decision-making about student learning. Many schools mentioned the use of a whole-school approach as making a significant contribution. Consistent literacy practices, such as dedicated literacy blocks and the use of a common language right across the school were common. However, one of the main findings is that changing teaching practices takes both time and support and that schools systems need to allow time for changes to be embedded rather than demanding instant results.

Leadership support. Leadership support for literacy improvement comes in many forms and the PALL research highlighted the importance of some of them. There was a constant reference from school leaders across the studies of the need for leaders to have the knowledge required to make the changes necessary for improved literacy results. The PALL program itself was one of these forms of support, providing school leaders with BOTH what they needed to know to support strong reading outcomes in their schools AND ways in which the leadership of this effort might be shared by teachers and others in the school. They were provided with evidence-based knowledge about the BIG 6 and the need for, and strategies to support, the development of a shared moral purpose as it applied to literacy through improved data collection and use in a disciplined way. The literacy advisers also played a critical role in the process, providing ongoing support for the building of knowledge and skill over the program's life. A clear implication is that focused professional development for principals can have a substantial impact on how schools approach their leadership of school improvement. A second implication is that the ongoing use of mentors, particularly for new principals might heighten growth within schools when specific goals are being identified by systems or by local authorities.

The value of PALL. Dempster et al. (2017, p. 191) argue that the research data, collected now over 7 years, suggest four themes that emerge from the program:

1. *Enhanced leadership for learning expertise;*
2. *Leadership partnerships and learning pathways;*
3. *The power of blended leadership learning programs; and*
4. *The need to rethink parent engagement.*

From the first of these, the data suggest that those involved in the program experienced a heightened competence, and confidence, to implement approaches to improving literacy in their schools. The second theme encompasses the PALL position that leadership is an activity, not one that is solely defined by the position one holds. What the research has found is that the vast majority of leaders are willing and supportive of sharing leadership of activities related to the improvement of literacy and that the majority of teachers are willing to take on these leadership actions as well. The third theme draws attention to the value of ensuring that leadership at its best is connected to the daily work of the school. Generic leadership actions, when specifically applied to student learning and with the focus of having research-validated knowledge about literacy, raises the profile of the literacy

development activity right across the school. Blending generic leadership with knowledge of reading enables school leaders to work side by side with teachers as they work out ways to support their students' needs to become more literate. Finally, the fourth theme clearly focuses on one of the critical issues that has yet to be solved, how schools can engage parents in the activity of supporting their child's learning. It is historically hard to engage parents in this, for many reasons, some school-based and some home-based. However, it is clear that for parents, families and communities, especially those in disadvantaged circumstances, home-based support to assist their children to learn would be welcomed. Dempster et al. (2017, p. 191) report:

> A clear need revealed in our research is that although parents have relevant talents, their knowledge of schools and the learning that is taking place in them is often lacking when juxtaposed with "disengaged parents" data. Understanding this finding, the approach taken must be strengths based if the ultimate goal is for parent- and community-led engagement initiatives to become a reality.

The ultimate test of PALL will come later, when we start to see data from the schools that have been involved in the PALL research over time, particularly from schools where PALL, and the BIG 6 approach, has been given time to embed itself into the culture of the school. Early findings suggest there are significant improvements (Townsend, 2017).

As Dempster et al. (2017, p. 193) conclude:

> . . .as we start to see data on students' enjoyment of reading, their reading habits in their spare time at school and at home, and the confidence of their parents in supporting their children's reading. If we are able to see these things, we are also likely to see positive changes in student reading performance on whatever measure is employed. Although this was not the primary purpose of PALL, it is a result to which we aspire.

The work of PALL continues, with professional learning and research being conducted in both Tasmania and South Australia. From the original base of 60 primary principals from disadvantaged communities, PALL has grown and expanded its reach to secondary and early childhood leaders and has developed programs specifically aimed at Indigenous communities, the middle years and for schools wishing to encourage greater engagement with families. Further quantitative data from participants and teachers from their schools and more longitudinal case studies will help us to further refine our understanding of how school leaders translate their learning, about leadership and literacy, into strategic programs within their schools, in ways that will support teachers to improve student engagement, involvement and learning, ultimately to see them improve their achievement in this critical area of human development.

12.5 New Zealand: Improving Contexts for Learning for Maori Students

Achievement disparities between specific groups of students in New Zealand education continue, over time, to be of concern. Of particular concern, as in many other countries internationally, is the achievement of Indigenous students. New Zealand's Indigenous Māori students do not do as well in the education system as do other students. The Office of the Auditor-General consistently reports that Māori students do not remain in schooling as long as other students nor are they achieving as highly (Auditor-General, 2012, 2013). In 2016, across all ethnicity groupings, Māori students were the lowest proportion of students remaining at school to age 17 (70.9%). This compares with a retention rate of 85.4% for European students (Ministry of Education, 2018a). Māori are also over-represented in our national stand-down[1] and exclusion[2] figures. In 2016, the age-standardised stand-down rate for Māori (37.3 stand-downs per 1000) was 2.4 times as high as Pākehā (students of the colonial settlers) (15.7 stand-downs per 1000). In the same year, the Ministry of Education reported that the age-standardised exclusion rate for Māori (3.0 exclusions per 1000) was 3.4 times as high as for Pākehā (0.9 exclusions per 1000) (Ministry of Education, 2018b).

Despite many initiatives to raise Māori student achievement, English-medium schooling continues to return lower achievement rates for Māori than for European students (Udahemuka, 2016). In 2016, 66% of Māori students left school with NCEA[3] Level 2 or above compared to 84% of European students (Ministry of Education, 2018c).

International measures confirm this picture. The Program for International Student Assessment (PISA) testing across the Organisation for Economic Cooperation and Development (OECD) countries continues to show New Zealand's education system as one that, in terms of education outcomes, achieves high levels of achievement for many students but not for all (Organisation for Economic Co-operation and Development, 2004, 2007, 2010). From the 2012 PISA survey, while New Zealand achievement overall was above the OECD average in reading, mathematics and science, the achievement of Māori students was both below the New Zealand average and the OECD average (May, Cowles, & Lamy, 2013).

PISA describes the New Zealand situation, where some students do well but there is a large gap between high and low achievers, as being one of high quality and low equity. Descriptions of high quality and low equity education systems, driven by deficit-oriented approaches, are familiar to educators across the world (Sleeter, 2011). The learners disproportionately underserved in New Zealand's secondary

[1]Formal removal of a student through a stand-down from school for a period of up to 5 school days.

[2]Where an enrolment of a student aged under 16 is terminated, with a requirement that the student enrols elsewhere.

[3]National Certificate of Educational Achievement (NCEA) is the official secondary school qualification in New Zealand.

schools continue to be Māori and sadly the marginalisation of this group of students is neither a recent phenomenon nor is it confined to education (Bishop, Berryman, & Wearmouth, 2014). Māori students leave school with lower qualifications and fewer life choices for their own futures and for the future well-being of New Zealand as a whole.

12.5.1 The New Zealand Policy Response: Ka Hikitia

Charged with improving Māori student experiences in the education system, the Ministry of Education launched Ka Hikitia - Managing for Success: Māori Education Strategy 2008–2012 (Ministry of Education, 2008). This strategy challenged educators to collaboratively focus on making the difference by ensuring that Māori students, "in their early years and first years of secondary school are present, engaged and achieving, and strong relationships with educators, whānau (family) and iwi (community) are supporting them to excel" (p. 5). The term Ka Hikitia, defined as a means to "'step up', 'lift up', or 'lengthen one's stride'" (Ministry of Education, 2008, p. 10), was positioned as "a call to action" (p. 11) in order to step up "the performance of the education system to ensure Māori [students] are enjoying education success as Māori" (p. 10). Within this strategy was a challenge to schools, education centres, educators, communities and the education system itself to step up so as to more effectively ensure the potential of its Māori learners. In so doing, the Ministry of Education recognised the need for an extensive change in positioning, expectations and practices across the entire education sector, "[i]t is about a shift in thinking and behaviour, a change in attitudes and expectations" (Ministry of Education, 2008, p. 4). The Ka Hikitia policy was refreshed in 2013 (Ministry of Education, 2013a).

Expectations and Responses. In 2013, under a working title of Building on Success, the Ministry of Education (2013b) sought a response that would generate equity by building in-school leader and teacher capability to embed what works for Māori learners within classrooms, leadership, school governance and school-wide practices. Within the contextual landscape of the re-launch of Ka Hikitia and the growing sense of urgency around Māori student achievement, the resulting initiative should also bring together the learnings from over a decade of discrete and varied research and professional learning and development initiatives addressing aspects of school life that impacted on Māori students' school experiences and their achievement (Ministry of Education, 2013b).

The new initiative Kia Eke Panuku (Building on Success) undertook to use an inquiry, evidence-based approach that would be responsive to each individual school and would aim to accelerate and lift the levels of achievement and education success of Māori students as Māori. Kia Eke Panuku focussed on strengthening Māori students' participation and achievement and thus their potential and future as productive citizens at a family and tribal level and at the level of New Zealand and the global community. The model focused on what schools' leadership, teachers, and

in turn Māori students and their families, could do in response to the strengths and/or challenges identified from within the range of contexts and settings in which they each engaged. As such, the aim was that schools would become inextricably connected through Māori students to their homes and communities (Alton-Lee, Robinson, Hohepa, & Lloyd, 2009). Working more effectively with Māori communities would enable schools to benefit from the funds of cultural knowledge (Moll, Amanti, Neff, & Gonzalez, 1992) and expertise that continues to be marginalised and under-utilised by many schools. Critical theories were used alongside kaupapa Māori theories[4] in support of the more effective inclusion of Māori and other marginalised students (Berryman, Nevin, SooHoo, & Ford, 2015).

Shared agenda and moral purpose. Central to Kia Eke Panuku was the agenda of secondary schools giving life to Ka Hikitia and addressing the aspirations of Māori communities by supporting Māori students to pursue their potential. Positioning as partners within the Treaty of Waitangi[5] called on leaders to confront and address the major imbalances of power and privilege that existed as a result of their shared history, either as Māori, or descendants of the colonisers of Aotearoa, or as economic migrants. It also required the acknowledgement and critical reflection of the historical and continuingly destructive impact that loss of land, loss of language, loss of rangatiratanga (self-determination) and loss of mana (ascribed personal prestige and power), continues to have on the wellbeing and success of Māori today. Seeking power-sharing relationships between Treaty partners in the field of educational professional learning and development was therefore essential to avoiding the imposition of yet more inadequate education theories that position Māori epistemologies and world-views at the margins of educational policy and practice rather than at the centre (Smith, 1999).

Leadership was shared with others. School principals were asked to enlist the support of leaders and teachers from across the school to develop a Strategic Change Leadership team. In some cases, this team included senior Māori students and people from their home communities. Some people within these teams described strategic change leaders as:

> A group of people who can lead this new way of being. We're leveraging off multiple voices, groups of people, both inside and outside the school, who've got a vested interest in raising Māori student achievement.

> We have people from different spaces and places. We've got many views coming in and that makes the learning a lot richer and a lot stronger.

[4]Kaupapa Māori research is done by Māori, with Māori and about Māori

[5]The Treaty of Waitangi in 1840, signed by Māori tribal leaders and British Government representatives mandated a partnership relationship and established British governance in return for Māori tribal ownership and protection of their land interests and cultural treasures. However, the sovereignty guaranteed to Māori was increasingly ignored, with dire consequences for Māori cultural, social and economic wellbeing, well into the twentieth century.

> There's a different dynamic when you've got different voices, and voices that may have not necessarily been heard that much before, around the table. It certainly changes the conversations.

The specific factors attended to in Kia Eke Panuku. The Strategic Change Leadership Team Co-Constructed an Entry Point to Align with the Individual School Evidence of Māori Students' Engagement and Achievement As Well as Other Related, Contextual Factors. This Initial Profiling Process Was Based on the Following Five Inter-Related Dimensions that Became the Hallmark of Kia Eke Panuku

1. Leadership
2. Evidence-based inquiry
3. Culturally responsive and relational contexts for learning
4. Educationally powerful connections amongst schools, whānau, hapū, iwi and Māori organisations, and
5. Literacy, numeracy and Māori language across the curriculum.

Support to improve learning. Kia Eke Panuku facilitators worked with Strategic Change Leadership teams to develop individualized action plans. Together they worked across these dimensions towards critical and sustainable change within the school. Their work was about developing and growing the 'skill and the will' within school personnel, to improve outcomes for Māori students. Ultimately the goal was to provide in-school capability along with fit for purpose tools and resources thus removing the need for external support and facilitation (Berryman, Eley, Ford & Egan, Berryman, Eley, Ford, & Egan, 2016).

Planning to do things differently requires learning, unlearning, relearning. In the work with schools, an essential part of planning to do things differently, required unlearning (Wink, 2011) or disrupting much of the embedded discourses and beliefs that formed the status quo (Apple, 2013) about Māori students and their home communities. Unlearning was an essential precursor to relearning and learning more emancipatory discourses of potential and social justice. Leaders and teachers began to question what they were doing and how this might contribute to or resist the current hegemony in their schools and then out into society. This created contexts where new discourses began to exemplify the dynamic interplay between the critical principles of conscientisation, resistance and transformative praxis (Freire, 1972; Smith, 2003)

Evidence of outcomes for Māori learners, alongside evidence of current leadership and/or classroom practices, informed new theorising and practices (*conscientisation*) towards the creation of more socially just learning contexts. Leaders and teachers could then decide what practices were most effective and therefore needed to be sustained; what practices were ineffective and needed to be discontinued; and what practices needed to change in order to become more effective for Māori learners. These practices were understood within Freire's (1972) concept of *resistance*. Leaders and teachers then implemented and reflected on the changes that would lead to accelerating improved outcomes for *Māori learners as Māori* in order to bring about *transformative praxis*.

Leadership discussions. As discussed by members of a school's strategic change leadership team, relevant evidence and ongoing critical reflections informed the continuing clarification of the focus and implementation of the school's strategic action plan:

> It's using evidence as a lens through which people can critically reflect on the influence of their current practice for Māori leaners.

> The important thing about Kia Eke Panuku action plans were that they were iterative documents, and because people were discovering new things as they progressed they should be constantly changing. As a new revelation comes up or a new set of data presents itself, there's a part of this action plan that we didn't know about before, and it's not reflected in it, so we need to change it. These action plans were living documents, quite fluid. They were not being completely changed all the time but they were being modified and tweaked as new realisations or new layers of understanding revealed themselves to the strategic change leadership team.

Through the process of making links between evidence of practice (what was observed) and theory (what we understand about this in terms of effective pedagogy) participants were able to deepen their own understandings about culturally responsive and relational pedagogy. Deepening understandings of the kinds of sustainable change required across the Kia Eke Panuku dimensions are exemplified in the voices of these leaders who are reflecting upon their own agency to bring about the reform in their own school. They first understood that the reform needed to engage across multiple contexts:

> To manage a change like this you actually need both elements; you need the structural element, which is the way the school conceives its goals and its priorities and its leadership and how it does things, but you also need what happens inside the classrooms.

Another leader talked about how this involved learning and helping others to be learners on this journey as well:

> If you think that you've got no room to grow, then it's time to quit. There's always room for improvement and fine-tuning. Shadow coaching will provide a really good opportunity for people to extend themselves.

Others talked about the need to challenge deficit theorising and focus on one's own professional agency to bring about the reform:

> I certainly was very upfront about challenging deficit theorising and repositioning ourselves over time; in our heads and in our hearts. . . to be agentic as professionals.

> We have to ensure that the sense of direction of the organisation is very clear and the leader's job is to define it and articulate it. Repeatedly.

> I would expect if this approach to changing teaching practice has got integrity and we apply it sincerely, it will speak for itself and the teachers who experience it will experience and see changes themselves and will tell other teachers about it and others will be drawn into this mahi (work) . . . as time goes by we expect more and more people to get involved in it.

While this belief in the integrity of this work to draw others in was very humbling, there was increasing evidence to suggest that other school leaders were beginning to

share similar sentiments. However, school leaders are unlikely to do this on their own or without sector imperative and support to do so.

12.5.2 Critical Leadership Leading Transformative Reform

Through the introduction of Kia Eke Panuku in 94 secondary schools, some very pleasing early results emerged. However, the results could not be attributed to either the policy mandate to bring about the change, or to a new set of skills or strategies provided to school personnel. While the good intentions of policy-makers, school leaders and teachers, and a number of discrete interventions aimed at *fixing the Māori student problem*, may be necessary conditions for change, they were not in themselves sufficient.

A desire and a policy mandate for change (the *will* of reform) is an essential prerequisite but is not sufficient on its own. Essential but also insufficient on their own are effective professional learning and development programmes in schools that support leaders and teachers to improve their practices around Māori student experiences (the *skill* of reform). Combined, these will make a difference for many of the students lucky enough to be in those schools and in the classrooms of teachers who are committed to the policy for the reform.

However, for sustained systemic change, a further critical factor was required – widespread ownership of the personal and the public responsibility to use power, privilege, and position within schools to promote social justice and enlightenment for the benefit, not only of individuals and the organisation, but of society as a whole (Quantz, Rogers & Dantley, Quantz, Rogers, & Dantley, 1991; Shields, 2010). Fullan (2003, 2007) refers to leadership for sustainability as public service with a moral purpose. Deep, sustainable change that truly leads to Māori students enjoying and achieving educational success as Māori requires *will* (school leaders deliberate engagement with the policy mandate), *skill* (school leaders and teachers learning from the research and professional development about what works for Māori) underpinned by a relentless *moral imperative* for change (Berryman et al., 2016; Berryman & Eley, 2017).

Beyond the will (the mandate to change) and the skills of school leaders and personnel, the driver for reform rests with leaders who embrace the moral imperative to be the agents for change and who underpin their leadership with a refusal to tolerate a status quo that includes disparity for groups of students within their school, in this case Māori. This ensures that the work is led with a real sense of urgency and with courage to persist with the shared agenda by reframing the situation so that new emancipatory possibilities can be revealed. In many Kia Eke Panuku schools, *critical* leadership such as this had begun to give life to Ka Hikitia. Leaders such as these demonstrated on a daily basis:

- the courage to persist with a vision for Māori students enjoying and achieving educational success as Māori despite what society and history would portray as the 'norm'
- the ability to reframe the situation in order to see new realities and possibilities – that is to learn, unlearn and relearn all aspects of practices and beliefs within a school
- a sense of urgency – the belief that things must change for the students we have in front of us today, tomorrow and into the future.

Given that cohorts of schools entered Kia Eke Panuku one third at a time, with the final cohort not starting till the second year, the time frame was insufficient for deeply embedded reform to take place. All too soon, when funding stopped, the initiative in many schools was halted because new institutions and practices had not had time to become embedded. None-the-less, like the following leader, many leaders had begun to understand their own agency in this work and had begun to get at the heart of the matter:

> You've got to get right down into those deep layers of who you are, of what you're doing and the impact it's having on others. Within strategic change leadership teams and within the face-to-face sessions that we have, there is a space for people to feel comfortable enough for them to start peeling back some of those layers. Don't need that one, don't need that one, oh now that one's a bit challenging, that's going to hurt a bit so I'm going to sit back for a while, just listen. And as they grow in confidence, so they're able to really get to the heart of it. And that's where Māori students are, at the heart of it.

12.6 Discussion

This chapter has described the qualities and characteristics of principals that are regarded as being successful and who lead successful schools, and then provided two examples of how leadership makes a difference: to literacy development and to Māori learning in mainstream schooling. It provides a compelling argument for the positive connection between leadership and learning.

Both the international study and the country studies identified common elements that support successful leadership, even though they might use different terminology and operate within different contexts. The development of vision is talked about in the ISSPP and could be considered the equivalent to developing a shared moral purpose identified in both the Australian and New Zealand case studies; the leadership skills and qualities discussed by the ISSPP are reflected in the support for shared leadership in Australia and critical leadership in New Zealand; and the need to build capacity identified in the ISSPP is reflected by the implementation of the Leadership for Learning Blueprint of PALL and the use of Ka Hikitia in Kia Eke Panuku in New Zealand. In each case intervention is identified as the pathway to focus the will, skill and shared moral purpose of leaders, as the factors that respond to the evidence, circumstances and context in which the school finds itself.

We know that context does matter in terms of educational success (e.g. Teese & Polesel, 2003), and that context is important for how leadership is expressed (Day, Sammons, Hopkins, Harris, Leithwood, Gu & Brown, Day et al., 2010; Gurr, 2014; Maran & Pascual, 2018; Schwarz & Brauckmann, 2015), but the evidence within this chapter suggests that leadership can adaptively respond to context and not be subservient to it. The research on successful principals indicates that they seem to be able to adapt, use and influence context to foster success. In reviewing ISSPP contributions from several countries new to the ISSPP, Drysdale (2011) found that whilst context did impact on what successful leaders did across different country contexts, they were found to be adaptive and reflective, and able to learn from their practice and experience to ensure school success. Early in the story of the ISSPP, Day (2005, p. 581) noted the principals demonstrated the ability to:

> ...not be confined by the contexts in which they work. They do not comply, subvert, or overtly oppose. Rather they actively mediate and moderate within a set of core values and practices which transcend narrowly conceived improvement agendas.

The interplay between context and leadership is complex however. Within the ISSPP there are cases of principals who have changed school contexts, not altered their leadership substantially, and still been successful leaders (e.g. Gurr, 2007). Within the one system, there are examples of principals who have operated in the same policy environment, yet their change interventions/practices are very different with some embracing continuous and often rapid change, and others more circumspect (Drysdale, Goode, & Gurr, 2009, 2011; Goode, 2017; Gurr et al., 2018a, 2018b; Longmuir, 2017). Sometimes the context means that leadership for success becomes a long and difficult path as shown by the New Zealand initiatives to improve the education of Māori students. In Australia, we see principals who have been successful principals in the past, finding themselves struggling to make the impact they would like when they have taken on the challenge of improving a struggling school in challenging circumstances (Gurr, Drysdale, Clarke, & Wildy, 2014; Gurr et al., 2018a, 2018b). Yet, the Australian literacy leadership example showed how common understandings and practices could be used to produce positive outcomes in a range of contexts. Within the ISSPP, the well-articulated general leadership dimensions of setting direction, developing people, developing the school and managing the instructional program (e.g. Leithwood, Day, Sammons, Harris, & Hopkins, 2006), are evident across the cases; they are not sufficient in all contexts, but they do provide guidance on practices that can be applied usefully in all contexts.

12.7 Conclusion

Context can be perceived on many levels, but effective and successful principals are able to make sense of the complex forces that make up their environment. Our research shows that they are not held captive by their context, but work successfully

with the challenges, opportunities and threats that are presented. In some cases, in this chapter, the leaders were able to overcome significant barriers to bring about positive change, despite the context.

What can our examples from Australian and New Zealand research projects tell us about leadership for learning? We can identify five characteristics, and one over-arching factor, that contribute to our understanding of leadership for learning.

First, leadership for learning needs to be broadly defined (MacBeath & Townsend, 2011). It encompasses many aspects which require untangling to understand both the explicit and more subtle dimensions. It can be defined as any activity or action designed to promote growth in student learning. It is the planning, organising, developing, implementing and measuring that goes on the support learning in the school.

Second, leadership for learning is learning for all. It is assumed that leadership for learning focuses on student learning, but this requires learning at all levels: organisational learning, professional learning, individual learning and community learning. Importantly, principals are leaders *in* and *of* learning. This necessitates a sound understanding of pedagogy as evidenced by our examples.

Third, defining what is student learning is also important, be it traditional forms of achievement such as performance in academic areas, or more authentic outcomes (Newman et al., 1996) such as social competence, community values and citizenship. We have examples of both these perspectives with PALL focused on literacy learning and the Maori and ISSPP taking a broader perspective of student success.

The fourth lesson is that leaders have a skill set and personal characteristics that help them to navigate the terrain and influence others to achieve vision for a better future. These skills and qualities are not predictive of success but the leaders in our examples certainly have acquired them and use them to full advantage. However, there are some common factors, such as vision, values, capacity building and strategic interventions, key practices that are characteristic of leaders of learning.

Finally, leadership for learning is not the act of a single person. Leadership is shared or distributed. While there are many leadership styles and approaches that reflect different situations, the principals in our cases empower, engage and enable others to share responsibility in the quest for improved student learning. Leadership for learning is inclusive.

The over-arching factor, one that must be taken into account in order to approach successful leadership as it relates to these five characteristics is the understanding that, as discussed previously, context matters. How much it matters is complex. It takes in factors of geography, demography, economics, culture, community and language. The language we use to work within our context is critical for, as Williams (1972) shows us, even within the same country, and even when we might speak the same national language, the nuances of language are important. Some countries don't have a word for leadership and other countries use the same word for both leadership and administration. Successful principals learn the language of the community they work in and reach out to bridge cultural and other differences that might exist within the school and the school system, which always reflects the dominant community, with the local community, which in many cases might have a very

different view of the world. To do so involves respect, invitation and support. This is particularly the case when working with diverse and often disadvantaged communities and students. This has been highlighted in all three examples in this chapter as an unending commitment to developing a strong, shared, underlying moral purpose for education, one that considers a positive educational outcome for every single student, both in single schools and across education systems, as being the sole measure of success.

Over time we have moved from using the term school effectiveness, where the principal was seen as the key agent for successful student achievement, through classroom effectiveness, which gave priority to the work of teachers, and we are now using the term educational effectiveness, which reflects the complex interactions that lead to the vision of success for all students. As Creemers and Kyriakides (2012) argue, the interactions between student, teacher, school leaders and families, on the one hand and classroom, school, education system and community, on the other, are dynamic and fluid. A dynamic approach to school improvement is required if we are to achieve both quality and equity in education. All need to play a role if equity is to be achieved.

Whether we are looking for curriculum or system reform, contexts for effecting positive change for those students who we continue to fail or marginalise throughout our education systems, remain our biggest challenge. This requires leadership that is also prepared to engage with challenging and changing the deeply entrenched, unconscious and conscious bias in schools and society at large. These emerging discourses continue to reinforce a particular deficit view of groups and an ongoing status quo of underachievement. Contexts such as these require leadership to be both critical and political. Furthermore, until leaders have a determined and urgent focus on these students, the depth of our moral purpose may yet to be fully realised.

References

Apple, M. (2013). *Knowledge, power, and education*. New York, NY: Routledge.

Alton-Lee, A., Robinson, V., Hohepa, M., & Lloyd, C. (2009). Creating educationally powerful connections with family, whānau, and communities. In V. Robinson, M. Hohepa, & C. Lloyd (Eds.), *School leadership and student outcomes: Identifying what works and why* (pp. 142–170). Wellington, New Zealand: Ministry of Education.

Auditor-General. (2012). *Education for Māori: Context for our proposed audit work until 2017*. Wellington, New Zealand: Office of the Auditor-General.

Auditor-General. (2013). *Education for Māori: Implementing Ka Hikitia – Managing for success*. Wellington, New Zealand: Parliamentary paper.

Australian Primary Principals Association (APPA). (2013). *Principals as literacy leaders (project extension no 1) final report*. Kingston, ACT, Australia: APPA.

Barth, R. S. (1999). *The teacher leader*. Providence, RI: Rhode Island Foundation.

Bassett, G. W., Crane, A. R., & Walker, W. G. (1967). *Headmasters for better schools (2nd ed.)*. St Lucia, Australia: University of Queensland Press.

Beare, H., Caldwell, B. J., & Millikan, R. H. (1989). *Creating an excellent school*. London, UK: Routledge.

Belchetz, D., & Leithwood, K. (2007). Does context matter and if so, how? In K. Leithwood & C. Day (Eds.), *Successful school leadership in times of change* (pp. 117–138). Toronto, Canada: Springer.

Berryman, M., & Eley, E. (2017). Succeeding *as Māori*: Māori students' views on our *stepping up* to the Ka Hikitia challenge. *New Zealand Journal of Education Studies,* 1–15. https://doi.org/10.1007/s40841-017-0076-1

Berryman, M., Eley, E., Ford, T., & Egan, M. (2016). Going beyond the personal will and professional skills to give life to Ka Hikitia. *Journal of Educational Leadership, Policy and Practice, 30*(2), 56–68.

Berryman, M., Nevin, A., SooHoo, S., & Ford, T. (2015). A culturally responsive framework for social justice. In K. Esposito & A. Normore (Eds.), *Inclusive practices for special populations in urban settings: The moral imperative for social justice leadership* (pp. 143–164). Charlotte, NC: Information Age Publishers (IAP).

Bishop, R., Berryman, M., & Wearmouth, J. (2014). *Te Kotahitanga: Towards effective education reform for indigenous and other minoritised students*. Wellington, New Zealand: NZCER Press.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education: Dynamic approaches to school improvement*. London, UK: Routledge.

Day, C. (2005). Sustaining success in challenging contexts: Leadership in English schools. In K. Leithwood & C. Day (Eds.), *Successful school leadership in times of change* (pp. 59–70). Dordrecht, The Netherlands: Springer-Kluwer.

Day, C., Harris, A., Hadfield, M., Tolley, H., & Beresford, J. (2000). *Leading schools in times of change*. Buckingham, UK: Open University Press.

Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., & Brown, E. (2010). *10 strong claims about successful school leadership*. Nottingham, UK: National College for Leadership of Schools and Children's Services.

Dempster, N., Konza, D., Robson, G., Gaffney, M., Lock, G., & McKennariey, K. (2012). *Principals as literacy leaders: Confident, credible and connected*. Kingston, Autralia: Australian Primary Principals Association.

Dempster, N., Johnson, G., & Stevens, E. (2014). *The Principals as Literacy Leaders (PALL) project: (PALL: Tasmania): Final report, October 2014.* Unpublished report, Brisbane, Queensland: Griffith Institute for Educational Research, Griffith University.

Dempster, N., Townsend, T., Johnson, G., Bayetto, A., Lovett, S., & Stevens, E. (2017). *Leadership and literacy: Principals, partnerships and pathways to improvement*. Cham, Switzerland: Springer, 209pp. (available in 113 libraries worldwide as of august 2017 – worldcat.org)

Dinham, S. (2007). The secondary head of department and the achievement of exceptional student outcomes. *Journal of Educational Administration, 45*(1), 62–79.

Drysdale, L. (2011). Special Issue. *Leadership and Policy in Schools, 10*(4).

Drysdale, L., Goode, H., & Gurr, D. (2009). An Australian model of successful school leadership: Moving from success to sustainability. *Journal of Educational Administration, 47*(6), 697–708.

Drysdale, L., Goode, H., & Gurr, D. (2011). Sustaining school and leadership success in two Australian schools. In L. Moos, O. Johansson, & C. Day (Eds.), *How school principals sustain success over time: International perspectives* (pp. 25–38). Dordrecht, The Netherlands: Springer-Kluwer.

Drysdale, L., & Gurr, D. (2011). The theory and practice of successful school leadership in Australia. *School Leadership and Management, 31*(4), 355–368.

Duignan, P., & Gurr, D. (Eds.). (2007). *Leading Australia's schools*. Sydney, Australia: ACEL and DEST.

Duignan, P., Marshall, A. R., Harrold, R. I., Phillipps, D. M., Thomas, E. B., & Lane, T. J. (1985). *The Australian school principal: A summary report*. Canberra, Australia: Commonwealth Schools Commission.

Edmonds, R. (1978). *A discussion of the literature and issues related to effective schooling*. A paper presented to National Conference on Urban Education, CEMREL, St Louis, USA.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15–27.

Enomoto, E. (1997). Schools as nested communities: Sergiovanni's metaphor extended. *Urban Education, 32*(4), 512–531.
Freire, P. (1972). *Pedagogy of the oppressed*. New York, NY: Continuum.
Fullan, M. (2003). *The moral imperative of school leadership*. Thousand Oaks, CA: Corwin Press.
Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York, NY: Teachers College Press.
Goode, H. (2017). *A study of successful principal leadership: Moving from success to sustainability*. Doctor of Philosophy Thesis, The University of Melbourne.
Gronn, P. (2000). Distributed properties: A new architecture for leadership. *Educational Management and Administration, 28*, 317–338.
Gronn, P. (2002). Distributed leadership as a unit of analysis. *Leadership Quarterly, 13*, 423–451.
Gurr, D. (2007). We can be the best. In P. Duignan & D. Gurr (Eds.), *Leading Australia's schools* (pp. 124–131). Sydney, Australia: ACEL and DEST.
Gurr, D. (2009). Successful school leadership in Australia. In N. Cranston & L. Erlich (Eds.), *Australian educational leadership today: Issues and trends* (pp. 369–394). Brisbane, Australia: Australian Academic Press.
Gurr, D. (2014). Successful school leadership across contexts and cultures. *Leading and Managing, 20*(2), 75–88.
Gurr, D. (2015). A model of successful school leadership from the international successful school Principalship project. *Societies, 5*(1), 136–150.
Gurr, D., & Drysdale, L. (2007). Models of successful school leadership: Victorian case studies. In K. Leithwood & C. Day (Eds.), *Successful school leadership in times of change* (pp. 39–58). Dordrecht, Netherlands: Springer-Kluwer.
Gurr, D., & Drysdale, L. (2016). Successful and effective school leadership: Insights from Australasia. In P. Pashiardis & O. Johansson (Eds.), *Successful school leadership: International perspectives* (pp. 139–154). London, UK: Bloomsbury.
Gurr, D., & Drysdale, L. (2013). Middle-level school leaders: Potential, constraints and implications for leadership preparation. *Journal of Educational Administration, 51*(1), 55–71.
Gurr, D., Drysdale, L., Clarke, S., & Wildy, H. (2014). High needs schools in Australia. *Management in Education, 28*(3), 86–90.
Gurr, D., Drysdale, L., Di Natale, E., Ford, P., Hardy, R., & Swann, R. (2003). Successful school leadership in Victoria: Three case studies. *Leading and Managing, 9*(1), 18–37.
Gurr, D., Drysdale, L., Longmuir, F., & McCrohan, K. (2018a – In press). Successful school leadership that is culturally sensitive but not context constrained, in E. Murakami, D. Gurr, & R. Notman (Eds), *Leadership, culture and school success in high-need schools*. Washington, DC: Information Age Publishing, pages to be confirmed on publication.
Gurr, D., Drysdale, L., Longmuir, F., & McCrohan, K. (2018b – In press). Leading the improvement of schools in challenging circumstances, *International Studies in Educational Administration, 46*(1), pages to be confirmed on publication
Gurr, D., Drysdale, L., & Mulford, B. (2005). Successful principal leadership: Australian case studies. *Journal of Educational Administration, 43*(6), 539–551.
Gurr, D., Drysdale, L., & Mulford, B. (2006). Models of successful principal leadership. *School Leadership and Management, 26*(4), 371–395.
Hallinger, P. (2018). Bringing context out of the shadow. *Educational Management Administration & Leadership, 46*(1), 5–24.
Hallinger, P., & Heck, R. H. (2010). Leadership for learning: Does collaborative leadership make a difference in school improvement? *Educational Management Administration & Leadership, 38*, 654–678.
Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal, 86*(2), 217–248.
Hopkins, D. (2010). *Realising the Potential of System Leadership Studies in Educational Leadership, 10*, 211–224.

Hopkins, D., Harris, A., & Jackson, D. (2010). Understanding the School's capacity for development: Growth states and strategies. *School Leadership & Management, 17*(3), 401–412. https://doi.org/10.1080/13632439769944

Hord, S. (1997). *Professional learning communities: Communities of continuous inquiry and improvement*. Austin, TX: Southwest Educational Development Laboratory.

Johnson, G., Dempster, N., McKenzie, L., Klieve, H., Fluckiger, B., Lovett, S., … Webster, A. (2014). *Principals as literacy leaders with indigenous communities: Leadership for learning to read – 'Both ways'*. Canberra, Australia: The Australian Primary Principals Association.

Lambert, L. (2002). A framework for shared leadership. *Educational Leadership, 59*(8), 37–40.

Leithwood, K., & Janzti, D. (2000). The effects of transformational leadership on organisational conditions and student engagement with school. *Journal of Educational Administration, 38*(2), 112–129.

Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. Nottingham, UK: National College of School Leadership.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91*, 509–536.

Longmuir, F. (2017). *Principal leadership in high-advantage, improving Victorian secondary schools*. Doctor of Philosophy Thesis, The University of Melbourne.

MacBeath, J. (2010). Leadership for learning: Concepts, principles and practices. *Leadership for Learning, the Cambridge Network*. Retrieved from http://oer.educ.cam.ac.uk/w/images/9/98/LfL-Concepts_Principles_Practices_cc_1.pdf

MacBeath, J., Frost, D., & Swaffield, S. (2008). Editorial. *School Leadership & Management, 28*(4), 301–306.

MacBeath, J., & Dempster, N. (2009). *Connecting leadership and learning: Principles for practice*. London, UK: Routledge.

MacBeath, J., & Townsend, T. (2011). Thinking and acting both locally and globally: What do we know now and how do we continue to improve? In T. Townsend & J. MacBeath (Eds.), *The international handbook of leadership for learning* (pp. 1241–1259). New York, NY: Springer.

Maran, J., & Pascual, J. (2018). Comparative study of school principals' leadership practices: Lessons for Chile from a cross-country analysis. *Educational Management Administration and Leadership, 46*(2), 279–300.

Marzano, R. J., Waters, J. T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

May, S., Cowles, S., Lamy, M., & Research Division, Ministry of Education. (2013). *PISA 2012: New Zealand summary report*. Wellington, New Zealand: Ministry of Education.

Ministry of Education. (2008). *Ka Hikitia-managing for success: The Māori education strategy 2008–2010*. Wellington, New Zealand: Ministry of Education.

Ministry of Education. (2013a). *Ka Hikitia – Accelerating success 2013–2017*. Wellington, New Zealand: Ministry of Education.

Ministry of Education. (2013b). *Registration of interest 397–5609, professional development to generate equity and accelerate the achievement of Māori learners in secondary education (working title: Building on success)*. Wellington, New Zealand: Ministry of Education.

Ministry of Education. (2018a). *Education Counts*. Retrieved April 14, 2018, from Highest Attainment Numbers: http://www.educationcounts.govt.nz/statistics/indicators/main/student-engagement-participation/Stand-downs-suspensions-exclusions-expulsions

Ministry of Education. (2018b). *Education Counts*. Retrieved April 14, 2018, from Highest Attainment Numbers: http://www.educationcounts.govt.nz/statistics/indicators/main/student-engagement-participation/1955

Ministry of Education. (2018c). *School Leaver Attainment*. Retrieved April 16, 2018, from Education Counts: http://www.educationcounts.govt.nz/statistics/schooling/senior-student-attainment/school-leavers2/ncea-level-2-or-above-numbers

Mitra, D. (2007). Student voice in school reform: From listening to leadership. In D. Thiessen & A. Cook-Sather (Eds.), *International handbook of student experience in elementary and secondary school* (pp. 727–744). New York, NY: Springer.

Moll, L., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice, 31*(2), 132–141.

Møller, J. (2002): Democratic Leadership in an Age of Managerial Accountability. *Improving Schools,* 5(1).

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Somerset, UK: Open Books.

Mulford, B., & Johns, S. (2004). Successful school principalship. *Leading and Managing, 10*(1), 45–76.

Murphy, J., Elliott, S., Goldring, E., & Porter, A. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership & Management, 27*(2), 179–201.

National Institute of Child Health and Human Development (NICHD). (2000). *National reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* https://www1.nichd.nih.gov/publications/pubs/nrp/Pages/smallbook.aspx

Newman, F. M., Marks, H. M., & Camoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*, 280–312. https://doi.org/10.1086/444136

Notman, R. (Ed.). (2011a). *Successful educational leadership in New Zealand: Case studies of schools and an early childhood centre*. Wellington, New Zealand: NZCER Press.

Notman, R. (2011b). Building leadership success in a New Zealand context. In R. Notman (Ed.), *Successful educational leadership in New Zealand: Case studies of schools and an early childhood centre* (pp. 135–152). Wellington, New Zealand: NZCER Press.

Notman, R. (2012). Intrapersonal factors in New Zealand school leadership success. *International Journal of Educational Management, 26*(5), 470–479.

Notman, R. (2014). A values-led principalship: The person within the professional. In C. Day & D. Gurr (Eds.), *Leading schools successfully: Stories from the field* (pp. 117–128). London, UK: Routledge.

Notman, R., & Henry, A. (2009). The human face of principalship: A synthesis of case study findings. *Journal of Educational Leadership. Policy and Practice, 24*(1), 37–52.

Organisation for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: PISA results for 2003.* Retrieved January 13, 2016, from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/34002216.pdf

Organisation for Economic Co-operation and Development. (2007). *PISA 2006 results: Science competencies for tomorrow's world.* Retrieved January 13, 2016, from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/pisa2006results.htm

Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: What students know and can do in reading, mathematics and science.* Retrieved January 13, 2016, from http://www.oecd.org/pisa/pisaproducts/pisa2009keyfindings.htm

Quantz, R., Rogers, J., & Dantley, M. (1991). Rethinking transformative leadership: Toward democratic reform of schools. *Journal of Education, 173*(3), 96–118.

Reynolds, D. (1982). The search for school effectiveness. *School Organisation, 2*(3), 215–237.

Robinson, V., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: an analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 564–588.

Robinson, R., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why. Best evidence synthesis iteration [BES]*. Wellington, New Zealand: New Zealand Ministry of Education. Retrieved from http://www.educationcounts.govt.nz/publications/series/2515/60169/60170

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. Retrieved from http://www.standards.dfes.gov.uk/rosereview

Routman, R. (2014). *Read, write, lead: Breakthrough strategies for schoolwide literacy success*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Rowe, K. (2005). *Teaching reading: Report and recommendations of the National Inquiry into the teaching of literacy*. Canberra, Australia: Department of Education, Science and Technology.

Rutter, M., Maughan, B., Mortimore, P., & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and effects on children*. Boston, MA: Harvard University Press.

Schwarz, A. & Brauckmann, S. (2015). Between facts and perceptions: The area close to school as a context factor in school leadership, *Schumpter Discussion Papers*, *2015*(3), Schumpter School of Business and Economics, University of Wuppertal, Germany.

Seashore-Louis, K., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). Learning from leadership. *Investigating the links to improved student learning*. Final report of research to Wallace Foundation. Minneapolis, MN: University of Minnesota.

Sergiovanni, T. L. (1987). The change process: Management and leadership for school improvement. *The principalship: A reflective practice perspective* (pp. 276–294).

Sergiovanni, T. J. (1991). *The principalship – A reflective practice perspective* (3rd ed.). Boston, MA: Allyn and Bacon.

Sergiovanni, T. (1992). *Moral leadership: Getting to the heart of school improvement*. New York, NY: Jossey Bass.

Shields, C. (2010). Transformative leadership: Working for equity in diverse contexts. *Educational Administration Quarterly, 46*(4), 558–589.

Silins, H. (1994). The relationship between transformational and transactional leadership and school improvement outcomes. *School Effectiveness and School Improvement, 5*(3), p272–p298.

Sleeter, C. (Ed.). (2011). *Professional development for culturally responsive and relationship-based pedagogy* (1st ed.). New York, NY: Peter Lang Publishing.

Smith, G. H. (2003). *Kaupapa Māori theory: Transforming indigenous transformation of education and schooling*. Kaupapa Māori Symposium, New Zealand Association for Research in Education/Australian Association for Research in Education Conference, Auckland.

Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. London, UK: Zed Books.

Southworth, G. (2005). School leadership and system leadership. *The Educational Forum, 69*(2), 212–220.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher, 30*, 23–28.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies, 36*, 3–34.

Starratt, R. (2001). Democratic leadership theory in late modernity: An oxymoron or ironic possibility? In P. Begley & O. Johansson (Eds.), *The ethical dimensions of school leadership* (pp. 12–31). Boston, MA: Kluwer.

Starratt, R. J. (2004). *Ethical leadership*. San Francisco, CA: Jossey-Bass.

Teese, R., & Polesel, J. (2003). *Undemocratic schooling: Equity and quality in mass secondary education in Australia*. Carlton, Australia: Melbourne University Press.

Townsend, T. (2016). Leading schools in the 21st century: Careful driving in the fast Lane. In G. Johnson & N. Dempster (Eds.), *Leadership in diverse learning contexts* (pp. 411–430). Cham, Switzerland: Springer.

Townsend, T. (2017). *Sustainability and impact*: *A conversation with schools where the Principals as Literacy Leaders (PALL) program and the Reading BIG 6 have been used over three or more years*. Victorian Principals Association, North Melbourne

Townsend, T. (2018). *Milestone report on 2017 data collection from the Tasmanian principals as literacy leaders (PALL) research* (1st year of a three-year study) (unpublished report)

Townsend, T., Dempster, N., Johnson, G., Bayetto, A. & Stevens, E. (2015a). *Leadership with a purpose: A report on five case studies of Principals as Literacy Leaders (PALL) schools*

(Tasmania). Unpublished report. Brisbane, Queensland: Griffith Institute for Educational Research, Griffith University.

Townsend, T., Wilkinson, J., & Stevens, E. (2015b). *Leadership with a purpose: A report on four case studies of Principals as Literacy Leaders (PALL) schools (Victoria).* Unpublished report. Brisbane, Queensland: Griffith Institute for Educational Research, Griffith University.

Townsend, T., & MacBeath, J. (2011). *The international handbook on leadership for learning*. Dordrecht, The Netherlands: Springer.

Udahemuka, M. (2016). *Signal loss: What we know about school performance*. Wellington, New Zealand: The New Zealand Initiative.

Waters, J. T., & Marzano, R. J. (2007). School District leadership that works: The effect of superintendent leadership on student achievement. *ERS Spectrum, 25*(2), 1–12.

Weber, G. (1971). *Inner City children can be taught to read: Four successful schools*. Washington, DC: Council for Basic Education.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council: Dallas, TX.

Wenner, J. A., & Campbell, T. (2017). The theoretical and empirical basis of teacher leadership a review of the literature. *Review of Educational Research, 87*(1), 134–171.

Williams, Robert L (September 1972). *The BITCH-100: A Culture-Specific Test*. Paper presented at the American Psychological Association Annual Convention. Honolulu, Hawaii.

Wink, J. (2011). *Critical pedagogy: Notes from the real world* (4th ed.). Upper Saddle River, NJ: Pearson.

Chapter 13
A National Evaluation of Kindergarten Outcomes: Findings from Uruguay

Alma Y. Lopez and J. Douglas Willms

13.1 Introduction

One of the targets for the 2030 Sustainable Development Goals (SDGs) set out by UNESCO is "By 2030, ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education" (United Nations [UN], 2015, p.17). However, only about one in five children in low-income countries attend preschool (World Bank, 2018a) and only a small number of low- and middle-income countries have achieved universal provision at the kindergarten level. A number of studies in high-income countries affirm that attending pre-kindergarten and kindergarten programs has a positive effect on children's skill development and reduces the risk of school failure (Ehrlich, Gwynne, & Allensworth, 2018; Hall et al., 2013; Hatfield, Burchinal, Pianta, & Sideris, 2016; Holod, Ogut, de los Reyes, Quick, & Manship, 2018; Skibbe, Connor, Morrison, & Jewkes, 2011). Findings from the Early Childhood Longitudinal Study (ECLS), which followed a large nationally representative cohort of United States children from an early age, found significant effects of children's skills on later reading achievement when they enter kindergarten (Chatterji, 2006). However, there have been only a few comparable studies in low- and middle-income countries. Also, we know relatively little about the range of pre-literacy skills children have when they begin kindergarten or whether their increase these skills during the kindergarten year.

Uruguay was one of the first countries in Latin America to have universal provision of early childhood programs for children aged 3–5 as part of its public-school offering. In 2017, the Ministry conducted a national study of kindergarten outcomes using the *Early Years Evaluation* (*EYE*; The Learning Bar, 2011a, 2011b)

A. Y. Lopez (✉) · J. D. Willms
The Learning Bar Inc, Fredericton, Canada
e-mail: almayadhira@yahoo.com.mx; doug.willms@thelearningbar.com

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_13

or *Evaluación Infantil Temprana (EIT)* in Spanish, which assesses the developmental skills of children as they prepare for and make the transition to formal schooling. In Uruguay, children aged 3, 4 and 5 are referred to as 'kindergarten' students. The target population for the 2017 data collection was all children enrolled in kindergarten at ages 4 and 5. The analyses in this paper pertain only to kindergarten students in the age 4 cohort, which we refer to as 'Kinder-4' students.

This is the first population-based national study of kindergarten outcomes in Latin America, and as such, it provides information on the variation in skill development when children enter kindergarten and the effects of kindergarten attendance for vulnerable children. Uruguay is classified as an upper middle-income country with a gross national income (GNI) of $21,870 in 2017 (World Bank, 2018b). However, levels of socioeconomic status vary considerably across its 19 geographic departments. For this study, data collected by the Ministry on the socioeconomic status (SES) of the schools were merged with the *EIT* data, enabling us to examine the extent to which classroom effects vary with SES. The observed effects in the lowest SES schools provide some indication of whether attendance in a kindergarten classroom can have positive effects in middle-income countries.

13.2 Literature Review

The assessment framework for Uruguay is in many respects aligned with the Educational Prosperity framework described by Willms (2018a). The framework, which is based on a life-course approach, includes six stages of development: prenatal, early development (age 0–2), pre-primary (age 3–5), early primary (age 6–9), late primary and lower secondary (age 10–15), and upper secondary (age 16–18). Willms (2018a) stresses the importance of two critical transitions: from pre-primary to early primary (school entry), which in Uruguay and many other countries occurs at age 5 or 6, and the transition from early primary to late primary, which he refers to as the transition from learning-to-read to reading-to-learn. His report uses the 2017 Uruguayan data to show the variation in cognitive and language skills at ages 4 and 5. The results show that the variation is substantial for both domains.

This study is situated in the middle of the pre-primary period. It is a crucial period for the development of skills that enable a child to be successful at school. Two types of studies have been conducted that pertain to the effects of attending pre-schools or kindergarten programs. One considers the specific skills that are related to later success in reading and mathematics and whether early childhood education and care (ECEC) and kindergarten programs are effective in developing these skills. The other focuses on the long-term effects of attending a pre-school or kindergarten program.

13.2.1 The Skills That Matter

The successful acquisition of reading skills during the early primary school period depends on two components: decoding and linguistic comprehension (Gough & Tunmer, 1986). This is referred to as the "Simple View of Reading". The majority of children who fail to become successful readers fail to learn how to decode words; that is, the ability to recognize familiar and unfamiliar words (Storch & Whitehurst, 2002; Verhoeven, van Leeuwe, & Vermeer, 2011). Linguistic comprehension – the ability to understand and interpret spoken and written language when they are parts of sentences or other discourse – is reinforced when children develop strong language skills and begin to understand what they are reading (Perfetti, Landi, & Oakhill, 2005; Vellutino & Scanlon, 1987).

The seminal work of Scarborough (1989), which summarized the results of 61 studies, found that children's letter identification and phonological awareness skills when they started early primary school were predictive of their ability to decode words when they were in the first or second grade, while vocabulary, concepts about print, and sentence recall were predictive of linguistic comprehension skills. Scarborough (2001) identified two strands of skills that reinforce each other as children learn to read: word recognition, which includes phonological awareness, the ability to decode words, and sight recognition of familiar words; and language comprehension, which includes vocabulary, verbal reasoning, language structures, background knowledge, and literacy knowledge.

Several studies over the past decade support Scarborough's findings and are consistent with the Simple View of Reading. A comprehensive review conducted by the National Early Literacy Panel (2008) identified alphabet knowledge, concepts about print, phonological processing skills, and oral language skills as strong predictors of subsequent reading skills. An important aspect of this review is that the findings indicated that these skills were predictive of later reading achievement after controlling for IQ and SES. This suggests that ECEC and kindergarten programs that explicitly teach key pre-literacy skills can enable children with differing ability and family backgrounds to become successful readers (Schatschneider, Petscher, & Williams, 2008). A meta-analysis conducted by La Paro and Pianta (2000), a summary of six longitudinal studies by Duncan et al. (2007), and a comprehensive review by Linan-Thompson (2014) provide further support for the development of pre-literacy skills during the pre-primary period.

An important caveat, discussed by Willms (2018a) and relevant to this study, is that the rate at which children become proficient in decoding words depends on the orthographic depth of the language, which refers to the relationship between letters and speech sounds (phonemes). Spanish has a very shallow orthography: the connection between letters and their speech sounds is consistent, and therefore the acquisitions of decoding skills is considerably easier than for English or French, which have a deep orthography. Two studies conducted in Latin America, based on the simple view of reading, have concluded that linguistic comprehension is more

important than decoding as children are learning to read (Polo, Araujo, & Salceda, 2017; Ripoll, Aguado, & Castilla-Earls, 2014).

A number of studies have stressed the importance of the quality of these programs (Lehrl & Smidt, 2018). An early review by Ramey and Ramey (1998) set out six principles that remain important markers of program quality:

1. timing – programs that begin early, preferably soon after birth;
2. intensity – programs that frequently engage experienced teachers;
3. directedness – programs that provide learning experiences that have direct effects on later outcomes;
4. comprehensiveness – programs that take a broad, multi-pronged approach;
5. matching – programs that take into account children's individual needs;
6. maintenance – programs that identify opportunities to support early gains after children enter school.

A study of 76 children by Skibbe et al. (2011) examined whether the effects of preschool on children's self-regulation, decoding, and letter knowledge skills were associated with the duration of pre-school. They found that children who had received 2 years of pre-school had higher scores in decoding and letter knowledge than those who had received only 1 year of preschool. This was not the case for vocabulary or self-regulation, however. A limitation of their study is that they could not control for the potential effects of maturity. A study conducted in Bolivia examined the effects of a preschool program designed for disadvantaged children, with a sample of approximately 4000 children (Behrman, Cheng, & Todd, 2004). An important aspect of their study is that they matched the children attending the preschool with a control group of children in the area who were not in the program. They found significant effects on children's cognitive and psychosocial outcomes for children that had attended the program for at least 7 months. These studies emphasize the importance of controlling for the maturation of children in the absence of preschool attendance and the duration of the program.

13.2.2 The Enduring Effects of Attending Pre-schools Programs and Kindergarten

A number of studies have examined the impact of attending ECEC programs or kindergarten during the pre-primary period. These studies have used different methodologies to assess ECEC and kindergarten effects on children's skills during the early grades of primary education, and in some cases their long-term effects. Reviews of this literature have generally supported two conclusions: attending high-quality ECEC and kindergarten programs leads to better schooling outcomes (Yoshikawa et al., 2013) and the effects tend to be stronger for children from disadvantaged backgrounds and those with lower ability (Barnett, 2011).

A meta-analysis of three types of programs found that their effects vary: model programs designed by researchers have effects on cognitive development of 0.57

standard deviations; state and local public-school programs have effects of 0.32 standard deviations; and Head Start programs have effects of 0.17 standard deviations (Kay & Pennucci, 2014). These results are consistent with Hattie's (2009) synthesis of 16 meta-analyses of early interventions; he reported an average effect size for pre-school programs of 0.50.

Longitudinal studies that have followed large cohorts of children also provide evidence that attending ECEC and kindergarten programs have positive, enduring effects. Findings from the Early Childhood Longitudinal Study (ECLS), conducted by the National Center of Education Statistics (NCES) with a nationally representative cohort of U.S. children, indicated significant positive effects on children's reading achievement in the first grade of elementary school (McCoach, O'Connell, Reis, & Levitt, 2006), and on children's growth in mathematics through to the third grade (DiPerna, Lei, & Reid, 2007). Similarly, analysis of data from the Effective Pre-school, Primary, and Secondary Education Project (EPPSE 3–16+), which followed 2800 children in the U.K. that attended kindergarten and 300 that did not attend kindergarten, found positive effects for children's outcomes in reading and mathematics at ages 7, 11, and 16 (Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2008; Taggart, Sylva, Melhuish, Sammons, & Siraj, 2015). Similar findings pertaining to the development of numeracy skills were found in a study involving 97 German preschools (Anders, Grosse, Rossbach, Ebert, & Weinert, 2013).

Studies based on data collected in Latin American countries have also confirmed the positive effects of attending ECEC and kindergarten programs. An Argentinian study of the effects of attending formal preschools, called *Jardines*, had positive effects on children's receptive vocabulary, with larger effects associated with the number of years children attended the program (Cueto et al., 2016). Berlinski, Galiani, and Gertler (2009) found for a large sample of children in Argentina that 1 year of pre-primary school attendance was associated with an increase of third grade test scores of 0.23 standard deviations. Two longitudinal studies with large samples and a longer duration have also reported positive, enduring effects of attending kindergarten. A study of 17,430 children in Uruguay found strong positive effects of kindergarten attendance on children's achievement in grade 1. These effects were weaker, but still statistically significant, in grade 6 (Aguilar & Tansini, 2012). A study in Chile examined the effects of ECEC attendance on academic achievement in grade 4, using data from a national assessment administered in 2008. The analysis compared the results for a sample of 31,947 children who attended ECEC programs provided by an educational service that covered ages 2–5 years with a sample of 54,571 children who did not attend an ECEC program. Children who attended an ECEC program had better performance in math, reading, and social science than those who did not attend, with children from low-middle SES families benefiting the most (Cortázar, 2015).

13.3 Research Questions

This study considers two broad sets of questions. The first set is about the variation in children's language and cognitive skills when they enter Kinder-4:

1. What is the variation in language and cognitive skills among children within classrooms?
2. Is this variation related to the gender and age of the children?
3. To what extent do average levels of children's skills vary among classrooms?
4. Is the variation among classrooms in average skill levels related to school mean SES?

Analysis of the first question will give an indication of the range of learning needs that a typical kindergarten teacher needs to contemplate; it is relevant to the curriculum, the organisation of the classroom, and the level of support required to meet all children's developmental learning needs. The second question not only considers the relationship of school-entry skills with gender, but also the relationship with age. Our estimate of the 'age effect' is an estimate of the expected growth in skills if children did *not* attend Kinder-4. We refer to this as the 'maturity effect' and later we use this to assess the effect of attending Kinder-4. The last two questions are relevant to the degree to which children are segregated into different classrooms based on their initial skill levels. In all low- and middle-income countries, skill levels at the late primary and lower secondary levels vary between rural and urban schools and between poor and rich areas of the large cities (Murillo, 2007; Willms & Somers, 2001). However, we know relatively little about the extent of between-classroom variation in skill levels when children first enter school. These can be seen as baseline estimates of *vertical* segregation for a schooling system (Willms, 2010, 2018a).

The second set of questions will discern with whether attendance in kindergarten makes a difference. Several studies have asked whether certain 'readiness skills' or attending kindergarten are predictive of later school success. We discuss the main findings of these two types of studies in the literature review. Generally, at least in high-income countries, a number of cognitive and language skills measured in kindergarten or grade 1 are predictive of reading and math skills measured at the end of the early primary period. Also, the effects of kindergarten attendance tend to be greater for low-SES and low-ability students. The less studied piece in the literature, however, is whether children improve their levels of predictive readiness skills over the course of a kindergarten year. Accordingly, this study considers five questions:

1. For children who are 'vulnerable', to what extent do their pre-post growth scores in cognitive and language skills vary within classrooms?
2. Are their rates of growth related to their gender or their age when they started the school year?
3. To what extent do rates of growth of vulnerable children vary among classrooms?
4. Is this variation related to the mean SES of the school?
5. Does the rate of growth exceed the maturity effect for the majority of classrooms?

The first two questions address the issue of whether kindergarten attendance has an effect, gauged by whether children improve their levels of predictive readiness skills over the course of the kindergarten year. The last three questions are similar to those used in traditional 'school effects' analyses which contemplate whether some classroom teachers are more successful than others in improving children's skills. We do not couch the question in those terms, however, as we do not have adequate data for controlling for children's family background or the potential effects of various community factors. Ideally, from a research perspective, one would like to conduct a 'true' experiment, with children randomly assigned to treatment and control conditions. Given that this is impossible with a national study, and depending on the study design, it could be considered unethical, we address this problem by estimating the 'maturity effect', which is the expected growth in skill development if children did not attend the kindergarten program. This enables us to estimate the extent to which gains in children's pre-literacy skills exceed the maturity effect. These results are important as they can indicate the need for further studies that consider in detail teachers' approaches to classroom instruction. They can also serve as a baseline for assessing the effects of a kindergarten intervention. In Canada, for example, a study based on the *EYE* is in progress to evaluate the effects on skill development of providing teachers with play-based learning activities aimed at strengthening children's cognitive and language skills.

13.4 Method

13.4.1 Data Sources and Measures

Evaluación Infantil Temprana (EIT) The *EIT* is an assessment tool designed to identify children at ages 4 and 5 that are likely to struggle learning to read during the early primary school period. The *EIT* evaluates aspects of early child development in five developmental domains that are closely related to emerging literacy skills and children's success at school: Awareness of Self and Environment, Social Skills and Approaches to Learning, Cognitive Skills, Language and Communication (hereafter, called Language Skills), and Physical Development (The Learning Bar [TLB], 2017a).

The *EIT* was nationally administered in Uruguay during the 2017 school year with children aged 4 and 5 enrolled in Kindergarten. Teachers observed their students for 3 weeks at the beginning of the school year, and assigned them a score for each item on the five domains assessed. We refer to this as the 'pre-assessment'. A second assessment was administered to those children identified in the pre-assessment as being 'vulnerable', referred to as the 'post-assessment'. A child's level of vulnerability is determined for the *EIT* with a formula that provides an estimate of the likelihood a child will be a struggling reader at the end of grade 2. The formula, based on a logistic regression analysis using longitudinal data from five school divisions, found that the strongest predictors for later reading

achievement were cognitive skills and language skills. For Uruguay, as in other countries, children with less than an 80% probability of being a successful reader, were considered vulnerable. The threshold is set high to ensure that the prevalence of 'false negatives' is relatively low. In other words, we are willing to risk providing extra support for a child even though it may not have been necessary than risking have a child 'fall through the cracks'.

Cognitive Skills The *EIT* measure of cognitive skills assesses foundational pre-literacy and numeracy skills (The Learning Bar [TLB], 2017b). The measure includes 8 items or skills per domain, and for each skill teachers indicate whether the child is: unable to do it (0), can do it partially or sometimes (1), can usually do it (2) or can do it consistently (3). The overall domain score is the average of the 8 items. The cognitive scale has a reliability of 0.90 for the Uruguayan sample.

Language Skills The *EIT* measure of language skills assesses children's receptive language skills, the ability to listen to and understand instructions, discussions, and stories; and expressive language skills, the ability to communicate and be understood by others. The measure of language skills also includes 8 items and is scored in the same way as cognitive skills. The language scale has a reliability of 0.93 for the Uruguayan sample.

Age It refers to the children's age when they were assessed. It was estimated with the date of birth and the date of the assessment. The age is calculated as a continuous variable, which ranges from 4.03 to 4.99.

Gender Gender is a binary variable representing the biological sex of the child.

Table 13.1 shows the descriptive statistics for the child-level variables.

Family Questionnaire A questionnaire was administered in September 2015 by the Council of Initial and Elementary Education (CEIP) and the Area of Research, Evaluation and Statistics (DIEE) of Uruguay, to a sample of parents of the children attending kindergarten. The questionnaire measures three components to estimate family SES: home education level, economic status, and home social integration (Administración Nacional de Educación Pública, Consejo Directivo Central, División de Investigación, Evaluación y Estadística and Departamento de Investigación y Estadística Educativa [ANEP, CDC, DIEE, DIEE], 2011). The individual-level data

Table 13.1 Descriptive statistics of the child sample (N = 22,582)

	Min	Max	Mean	SD	Skewness
Outcome measures					
Cognitive skills	0.00	3.00	1.00	0.70	0.74
Language skills	0.00	3.00	1.82	0.77	−0.32
Child-level covariates					
Age	4.03	4.99	4.54	0.26	
Gender (females = 1; males =0)	0	1	0.49	0.50	

were aggregated to the school level and classified into quintiles. Data on the SES quintile for each school were available for this research.

13.4.2 Sample

The study was based on data from the national sample of 4-year old children. These data include pre-assessment *EIT* scores for 22,582 children, aged 4.0–4.99, who were enrolled in 1591 public school classrooms in 1228 schools. Nine hundred and eighty-two schools – 80% of the sample – had only one classroom with age 4 children and an additional 171 schools had two classrooms. Only 75 schools had three or more classrooms. Figure 13.1 shows the distribution of classrooms by the ANEP SES quintiles. If the sample were representative of the national sample, then we would expect there to be 318 classrooms in each quintile. The distribution suggests that the sample has a lower number of classrooms than expected in the lowest two quintiles, and a higher number in the top two quintiles. The questions pertaining to student growth were based on the sub-sample of children who were deemed vulnerable based on the pre-assessment. This subsample included 13,994 children.

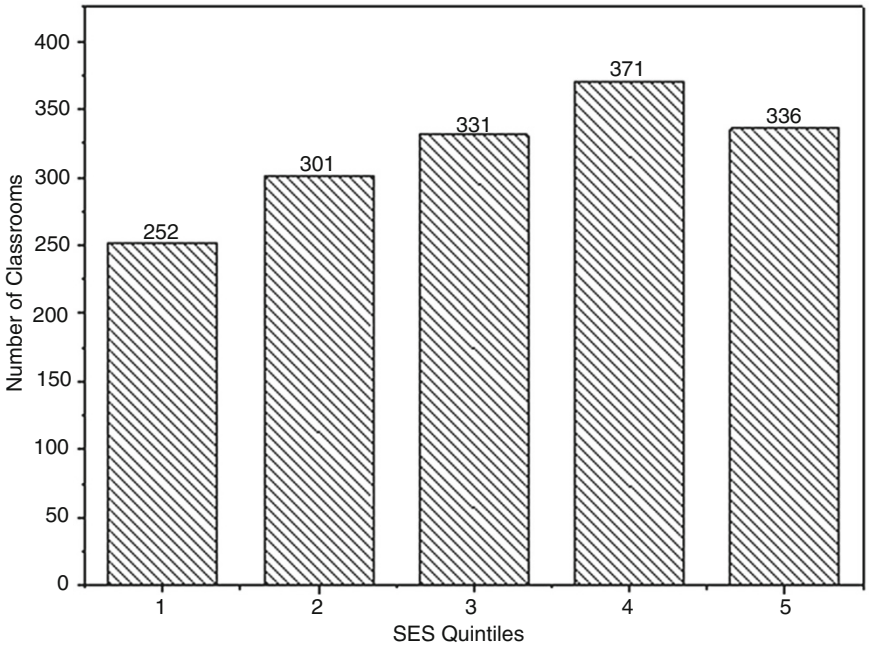

Fig. 13.1 Distribution of classrooms by ANEP SES Quintile

13.4.3 Analysis

Two sets of analyses were conducted to address the research questions pertaining to the variation in children's language and cognitive skills when they enter Kinder-4 and the extent to which children improve their levels of cognitive and language skills over the course of the school year. Both sets of analyses fit a series of hierarchical linear models (HLM), with children at Level 1, nested within classrooms at Level 2.

For the first set of questions, the outcome variables were the *EIT* scores for cognitive and language skills, respectively. The following models were fit for each outcome: a 'null model' with no covariates; a 'base model', which included gender and age at level 1; and a 'full model', which included gender and age at Level 1 and a set of four dummy variables representing the five SES quintiles at Level 2. The quintile variables were used to explain variation in the intercepts and the age slopes.

For the second set of research questions, the outcome variables were each child's annual growth in cognitive and language skills, respectively. A child's average daily growth was calculated by subtracting the domain pre-test score (γ_1) from the post-test score (γ_2) and dividing by the number of days between the pre- and post-assessment dates (*pre-post period*). The annual growth is the average daily growth multiplied by 270, which is the number of calendar days in the Uruguayan school year (Administración Nacional de Educación Pública [ANEP], 2018). The calculation is:

$$Annual\ Growth = \frac{(\gamma_2 - \gamma_1) * 270}{pre - post\ period}$$

The analyses for this set of questions was identical to those of the first set, except that the outcome refers to annual growth scores.

In conducting these analyses, we considered using a three-level model, with children nested within classes at Level 1, classes nested within schools at Level 2, and between school variation at Level 3. However, we found that the coefficients from the three- and two-level models were virtually identical, and thus, for simplicity, we preferred to report the results from the two-level model.

Finally, to address the question about whether the rate of growth exceeded the maturity effect for the majority of classrooms, we found that the maturity effect did not vary substantially among classrooms, and therefore we simply estimated the 'classroom effect' as the difference between the estimate of each classroom's annual growth and the national average maturity effect.

13.5 Results

13.5.1 Pre-literacy Skills upon Entry into Kinder-4

Table 13.2 shows the results of the HLM analyses for the first set of questions, for both language and cognitive skills.

Table 13.2 Estimates of regression coefficients and standard error for models pertaining to Cognitive and Language skills upon Kinder-4

	Cognitive skills						Language skills					
	Null model		Base model		Full model		Null model		Base model		Full model	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Intercept	**0.997**	(0.012)	**0.977**	(0.011)	**0.975**	(0.025)	**1.811**	(0.013)	**1.790**	(0.013)	**1.812**	(0.029)
SES Quintile 1					**−0.126**	(0.038)					**−0.151**	(0.043)
SES Quintile 2					−0.069	(0.036)					−0.067	(0.041)
SES Quintile 4					0.043	(0.035)					−0.008	(0.037)
SES Quintile 5					**0.119**	(0.036)					**0.080**	(0.040)
Female			**0.177**	(0.007)	**0.177**	(0.008)			**0.216**	(0.008)	**0.216**	(0.008)
Age (maturity effect)			**0.573**	(0.015)	**0.515**	(0.035)			**0.599**	(0.017)	**0.554**	(0.037)
SES Quintile 1					−0.023	(0.051)					0.070	(0.056)
SES Quintile 2					0.017	(0.049)					−0.004	(0.053)
SES Quintile 4					**0.142**	(0.048)					0.072	(0.050)
SES Quintile 5					**0.121**	(0.050)					0.082	(0.052)
Variance components	Var		Var	%	Var	%	Var		Var	%	Var	%
Among children	0.314		0.279	11.15%	0.280		0.394		0.355	9.90%	0.356	
Among classes	0.180		**0.176**	2.22%	**0.169**	4.00%	0.189		0.195	–	0.190	2.6%
Age slopes			**0.060**		**0.056**	6.67%			0.033		0.033	–

Note: Figures in bold font are statistically significant ($p < 0.05$)

Variation Among Children Within Classrooms Per the results in Table 13.2, the null model indicates that the average level of cognitive skills within classrooms across the 1591 classes is 0.997. The average within-classroom variance is 0.314, and therefore the average within-classroom standard deviation is ($\sqrt{0.314} =$) 0.56. Therefore, in a typical classroom, children's scores vary by more than 1 full point on the 3-point scale of the *EIT*. For the test of language skills, the average score within classrooms is 1.811. The variance is 0.394, corresponding to a standard deviation of 0.63. The average scores for the language test are on average considerably higher than those of the cognitive test, and the variation is slightly greater.

Results for the base model indicate that children's scores when they start Kinder-4 are related to their gender and age. On average, girls' scores for cognitive skills are 0.177 points higher than those of boys, and for language skills they are 0.216 points higher. The average within-class slope for age, which we refer to as the 'maturity effect', is 0.573 points for cognitive skills and 0.599 for language skills. These estimates indicate the difference in skills, on average, for a child who is 4 years old upon entry into Kinder-4 and a child who is 5 years old. These are large differences. Figures 13.2 and 13.3 show the average levels of cognitive and language skills for children by their age when the assessment was conducted. The age variable was divided into 12 equal intervals, such that an age of 4–0 refers to children who were aged 4.00–4.083 years old. The intervals roughly correspond to a child's month of birth, with children who were 4–0 being young for their cohort and those who were 4–11 being old for their cohort.

These results can be used to situate the gender effects in a 'maturity' perspective. For cognitive skills, the 0.177 point female advantage is comparable to about 2.5 months of cognitive skill development. For example, a boy who was 4 years 6 months old upon entry into Kinder-4, on average would have cognitive skills comparable to a girl who was 4 years 3.5 months old. For language skills, the female advantage is about 4.3 months.

Variation Among Classroom in Average Skills Levels The results from the null model indicate that classes vary significantly in their average levels of cognitive skills upon entry to Kinder-4. Per Table 13.2 results, the variance is 0.180, with a standard deviation of 0.424. Thus, while the average within-school mean is 0.997, the individual class means vary by about plus or minus 0.8 points. The inclusion of gender and age in the base model did not substantially reduce the variance; one would not expect this to be the case as gender and age tend to be evenly distributed among classes. The full model includes the SES quintiles. Schools in the lowest quintile had average scores that were 0.126 lower than those of the third quintile, while schools in the highest quintile were about 0.119 points higher. The age effect also varied by SES quintile, with slopes being steeper in the top two quintiles.

For language skills, the variance among classes was 0.189, approximately the same as for cognitive skills. Also, as with the cognitive skills, average scores were

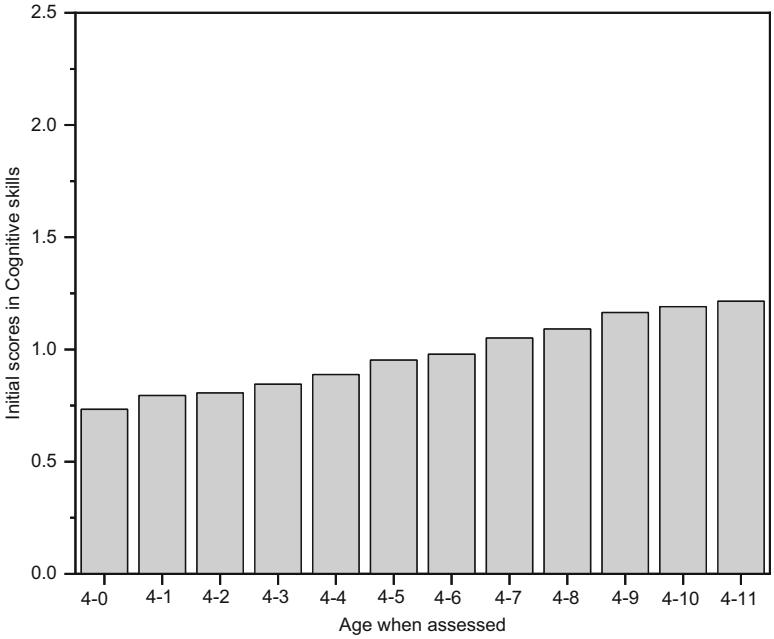

Fig. 13.2 The 'maturity effect' for cognitive skills

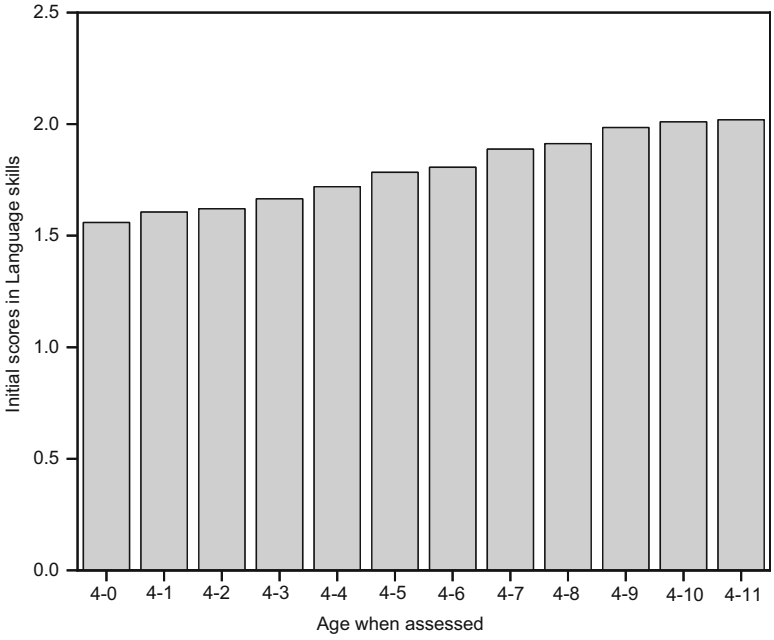

Fig. 13.3 The 'maturity effect' for language skills

lower in the lowest quintile and higher in the highest quintile, with differences similar to those observed with the cognitive test scores.

13.5.2 Annual Growth During Kinder-4

Table 13.3 shows the results of the HLM analyses for the annual growth in cognitive and language skills for the subsample of vulnerable children that was tested near the beginning and end of the school year. As noted earlier, the outcome variables were adjusted to provide estimates of annual growth.

Variation in Annual Growth Among Children Within Classrooms The average growth from the pre-test to the post-test is 1.586 points for cognitive skills and 1.184 points for language skills. The amount of growth varies significantly among children within classrooms – the variance is 0.762 and 0.601 for cognitive and language skills, respectively. Results for the base model indicate that the skill gains were greater for girls in cognitive skills, but this same effect was not evident for language skills. Results for the base model indicate that for cognitive skills, children who were relatively old for their cohort made greater gains than those who were relatively young for their cohort; while for language skills the effects were in the opposite direction.

Variation in Annual Growth Among Classrooms The null model indicates that the variance among classes in annual growth was 0.784 for cognitive skills and 0.643 for language skills. These variance components are comparable to the within-class variance estimates, and thus we can conclude that about one-half of the total variation in growth over the course of the school year is *within* classes and one-half is *between* classes. The full model results indicate that SES does not play a major role in explaining the variation in pre-post gains of children that had been identified as vulnerable.

Classroom Effects Adjusted for the Maturity Effect Given that the mean SES of the school accounted for only about 7% of variation of the maturity effects among classrooms for cognitive skills, and a negligible effect for language skills, we fixed the maturity effect at 0.573 for cognitive skills and 0.599 for language skills. The results indicate that the average annual growth for cognitive skills (1.586) exceeds the maturity effect (0.573) by nearly one full point. For language skills, the average classroom effect (1.184) exceeds the maturity effect (0.599) by about 0.6 points.

Table 13.3 Estimates of regression coefficients and standard error for models pertaining to annual growth in Cognitive and Language skills

	Cognitive skills						Language skills					
	Null model		Base model		Full model		Null model		Base model		Full model	
	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)	Effect	(SE)
Intercept	**1.586**	(0.026)	**1.576**	(0.026)	**1.567**	(0.058)	**1.184**	(0.024)	**1.198**	(0.024)	**1.128**	(0.052)
SES Quintile 1					−0.068	(0.088)					0.090	(0.078)
SES Quintile 2					0.014	(0.086)					0.062	(0.080)
SES Quintile 4					0.010	(0.078)					0.056	(0.070)
SES Quintile 5					0.072	(0.079)					**0.148**	(0.073)
Female			**0.104**	(0.017)	**0.104**	(0.017)			0.010	(0.016)	0.009	(0.015)
Age			**0.310**	(0.034)	**0.265**	(0.080)			**−0.252**	(0.030)	**−1.170**	(0.066)
SES Quintile 1					0.170	(0.112)					−0.071	(0.102)
SES Quintile 2					0.001	(0.111)					−0.001	(0.092)
SES Quintile 4					0.021	(0.108)					−0.093	(0.089)
SES Quintile 5					0.051	(0.107)					**−0.247**	(0.095)
Variance components	Var		Var	%	Var	%	Var		Var	%	Var	%
Among children	0.762		0.736	3.41%	0.736		0.601		0.584	2.83%	0.584	
Among classes	0.784		**0.780**	0.51%	0.781	–	0.643		0.656	–	0.654	0.30%
Age slopes			**0.260**		0.264	–			0.185		0.177	4.32%

Note: Figures in bold font are statistically significant ($p < 0.05$)

13.6 Discussion and Policy Implications

Findings from several studies on the effects of ECEC and kindergarten programs indicate that attending an early childhood program has a positive effect on early childhood outcomes and that the effect is strengthened when children have acquired particular pre-literacy skills before they begin the early primary school period. Logically, if an ECEC or kindergarten program has a positive effect on the requisite pre-literacy skills, children are more likely to have better schooling outcomes during the early primary school period and thereafter. This study is based on data from a national evaluation in Uruguay of four-year old children that were assessed at the beginning of the school year using the *Early Years Evaluation* (*EYE*, or *EIT* in Spanish). Children who were deemed vulnerable on the assessment were assessed again at the end of the school year.

Five key findings emanate from this research:

1. Children varied considerably within classes in their levels of skill development in the cognitive and language domains at the beginning of the school year.
2. The average levels of cognitive and language skill development varied significantly among classes at the beginning of the school year.
3. Children with relatively low scores at the beginning of the school year made significant growth in their cognitive and language skills.
4. The average level of annual growth in cognitive and language skills varied substantially within and among classes.
5. Annual growth exceeded the maturity effect, that is, the expected growth if children did not attend kindergarten. The average annual growth, net of the maturity effect, is nearly one full point for cognitive skills and 0.6 points for language skills (in a scale from 0 to 3).

The Uruguayan government has made great strides in meeting the SDG goal of ensuring that "all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education" (United Nations [UN], 2015, p.17). However, in low-income countries fewer than 20% of children aged 3- to 6-years old attend pre-primary education (World Bank, 2018b). In Latin America the enrollment rate is about 65% (Engle et al., 2007). The results of this study add to a growing body of literature that confirms access and attendance in kindergarten programs makes a difference.

Three key features of this study are that it is based on results of a universal kindergarten program for all children; it uses a reliable assessment that measures a set of outcomes relevant to children's success in school; and it assesses the growth of a vulnerable population over the course of the school year. The Uruguayan Ministry conducted an external evaluation of the *EIT* and found that 76% of teachers indicated it helped them plan their activities and led to positive changes in their day-to-day practice (Administración Nacional de Educación Pública [ANEP], 2017). Other factors that have contributed to its successful implementation are: the *EIT* has strong psychometric properties (The Learning Bar [TLB], 2016); it was contextualized to meet the needs of Uruguayan teachers; and it is population-based, involving all

kindergarten teachers in the assessment of their students (Lopez, 2016; Willms, 2018b).

This study is unique in that it provides estimates of the range of children's skills in cognitive and language development when children enter kindergarten at age 4. The range of development is substantial. This study found that almost two-thirds of the variation in children's skills is *within* classrooms. This finding emphasizes the challenge that most kindergarten teachers face: from the first day of school they need to meet the needs of children with a diverse range of ability. This underlies the need for pre-service and in-service teacher training to provide strategies for teaching in heterogeneous classrooms. This finding also brings attention to an inherent tension that many teachers feel: they are charged with teaching a state-mandated, age-based curriculum, but the children in their classes are at very different levels of development.

This tension is likely felt even more acutely by Grade 1 teachers if they feel that they are unable to meet the needs of students with low levels of pre-literacy skills. Willms (2018a, p. 20) noted that for the full population of Uruguayan children, "[i]f one considers all 5-year old children, who are set to enter primary school, those in the bottom quartile have skills that are comparable to those aged 4–4.5. In other words, they are more than one full year behind their peers in their cognitive skill development". The results call for alternative models for educating young children, including flexible curricular arrangements and ungraded classes (e.g., see Farrell, Manion, & Rincón-Gallardo, 2017).

About one-third of the variation in children's skills is *between* classrooms. Given that about 80% of the schools in the sample had only one classroom, most of this variation is also among schools. This level of 'vertical segregation' (Willms, 2010) is comparable to the extent of variation among schools in most Latin American countries at the end of Grade 4 (Murillo & Carrasco, 2011; Willms & Somers, 2001). A practical approach to reducing this variation is to use data describing children's skills at the beginning of kindergarten to allocate students to classes such that each class has a similar number of vulnerable students. At department and national levels the results provide information that can be used to determine which schools need additional resources. A vulnerability concentration plot of the results for Uruguay can show the concentration of vulnerable students among schools in the country (Willms, 2018a).

The variation in the annual growth among classrooms observed in this study could be attributed to many factors, including the family background of the students and various community factors that lie outside the control of teachers and school administrators. However, a large body of research suggests that quality of instruction is a key factor in explaining this variation. In some jurisdictions, this kind of data has been used to hold teachers and schools accountable, in many cases unfairly. The Educational Prosperity model calls for a shift towards the use of such data as 'leading indicators', rather than 'trailing indicators' for accountability (Willms, 2018a). This study shows that a national evaluation of early childhood skills can be used at the national level to identify strengths and weaknesses of the schooling system, inform policies aimed at improving outcomes and reducing inequalities, allocating

resources in ways that ensure equity of provision, and developing and evaluating interventions. At the school and classroom levels, a national evaluation can be used by educators to guide classroom practice, identify vulnerable students, set instructional goals and involve parents in meaningful ways.

The 1989 UN Convention on the Rights of the Child maintains that children have a right to a quality education, with a view to achieving these rights progressively and on the basis of equal opportunity. The evidence showing the positive effects of early childhood education and its potential for reducing inequalities between advantaged and disadvantaged groups supports the SDG goal for the provision of high-quality universal early childhood programs. It is a right for children and an obligation of governments. Achieving this goal will give all children the opportunity to thrive.

References

Administración Nacional de Educación Pública. (2017). *Evaluación Infantil Temprana 2017*. Encuesta. Montevideo, Uruguay: Author.

Administración Nacional de Educación Pública. (2018). *Monitor Educativo de Enseñanza Primaria. Estado de situación 2017*. Montevideo, Uruguay: Author.

Administración Nacional de Educación Pública, Consejo Directivo Central, División de Investigación, Evaluación y Estadística and Departamento de Investigación y Estadística Educativa. (2011). Relevamiento de Características Socioculturales 2010 de las escuelas del Consejo de Educación Inicial y Primaria. *Documentación de la metodología*. Montevideo, Uruguay: Autor.

Aguilar, R., & Tansini, R. (2012). Joint analysis of preschool attendance and school performance in the short and long-run. *International Journal of Educational Development, 32*(2), 224–231.

Anders, Y., Grosse, C., Rossbach, H. G., Ebert, S., & Weinert, S. (2013). Preschool and primary school influences on the development of children's early numeracy skills between the ages of 3 and 7 years in Germany. *School Effectiveness and School Improvement, 24*(2), 195–211.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science, 333*, 975–978.

Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *Review of Economics and Statistics, 86*(1), 108–132.

Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics, 93*(1–2), 219–234.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology, 98*(3), 489–507.

Cortázar, A. (2015). Long-term effects of public early childhood education on academic achievement in Chile. *Early Childhood Research Quarterly, 32*, 13–22.

Cueto, S., León, J., Miranda, A., Dearden, K., Crookston, B. T., & Behrman, J. R. (2016). Does pre-school improve cognitive abilities among children with early-life stunting? A longitudinal study for Peru. *International Journal of Educational Research, 75*, 102–114.

DiPerna, J. C., Lei, P. W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the Early Childhood Longitudinal Study–Kindergarten cohort. *Journal of Educational Psychology, 99*(2), 369–379.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

Ehrlich, S. B., Gwynne, J. A., & Allensworth, E. M. (2018). Pre-kindergarten attendance matters: Early chronic absence patterns and relationships to learning outcomes. *Early Childhood Research Quarterly, 44*, 136–151.

Engle, P. L., Black, M. M., Behrman, J. R., De Mello, M. C., Gertler, P. J., Kapiriri, L., … International Child Development Steering Group. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet, 369*(9557), 229–242.

Farrell, J., Manion, C., & Rincón-Gallardo, S. (2017). Reinventing schooling: Successful radical alternatives from the global south. In K. Bickmore, R. Hayhoe, C. Manion, K. Mundy, & R. Read (Eds.), *Comparative and international education: Issues for teachers* (pp. 59–87). Toronto, Canada: Canadian Scholars' Press.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Hall, J., Sylva, K., Sammons, P., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013). Can preschool protect young children's cognitive and social development? Variation by center quality and duration of attendance. *School Effectiveness and School Improvement, 24*(2), 155–176.

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly, 36*, 561–571.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Holod, A., Ogut, B., de los Reyes, I. B, Quick, H. E., & Manship, K. (2018). The effects of prekindergarten for Spanish-speaking dual language learners: Evidence from California's transitional kindergarten program. *Early Childhood Research Quarterly*. Advance on line publication. https://doi.org/10.1016/j.ecresq.2018.06.002

Kay, N., & Pennucci, A. (2014). *Early childhood education for low-income students: A review of the evidence and benefit–cost analysis* (Doc. No. 14-01-2201). Olympia: Washington State Institute for Public Policy.

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*(4), 443–484.

Lehrl, S., & Smidt, W. (2018). Differential effects of preschool quality on children's emergent literacy skills in the final preschool year in Germany. *Research Papers in Education, 33*(4), 492–514.

Linan-Thompson, S. (2014). *Screening of school readiness skills: A review of the literature*. Washington, DC: US AID. Retrieved from https://pdf.usaid.gov/pdf_docs/PA00JW1P.pdf

Lopez, A. Y. (2016). *Steps towards a national implementation of the early years evaluation in Uruguay: Actors and actions*. Paper presented at the inaugural symposium of the comparative and international education society, Scottsdale, AZ.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology, 98*(1), 14–28.

Murillo, F. J. (2007). School effectiveness research in Latin America. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 75–92). Dordrecht, The Netherlands: Springer.

Murillo, F. J., & Carrasco, M. R. (2011). ¿La escuela o la cuna? Evidencias sobre su aportación al rendimiento de los estudiantes de América Latina. Estudio multinivel sobre la estimación de los efectos escolares. *Profesorado Revista de Curriculum y Formación de Profesorado, 15*(3), 27–50.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Retrieved from http://lincs.ed.gov/publications/pdf/NELPReport09.pdf

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of Reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.

Polo, D. S., Araujo, N. P., & Salceda, J. C. (2017). La concepción simple de la lectura en alumnos de 4° de primaria de una escuela fiscal de Quito. *Alteridad, 12*(1), 228–235.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist, 53*(2), 109–120.

Ripoll, J. C., Aguado, G., & Castilla-Earls, A. P. (2014). The simple view of reading in elementary school: A systematic review. *Revista de Logopedia, Foniatría y Audiología, 34*, 17–31.

Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology, 81*(1), 101–108.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford Press.

Schatschneider, C., Petscher, Y., & Williams, K. M. (2008). How to evaluate a screening process: The vocabulary of screening and what educators need to know. In L. M. Justice & C. Vukelich (Eds.), *Solving problems in the teaching of literacy. Achieving excellence in preschool literacy instruction* (pp. 304–316). New York, NY: Guilford Press.

Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2011). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly, 26*(1), 42–49.

Storch, S., & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. https://doi.org/10.1037//0012-1649.38.6.934

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). *Effective pre-school and primary education 3-11 project (EPPE 3-11): Final report from the primary phase: Pre-school, school and family influences on children's development during key stage 2 (7–11)* (Research report DCSF-RR061). Retrieved from University of London website: http://eprints.bbk.ac.uk/7602/

Taggart, B., Sylva, K., Melhuish, E., Sammons, P., & Siraj, I. (2015). *Effective pre-school, primary and secondary education project (EPPSE 3-16+): How pre-school influences children and young people's attainment and developmental outcomes over time*. London, UK: Department for Education.

The Learning Bar Inc. (2011a). *Early years evaluation [Measurement instrument]*. Retrieved from http://www.thelearningbar.com/

The Learning Bar Inc. (2011b). *Early years evaluation teacher assessment [Measurement instrument]*. Retrieved from http://www.thelearningbar.com/

The Learning Bar Inc. (2016). *Informe Psicométrico. Evaluación Infantil Temprana. Uruguay 2016*. Fredericton, NB: Author.

The Learning Bar Inc. (2017a). *Early years evaluation psychometric report. Ecuador Pilot Study – 2016*. Fredericton, NB: Author.

The Learning Bar Inc. (2017b). *The early years evaluation – Theoretical and conceptual framework*. Fredericton, NB: Author.

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33*(3), 321–363.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the school years. *Scientific Studies of Reading, 15*(1), 8–25.

Willms, J. D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112*(4), 1008–1037.

Willms, J. D. (2018a). *Learning divides: Using data to inform educational policy*. Montreal, Canada: UNESCO Institute for Statistics.

Willms, J. D. (2018b). Educational prosperity: An assessment strategy for supporting student learning in low-income countries. In D. A. Wagner, S. Wolf, & R. F. Boruch (Eds.), *Learning at the bottom of the pyramid: Science, measurement and policy in low-income countries*. Paris, France: UNESCO-IIEP.

Willms, J. D., & Somers, M. A. (2001). Family, classroom and school effects on children's educational outcomes in Latin America. *International Journal of School Effectiveness and Improvement, 12*(4), 409–445.

World Bank. (2018a). *World development report 2018: Learning to realize Education's promise*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1096-1

World Bank. (2018b). *World Bank national accounts data, and OECD National Accounts data files*. Washington, DC: World Bank. Retrieved from https://data.worldbank.org/indicator/-NY.GNP.PCAP.PP.CD?locations=UY

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., … Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. Washington, DC: Society for Research in Children and Foundation for Child Development.

Chapter 14
Continuing Towards International Perspectives in Educational Effectiveness Research

James Hall, Pamela Sammons, and Ariel Lindorff

14.1 Introduction

This volume describes the history of Educational Effectiveness Research (EER), illustrates some of the main contemporary developments in EER theory and knowledge, and shows the broad variation in the approaches to, and foci of, EER as it currently exists across the world. Towards an international perspective, the chapters within this book have been written as part of the fourth and latest phase of EER – a phase that Reynolds et al. (2014) characterise as showing,

> . . .the internationalization of the field, together with the merger or synergy of approaches generated by having, for example, school effectiveness researchers in close intellectual proximity to school improvement researchers and practitioners. (p.199)

This book provides a range of examples of these mergers and synergies in action with the consequence that there are common themes running across chapters. This final chapter begins by considering these common themes for a fuller understanding

J. Hall (✉)
Southampton Education School, University of Southampton, Southampton, UK
e-mail: j.e.hall@soton.ac.uk

P. Sammons · A. Lindorff
Department of Education, University of Oxford, Oxford, UK
e-mail: pamela.sammons@education.ox.ac.uk; ariel.lindorff@education.ox.ac.uk

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3_14

of how EER is achieving an EER perspective (extending the discussion in Chap. 2[1]), It then reflects on future directions for EER and explores some future directions for an international perspective in EER.

14.2 Common Themes Within This Volume

Unsurprisingly, a common theme within this volume is that of the contact and collaboration that exists between Educational Effectiveness Research (EER) and International Large Scale Assessment (ILSA) studies. Of course, underling this contact and collaboration are shared understandings about how the world works and the research methods one can use to understand educational structures, processes and relationships (in other words, shared paradigms). Figure 14.1 illustrates the

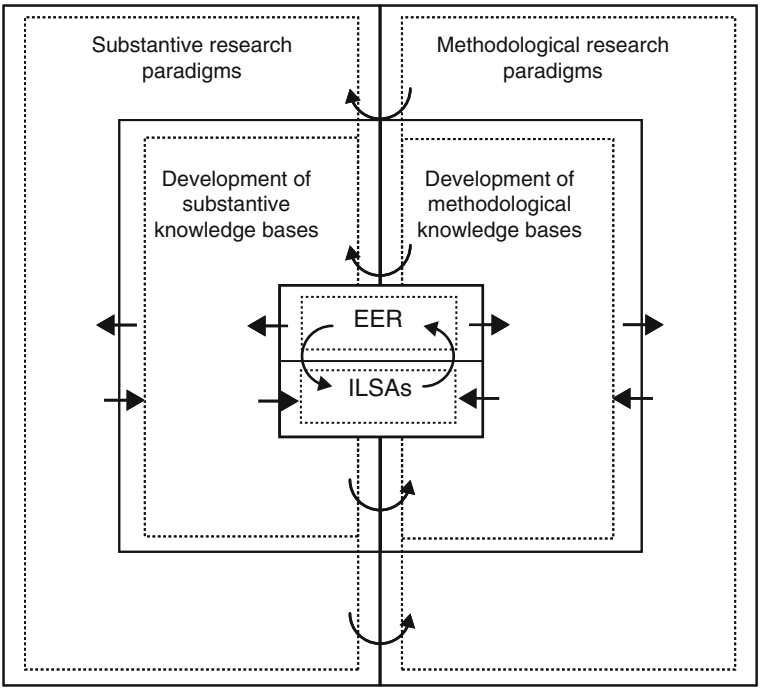

Fig. 14.1 The reciprocal relationship shared between Educational Effectiveness Research (EER) and International Large Scale Assessment studies (ILSAs) and the nesting of these relationships within substantive and methodological knowledge bases and research paradigms

[1]Lindorff, A., Sammons, P., & Hall, J. (n.d.). International perspectives in educational effectiveness research: A historical overview. In J. Hall, A. Lindorff, & P. Sammons. (Eds.). *International perspectives in educational effectiveness research*. Dordrecht, Netherlands: Springer

contacts between the two fields and differentiates these according to substantive knowledge (regarding education) and research methodology (in keeping with Creemers, Kyriakides & Sammons, 2010). Examples of the contact and collaboration between EER and ISLA studies can be found throughout this book, but it is particularly apparent within Chaps. 2, 6[2] and 7[3]. Lindorff and colleagues (Chap. 2) discuss the convergence of EER and ILSA research methods and knowledge bases via EER secondary analyses of ILSA data and, reciprocally, input from EER into conceptual frameworks for ILSA, Klieme (Chap. 7) provides examples of the contacts between EER and ILSAs, and Reynolds and colleagues (Chap. 6) discuss the reviews of EER evidence that have been commissioned by the World Bank, the European Economic Community (EEC), and the International Association for Evaluation of Achievement (IEA) – the organisation responsible for many of the early international comparisons of education.

Figure 14.1 also illustrates thatthe relationship between the knowledge and methods of EER and ILSAs is mutually informative. For example, on one hand, calls exist for questionnaires in ILSAs to be based on EER theories (e.g. Jude & Kuger, 2018) while on the other, ILSAs can trigger new studies (for example by revealing new examples or forms of educational inequalities), and provide new research tools that EER can use. We see an example of the latter within Chap. 9[4] where the results of an EER study carried out in mainland China are presented that use questionnaire items that were originally developed within an ILSA (the Teaching and Learning International Survey, TALIS; e.g. Jensen, Sandoval-Hernández, Knoll, & Gonzalez, 2012).

A second prominent theme within the book is '*context stripping'*. Here, context refers to the unique place and time within which a piece of research is carried out with both being understood to shape, in the case of educational research, the educational structures, processes, and drivers of student outcomes and educational progress. Thus, context stripping refers to the lack of awareness of the context within which a piece of research is designed, analysed, and then interpreted. For example, Chaps. 6 and 7 discuss the hazards of this practice in ILSAs, in the relationship

[2]Reynolds, D., Kelly, A., Harris, A., Jones, M., Adams, D., Miao, Z., & Bokhove, C. (n.d.). Extending educational effectiveness: A critical review of research approaches in international effectiveness research, and proposals to improve them. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

[3]Klieme, E. (n.d.). Policies and practices of assessment: A showcase for the use (and misuse) of International Large Scale Assessments in educational effectiveness research. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

[4]Thomas, S.M. (n.d.). School and teacher value added performance and the relationship with teacher professional development in mainland China. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

between ILSAs and EER[5], and in instances where evidence has been misunderstood or misused by some policy makers.

Fortunately, a solution to limit context stripping and the hazards of context stripping within EER is already happening. As the field continues to develop, share, and then test *theories* of effective education, so greater and more nuanced understandings of the importance of time and place (context) are coming to the fore. Take for example the dynamic model of educational effectiveness (Creemers & Kyriakides, 2006) and the dynamic approach to school improvement (Creemers, Kyriakides, & Antoniou, 2013; see Chap. 3). However, continual critical appraisal is still required from all those who engage with knowledge bases of educational research (including EER) as regards the context-specificity of both educational knowledge and the methodological tools used to generate this (currently seldom studied; see Chap. 10[6]). For example, Chap. 7 provides evidence of a positive shift in the rigour of the analysis and interpretation of ILSA data by researchers. While this shift towards greater analytic rigour is admirable, there still remain concerns regarding context stripping in the development and reporting of the measurement tools used in educational research. This includes the presence of equivalent terms in different languages (e.g. "leadership" versus "administration"; see Chap. 12[7]), and the presence of synonymous terms within a language that are non-synonymous across cultures and countries where this language is used. Furthermore, the qualitative methodologies that are prominent within studies of school improvement, and that are now being included more frequently in studies of educational effectiveness (often via the use of mixed methods designs; Teddlie & Sammons, 2010; Reynolds et al., 2014; Sammons, Davis, Day, & Gu, 2014), are particularly valued for their attempts to support resistance to context stripping (see Sect. 14.3.2). The educational improvement and educational change literature also provides relevant insight into the complexities of education systems and the contexts within which they operate; as Ainscow, Dyson, Goldrick, and West (2012) suggest, "the extent to which students' experiences and outcomes are equitable is not dependent only on the educational practices of their teachers, or even their schools", but "on a whole range of interacting processes that reach into the school from outside" (p. 198).

[5]The hazards of context stripping are of course also known elsewhere and are not particular to EER. For example, the failure to consider school truancy as simultaneously both a rational and irrational act by different stakeholders (e.g. by students [perhaps to avoid bullying] vs. parents and/or schools; see Schoon, 2006). Alternatively, the tendency for developmental resilience to be subject to doomed attempts to generate lists of universal personality states and traits, "*It is a fallacious approach, however, because resilience is not a single quality*." (p.4, Rutter, 2006).

[6]Opdenakker, M.-C. (n.d.). Three decades of educational effectiveness research in Belgium and the Netherlands: Key studies, main research topics and findings. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

[7]Townsend, T., Berryman, M., Gurr, D. & Drysdale, L. (n.d.). Leadership for Learning in diverse settings: School leaders setting the agenda in Australia and New Zealand. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

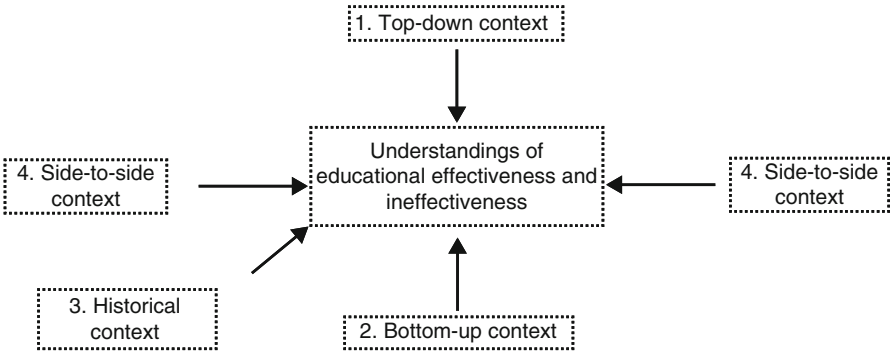

Fig. 14.2 Four main types of contextual awareness needed within EER studies and within the field of EER itself

A third theme apparent across the chapters of this book is also apparent within the two themes discussed so far. There is the need for greater awareness of context both within individual EER investigations and within the broader EER field itself (especially given the possibility of greater variation when carrying out international comparisons); as Chap. 12 states, "context matters". Further, we argue that this contextual awareness can be broken down into at least four different types (see Fig. 14.2):

1. **Top-down contextual awareness.** For example, the need for teacher effects to be studied net of, and also in relation to, broader school or neighbourhood effects;
2. **Bottom-up contextual awareness.** For example, the need to understand school network effects net of, and also in relation to, the unique effects of the particular educational institutions within each network;
3. **Historic contextual awareness.** For example, the need to measure and understand the antecedents and effects of differences in student attainment at school entry and to investigate patterns of change or trajectories of school effects over a period of time;
4. **Side-to-side contextual awareness**. For example, the need to identify and explore the effects of other educational environments such as the home learning environment, parenting, and other potential out of school sources of influence such as those from local neighbourhoods.

While the importance of producing contextualised understandings is core to EER, not all of the four types feature in our current understandings of educational effectiveness and ineffectiveness. For example, Contextual Value-Added (CVA) models of student progress demonstrate historic contextualisation (past attainment; the VA in CVA) and side-to-side contextualisation (student backgrounds; as part of the C in CVA) and in some instances out of school influences (neighbourhood context; again as part of the C in CVA). However, the implementation of CVA models with this explicit contextual awareness within educational practice is inconsistent – in other words, the CVA model is itself subject to top-down system-level contextualisation. Simply, some educational systems value and implement

CVA type models of student progress while others do not and the implications that are drawn from the use of these models also vary across contexts[8]. However, It is also important to bear in mind that the extent to which any of the four types of contextual awareness are important for a study to consider are bound by there being sufficient variability to permit effects to be identified in the first place. For example, educational systems can mandate educational practices, policies, structures, and processes to be consistent across networks, schools, teachers, facilities, classrooms, and lessons[9] thereby potentially limiting any of the four contexts shown in Fig. 14.2. Likewise, though in reverse, policy change at the level of the system (often in no short supply in education) may produce change(s) at all other levels and so permits a plethora of before-and-after comparisons permitting natural experiments to explore possible differences in educational effectiveness and/or ineffectiveness.

A fourth theme within this book is the expanding remit of EER – particularly to focus on a broader range of socioemotional student outcomes such as wellbeing, self-concept, and enjoyment of school (see Chap. 10 for how this has been experienced in Dutch and Belgian EER). While an argument can be made that this expanding remit represents EER returning to the broader set of outcomes deemed important in early seminal studies of primary and secondary schools (e.g. Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Rutter, Maughan, Mortimore, Ouston, & Smith, 1979), the academic landscape (the context) within which EER now considers educational effects for socioemotional student outcomes has, perhaps inevitably, changed. At the time of writing, this is an area of research that is heavily populated by theories and researchers from multiple fields within and outside education. These include theories and researchers aligned to Special Educational Needs and Disability (SEND), educational psychology, psychiatry, and developmental psychopathology (e.g. Hall & Kreppner, 2019). Thus, the agenda of EER increasingly touches or overlaps with other fields[10], and this puts pressure on all to develop theories, methods, and bodies of knowledge that do not 're-invent the wheel' (see also Chap. 7). This reflects the historic roots of earlier school effectiveness studies that involved a combination of researchers from psychological and sociological backgrounds (see Teddlie & Reynolds, 2000). While EER offers a unique additional perspective (particularly for evidence-based unpicking of the educational policies and practices that do and do not work, how, for whom, where, and when), theories, methods and findings still require further integration across fields. Moreover, to be successful such integration requires 'side-to-side awareness'

[8]Note that this portrayal of the varying existence and understanding of CVA models of student progress is in-keeping with a post-positivist worldview.

[9]For example, in an attempt to limit *ineffectiveness* – though such variation-limitation also risks the side effect of limiting *effective* policies, practices, and processes (i.e. inadvertent 'ceiling effects' when only 'floor effects' are intended).

[10]For example, Amartya Sen's Capability Approach (see Sen, 1992, 1999) has been separately applied to both SEND (Norwich, 2014) and to EER (see Chap. 4 & Kelly, 2012)

of parallel bodies of knowledge[11] where there is little historic overlap and communication, plus such integration is not a core remit of any of the fields. Thus, the potential for re-inventions of the wheel remains a very real risk.

Another example where the expanding (re-expanding) remit of EER produces overlap with other fields and so risks 're-inventing the wheel' is via differential educational effectiveness (Mortimore et al., 1988; Sammons, 1996). First, this idea is central to the educational equity agenda of EER that was present from the start, which means that studies of differential effectiveness also lie within the remit of the fields of research that share a focus on educational equity (again including SEND). Second, differential effectiveness offers a bridge from EER and school improvement to the school-based interventions that are developed by academics and professionals with an agenda that includes child development. In short, effective interventions can offer a means towards equity-boosting differential school effectiveness. However, how educational institutions identify, perceive, procure, adapt, deploy, and maintain effective interventions (and vice versa for interventions that are ineffective) as part of the manifestation of equity-enhancing differential school effectiveness (and improvement) is a topic in need of not only further interdisciplinary research but also further interdisciplinary theoretical development, drawing on a diversity of methods and approaches. This area also needs continual investigation because the contexts that surround schools, interventions, and educational effectiveness inevitably alter over time, and this is likely to limit the useful lifespan of particular research findings. In turn, this provides yet another prompt for the development of educational theory, as this is less subject to inevitable changes in educational policy and practice.

14.3 Future Directions...

14.3.1 ...For Educational Effectiveness Research

As the fourth wave of EER progresses, as speculations about future waves of EER are made (see Sect. 1 of this volume), as EER further internationalises and expands its remit, and as new theories are proposed, tested and refined, EER must continue to take stock (maintain historic contextual awareness) of its developing substantive and methodological bodies of knowledge. Towards this, Table 14.1 presents a new heuristic device for locating a study within EER according to the levels that it considers. By using this device, those engaged in EER are facilitated in making cumulative contributions to our understanding of educational effectiveness through targeted responses to gaps in in the current knowledge base. Further, the device does this while achieving the greater contextual awareness (incorporating ideas from

[11]For example, that consideration of socioemotional outcomes (such as student self-concept, motivation, and wellbeing) requires relaxing the assumption of linear expected progress and growth in these outcomes. A plateau or even decline in a developmental trajectory does not necessarily imply either problems in progress or developmental regression (e.g. Rutter & Rutter, 1992)

Table 14.1 Seven levels of educational effectiveness: A heuristic device for locating a study within Educational Effectiveness Research (EER)

Paradigms *(within Epistemes)*	Six hierarchical levels of educational effectiveness			External links	Level 7. Time
Level 1.	**1. Educational Systems** *(with policies & laws shaping levels 2–6; these systems can also be in a hierarchy e.g. region within province within country within international frameworks such as UN conventions)*			**Level 1.** *(e.g. policies for child safeguarding shared across social, health, and justice systems)*	Past events, structures & processes shape those in the future just as those now (and in the future) reframe the meaning of those in the past. For example, historic changes of leadership or staffing in schools; changes in school type or school intake; changes in national, local or school policies
Level 2.	**2. Institutional Groupings within systems** *(e.g. Networks based on shared school type, specific improvement goals, or locality)*			**Level 2.** *(e.g. goal-oriented collaborative working across schools and local education authorities to reduce educational inequity; see* Chapman et al., 2016 *for an example from Scotland)*	
Level 3.	**3. Educational institutions within networks** *(including institutional leadership, policies, practices, funding, & spaces)*			**Level 3.** *(e.g. via multiagency working, signposting, & referrals)*	
Level 4.	**4. Groupings within institutions** *(e.g. year groups [of students], subject departments [of teachers], buildings [containing classrooms])*			**Level 4.** *(e.g. cross-institution subject networks of teachers)*	
Level 5.	**5. Units within groupings** *(sometimes overlapping, sometimes with cross-classified student nesting; always with interplay between the units and with units having the potential for independent or joint effects)*			**Level 5.** *(e.g. learning environments shared with other institutions such as sports fields; professional teacher relationships across institutions)*	
	a. Student clusters *(e.g. ability sets & optioned modules)*	**b. Teachers**	**c. Learning environments** *(e.g. classrooms & exercise areas)*		

(continued)

Table 14.1 (continued)

Paradigms *(within Epistemes)*	Six hierarchical levels of educational effectiveness	External links	Level 7. Time
Level 6.	**6. Students** *(backgrounds & outcomes across: knowledge, skills, attitudes, behaviours, cognitions)*	**Level 6.** *(e.g. links with others outside the school)*	

Note: The variety of student outcomes shown within Level 6 are in response to the diversification of student outcomes that are studied by EER and that are being considered by educational systems internationally (e.g. Chaps. 9, 10, and 11)

Fig. 14.2 and context-specificity cf. Reynolds et al., 2014) that is needed to advance EER. This heuristic device identifies seven levels and has emerged inductively by studying the common themes of the particular contributions to this book, so it is important to note that this is not proposed as a substitute for other theoretical models or frameworks of EER. As a heuristic device for locating a study, it does not seek to outline the variable and dynamic (Creemers and Kyriakides 2006) structures and processes that exist either within or between levels (as do the models and theories presented in Chaps. 3, 10, and 12). Instead, the device is intended to act as a guide to aid the development, testing, and extension of theories of EER – a purpose that is in keeping with many of the chapters of this book. For example, the possibility of new paradigms of equity (Chap. 4[12]), the paucity of research at the level of "the middle tier" including school networks (Chap. 5[13]), and the possibility of multiple cross-level mechanisms by which institutional structures and processes may influence student outcomes (Chaps. 10 and 11). These are all fruitful foci for further investigation and each can be located within the heuristic device outlined in Table 14.1.

Although the levels within the heuristic device presented in Table 14.1 are informed by the preceding chapters and a number of well-established prior hierarchical depictions (including: Bronfenbrenner & Morris, 2006; Creemers, 1994; Creemers & Kyriakides, 2006, 2008; Palardy & Rumberger, 2008; Scheerens, 2016), there are four areas within the device that have been, by comparison, less frequently emphasised in existing EER literature. Each of these is now considered in turn with the intent of demonstrating the utility of the device for future studies in EER.

[12]Kelly, A. (n.d.). The Fifth Phase of Educational Effectiveness Research: The philosophy and measurement of equity. In J. Hall, A. Lindorff, & P. Sammons. (Eds.). *International perspectives in educational effectiveness research*. Dordrecht, Netherlands: Springer

[13]Muijs. (n.d.). Extending educational effectiveness: The middle tier and network effectiveness. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

The first of these four areas concerns the idea that underlying the paradigms governing educational systems and practices[14] (plus trends in educational research) are the "epistemes" (unconscious collective discussions; see Foucault, 1973; Best, 1994[15]) of the culture(s) within which the research paradigm is situated. For example, what groups of learners do the culture(s) under investigation perceive to exist and how is this perception altered when it is communicated across linguistic and cultural boundaries? This because issues of educational equity and differential educational effectiveness cannot be conceived of for groups that are not acknowledged or even perceived to exist in the first place – be this by researchers, policy makers, or educational practitioners[16]. Another example comes from Chap. 7 and the discussion of the variation that can exist between educational systems in the extent to which various actors within these different systems (national or local policy makers, school leaders, teachers, students, parents or communities) take the results of tests of student attainment seriously (which undermines attempts at international comparison).

That rationale for presenting paradigms (within epistemes) at each of the hierarchical levels within the device shown in Table 14.1 is that this is another manifestation of EER's awareness that it needs understandings of educational effectiveness to be both contextualised[17] and context-specific (see Sect. 14.2). In the interest of obtaining more nuanced (and thereby improved) context *sensitivity* in international perspectives on EER (i.e. greater contextual awareness and cultural specificity), future investigations would benefit from better understandings of the various conceptualisations of what education means and, of what, and who education is for, that may vary globally, regionally, locally, institutionally, and individually (embracing and extending the importance of paradigms that is common in school improvement studies, e.g. Hopkins et al., 2014). For example, the nature of interplay of effectiveness factors across the levels in the heuristic device may vary across the cultures and paradigms of the levels under consideration, and by the cultures and paradigms of the levels that may be in-between. The consequence of this culture- and paradigm-dependency is variation in the forms of, and conceptualisation of, what constitutes effective education for different individuals and groups. Thus, EER needs to map the extent of the cultural dependency of the existing knowledge base of

[14]With the paradigms governing educational policies and practices having been argued to feature more prominently in past school improvement studies rather than in investigations of school or teacher effectiveness (Hopkins, Stringfield, Harris, Stoll, & Mackay, 2014)

[15]Epistemes refer to, "*the fundamental codes of a culture – those determining its language, its schemes of perception, its exchanges, its techniques, its values, the hierarchy of its practices*" (p.xx Foucault, 1973; cited by Best, 1994)

[16]See Chap. 4 for more on the philosophies of equity, effectiveness and equality that exist within EER

[17]*Analytically,* this understanding contributed to the development of multilevel statistical models and to statistical models of student progress that are contextualised for salient background effects (see Sect. 14.2). *Theoretically,* this understanding is represented within the dynamic aspect of the EER model by Creemers and Kyriakides (see Chap. 3, Figure 3.1) and by the need for localised understandings of effectiveness in both international EER and in ILSAs (see Chaps. 6 and 7)

recognised educational effectiveness factors with international perspectives on EER being well positioned to contribute to this understanding. While there is existing awareness of this need to better understand cultural dependency[18] (e.g. in Reynolds et al., 2014), systematic investigations will be increasingly required as EER expands and grows (see Sect. 14.3.2). This would include investigations of educational effectiveness (differential or otherwise) as subject to culture- and paradigm-bound definitions and implementations of equity in education (see Chap. 4), and variable expectations for student achievement and progress across ages and stages (e.g. Chap. 13[19]) that prompt further overlap between the agendas of EER and other fields of research including SEND and educational psychology.

How therefore might EER achieve the improved context sensitivity that is needed across hierarchical educational paradigms and cultures? Again, school improvement studies provide one starting point (and a starting point that is less of a leap for the field than would be alternative starting points outside of EER). Those conducting future EER studies need to consider how they will seek to identify and reveal the implicit cultures and paradigms under investigation plus their impacts (potential or otherwise). The heuristic device shown in Fig. 14.1 serves to signpost, but what levels need to be considered, and what structures and processes are put in place to reveal these cultures and paradigms will, of course, vary across different studies. Common across EER studies though, and particularly for investigations operating across cultures and countries, there is the need for a pluralistic perspective in EER research if its findings are to be accurate to, and have credibility with, the individuals, groups, institutions, networks, and systems that are studied.

The second area within the heuristic device shown in Table 14.1 is that educational institutions are increasingly grouped and that these groupings may also demonstrate their own effectiveness and ineffectiveness. Understanding the structures and processes of collaboration and how these relate or respond to equity issues that exist within, between and beyond schools in context (Ainscow, 2016) is an important area for enquiry in EER beyond the level of the school. The relative neglect of EER at this level is touched upon in Chap. 2 and discussed further in Chaps. 3 and 5, with effectiveness at this level being both newer and therefore comparatively neglected compared to studies of effectiveness either at the level of the educational institution (particularly in the USA and UK; see Reynolds et al., 2014) or at the multifaceted level of the classroom/teacher (particularly in the Netherlands and Belgium; see Chap. 10).

[18]For example, Reynolds et al. (2014) separate the internal generalisability of EER findings across two levels: concept and operationalisation. Concepts were argued to be more generalisable than their more culture-bound (localised) operationalisations. For example, the cross-culturally identified importance of school leadership versus the culture-bound forms that is taken by school leadership in order for it to be effective.

[19]Lopez, A.Y., & Willms, J.D. (n.d.). A national evaluation of kindergarten outcomes: Findings from Uruguay. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

Third, that there is a hierarchical level divided into three parts within the heuristic device within which every student in an educational institution is nested. Students can be grouped in various ways – by their classes or different teaching groups or by teachers, by the learning environments in which they learn, and by student clusters that are independent of teacher and environment. The distinction between these units is particularly apparent in early years EER where a distinction is made between "process" and "structural" qualities. These differentiate, respectively, behavioural interactions (person-person & person-environment/resource) and the resources that are available to teachers in the learning environments that they share with their students (e.g. Hall et al., 2013; NICHD Early Child Care Research Network, 2003; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004). Further, while the effects from each of the parts within this multifaceted level are subject to complex theoretical interplay[20], they each also have the potential for their own independent effects (e.g. the effects on students of grouping structures such as from having highly motivated and academic advanced peers; see Chap. 11[21]). In particular, while extent EER literature emphasises teacher effectiveness, class effectiveness, and compositional effects, learning environments are also capable of exerting their own independent effects with these particularly apparent when considering educational *in*effectiveness. For example, some learning environments may impede student progress due to insufficient learning resources, disrepair of spaces and materials, or though the provision of materials that are inappropriate (e.g. to age, stage, or task) such as the use of e-books in classrooms where there are insufficient computer resources (Hall, Khalil, Kelly, & Galbraith, 2019).

Fourth, that there is a hierarchical level, again subject to its own level of effectiveness, that exists between the level of the educational institution and the multifaceted level within which are located teachers and clusters of students in their learning environments (often classrooms, but not always; for example, baby rooms in early years' settings and the varying locations for physical education). This intermediate level refers to groupings within educational institutions. These include student year groups, different 'ability' groupings, subject departments, and

[20]We deliberately avoid use of methodological concepts here, for example, interaction, mediation, and moderation. This for three reasons. First, because Table 14.1 illustrates an heuristic device rather than a framework, theory, or empirical model for testing. Second, because within this device the nature of the interplay between the three parts of level 5 is subject to variation across contexts and over time (as are processes and structures at any of the seven levels). For example, whether (or not) an educational system, institutional grouping, or institution affords its teachers the ability to control the student clusters that exist within the lessons that they teach (cf. educational realities concerned with, amongst others, ability sets, streams, differential teaching, inclusion, and integration). Third, methodological concepts are also subject to their own paradigms and paradigm shifts (e.g. Hall et al.,, 2020; Hall & Sammons, 2013; Creemers et al., 2010)

[21]Palardy, G.J. (n.d.).The impact of socioeconomic segregation in U.S. high schools on achievement, behavior, and attainment and the mediating effects of peers and school practices. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

subject-specific learning environments (e.g. for design technology, computing, physical education, and for laboratory-based science lessons).

Finally, for the heuristic device shown in Fig. 14.1 to have utility to researchers as a guide or mapping tool for development, testing, and extension of EER it must be readily adapted to different research settings. Towards this, investigations using the device would first need to appraise the nature of each of the seven levels within the real-world educational setting(s) of their studies. Second, the extent of variation in educational policy and practice within these levels would then need to be appraised (including restrictions on practice and policy due to policies and practices at 'higher' levels; see Sect. 14.3.2 below). Third, an appraisal would then be required of the relevance of each of the existing levels to the study's aims and research questions. For example, in an investigation of teacher effectiveness, if the real-world educational setting(s) focused on multiple teachers teaching in multiple common learning environments, then it may be beneficial for a study to evaluate the effectiveness of the teachers independently from the effectiveness of the learning environments. This followed by an estimation of the effects of this cross-classified nesting of pupils (within both teachers and spaces) on the student outcomes under investigation. This process would constitute an unpicking of the localised manifestation of the multifaceted nature of this level of educational effectiveness.

14.3.2 ...For International Perspectives in Educational Effectiveness Research (EER)

First, international perspectives in EER would benefit from a greater awareness, and critical appraisal, of past, present and likely future contact and collaborations that exist between ILSAs and EER (Fig. 14.1). This would help to limit context stripping of the cultural-specificity (Reynolds et al., 2014) of their substantive findings and limit the extent to which limitations in the capacity for research in these areas to share knowledge are overlooked. For example, the Programme for International Student Assessment (PISA) is an ILSA that is not designed to reveal evidence of effectiveness at the level of the school due to its cross-sectional design (Scheerens, 2016), although, and as illustrated in Chap. 7, it can be used in other ways to investigate EER topics. This greater awareness and critical appraisal has the potential to foster studies involving new areas of educational effectiveness and ineffectiveness at the system level (Level 1 in Table 14.1). Chapter 3[22] describes how evidence at this level is lacking as compared to evidence at the other levels, while

[22]Kyriakides, L., Creemers, B., & Charalambous, E. (n.d.). Developing and testing theories of educational effectiveness addressing the dynamic nature of education. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, Netherlands: Springer

Chap. 7 outlines how ILSAs have the potential for identifying educational effectiveness at this level through the comparison of multiple educational systems.

Of course the challenges in comparing national systems of education to yield evidence of effective and ineffective policy and practice are complex (with an overview provided by Chaps. 6 and 7, and some acknowledgement of recent efforts in this regard in Chap. 2). Further, the comparison of educational systems is not the only means of obtaining evidence of system-level educational effectiveness. Examining changes *within* systems over time can also yield evidence of system effectiveness with this being practically achieved through the use of natural experiments (also noted in Chap. 7). This because policies and practices at a higher level can support or, by contrast, may limit the effectiveness of the levels within them via placing constraints on policy and practice at lower levels. For example, if the policies and practices of a group of schools are shared due to their belonging to a common network then the effectiveness (or otherwise) of these policies and practices will be at the level of the network rather than the school – there may be reduced variation within and between schools so variations in student outcomes at the school level might be reduced. Chapter 12 provides several real-world examples of how change over time (as a basis for revealing effectiveness) can be linked to paradigm shifts at this level (see Fig. 14.1). Not only is culture change at the level of an educational institution (and the levels within) noted as essential for the success of the *Kia Eke Pamuku* ('Building on Success') initiative, but when not enough time was given for this culture change to become embedded in an institution then this limited the impact of the initiative.

Second, international perspectives in EER would benefit from more EER research being carried out in a wider variety of cultures and countries (e.g. the relative paucity in sub-Saharan Africa and mainland China, though there have been notable EER advances in these regions in recent years; see Chaps. 8[23] and 9). Moreover, this research would benefit by being designed and conducted by EER researchers from these cultures and countries (including mixed teams of international and local researchers; see Chap. 6), and by consideration of whether there are models that can be used for the systematic development of EER knowledge bases within cultures and countries where there is not yet a strong tradition of EER. The first two points (the need for wider variety and mixed teams) have been known for decades, hence the founding of ICSEI in the late 1980s, but remain evident today. As a consequence, this final chapter began by locating this volume within the fourth wave of EER research (characterised by internationalisation) first described over half a decade ago (see Sect. 14.1). However, the third point, whether systematic models for the development of EER within cultures and countries might be developed, requires further consideration by the EER community. Chapter 10 presents two historic

[23]Bambanota, G., Van Damme, J., De Fraine, B., Masimango, P. V., Mabela, G. K., Tshite, A. M., . . . Legono, J. P. B. (n.d.). Educational effectiveness research in Africa: The case of the Democratic Republic of the Congo (DRC). In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research.* Dordrecht, Netherlands: Springer

accounts of the development of EER traditions – one focussing on School Effectiveness Research (SER) in the USA and the other Teacher Effectiveness Research (TER) in the Netherlands and Belgium – but as yet it is not possible to judge to what extent either might serve as a model for the systematic development of EER elsewhere. However, common across these two examples is that the development of EER was in response to the educational needs of these countries, involved dialogue between researchers and educational professionals and policy makers from the start, and was undertaken (and continues to be undertaken) while engaging with the international EER community. As potential models for the development of EER in other countries then, might there be merit from building an EER evidence base from either a single level (see Table 14.1), and if so which, or from multiple levels? Further, to what extent should this development be directed and systematised versus given the freedom to develop as circumstances dictate (in other words driven by top-down versus bottom-up pressures)?

Third, international perspectives in EER would benefit from an even greater volume of mixed methods research being carried out, and from educational effectiveness researchers engaging more closely with mixed methods research organisations like the Mixed Methods International Research Association (MMIRA). Why? Because as the preceding sections of this chapter (and Chap. 2) have argued, international perspectives in EER require research findings that take into account the cultures and paradigms present at each of the levels of EER being investigated, which requires a combination of rich and thick description as well as associations between educational factors and student outcomes. Mixed methods can offer this insight via a wide variety of research designs that can incorporate the tradition of EER for large-scale quantitative research with qualitative research that is strongly resistant to the context-stripping that must be avoided in order to obtain accurate and credible findings and develop accurate and credible theories of education[24]. Thus, mixed methods Education Effectiveness Research (MMEER) may prove to be better able to engage with ILSA evidence in the formulation of new theories of the effectiveness of educational systems, and new fruitful EER enquiries within and across different country contexts, than either quantitative or qualitative approaches on their own. This due to the creative ways in which quantitative and qualitative approaches, data, analysis and findings can be combined and synthesised plus the potential for MMEER to buffer against context stripping, as evidenced by some of the previous efforts in this regard (e.g. Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002). As mentioned earlier in this chapter though, mixed methods research is not a panacea here. This approach to research comes with its own challenges as regards the practicalities of merging research methodologies and bridging epistemological positions (for details concerning EER see Teddlie & Sammons, 2010; Reynolds et al., 2014; Sammons et al., 2014; Sammons & Davis, 2017).

[24]For further discussion of the potential of MM designs and studies in EER see Sammons (2010), Teddlie and Sammons (2010), and Sammons and Davis (2017).

However, while evidence from EER and ILSAs has the potential to inform one another (bearing in mind the dangers of context stripping), a common risk remains: misunderstanding and potential misuse of evidence by education policy makers. Of course research evidence must not be cherry-picked, misinterpreted (wilfully or otherwise), or supressed (e.g. Torjesen, 2016; Hall, Lindorff, & Sammons, 2016) to produce overstatements, overgeneralisations, or simplifications (see Chap. 7), but questions remain concerning how such mistreatment and misuse can be minimised as EER engages in its core business of attempting to support evidence-based change to educational policy and practice (e.g. improving schools). With reference to Chap. 4 and to Anton Chekov, to what extent (and how) can EER bark the language of educational policy without being obliged to also wag its tail? For example, there is a common preference among contemporary policy makers for intervention-based evidence from experimental research (e.g. Haynes, Goldacre, & Torgerson, 2012) – this despite the limitations of such knowledge for understanding educational systems and processes being documented nearly 100 years ago (see James, 2013). While RCTs and quasi-experimental designs (QEDs) can prove useful – especially in evaluating specific interventions (e.g. the benefits of receiving pre-school or not in specific contexts such as the well documented Head Start programme in the US) – they cannot be used and are not useful for describing, investigating and understanding the ecological nature that underpins most educational systems and the naturally occurring variation in educational provision, institutional effectiveness, and student experiences that often reflect historic and cultural contexts and preferences. One way that experiments might prove fruitful is in evaluating the success of particular educational initiatives within specific school contexts before large-scale rollout (Muijs et al., 2014; compare also efficacy vs. effectiveness, e.g. Lindorff, Hall, & Sammons, 2019; Hall, Sammons, et al., 2016).

Fourth, the importance of methodological advances for the substantive findings of EER (both directly and indirectly via EER's reciprocal relationship with ILSAs; see Fig. 14.1) has been recognised (Creemers et al., 2010) but poses a risk to development of the field's substantive knowledge base. This with the consequence that the research methodologies of EER and ILSAs require continual critical appraisal (as Chaps. 6 and 7 do for ILSAs and for EER as is done by papers such as Reynolds et al., 2014). This need for continual critical appraisal of research methods is because of their potential to direct but also to constrain substantive findings. In short, the tools that we use may limit the knowledge that we can create. One contemporary example of this comes from the greater use of mixed methods approaches (rather than just quantitative) to understand the processes of educational effectiveness discussed above. Another example is our recent critique of the often taken for granted assumption that two well-known quantitative concepts (Moderation and Mediation) are mutually exclusive. The result was a reconsideration of methodological assumptions and a new side-to-side contextual awareness of methods and theories from fields close to EER (see Fig. 14.2). In turn, these yielded the specification of a new quantitative concept ('Airbag Moderation') that more faithfully represents a number of educational processes of interest to EER. These include: the impact of school selection effects for differential educational

effectiveness; the impact of educational policies and practices designed to facilitate equity in student progress (e.g. via the use of interventions that target); and the triggering of complacency after achievement deemed successful (see Hall et al., 2020).

Fifth, as the international perspective in EER continues to develop it prompts the need for greater reflection on how the notions of educational equity and inequity are being used within EER – both conceptually and when operationalised within research investigations (see Sect. 14.3.1 & Reynolds et al., 2014). This is because EER has long had a commitment to promoting *social* equity (reducing the equity gap in achievement or other outcomes between low and high SES students for example) and because international perspectives broaden the variety of *social* contexts within which studies of EER are carried out. Thus, operationalisations and conceptualisations of equity and inequity in EER are inclined to multiply as the field's international perspective grows. In turn, this prompts the need for conceptual discussions of the relationship between EER and educational equity to reflect on the accuracy to the increasing (and increasingly varied) operationalisations of equity that are being used. Of course, it is not just variation in researchers' concepts and operationalisations of equity that will diversify, but also the concepts and operationalisations of equity that exist within the systems, networks, institutions, groups, and individuals that feature in studies in EER. Further, such variations at the levels of research and research participants are exacerbated by the broadening remit of student outcomes that are being considered by EER and by educational policies and practices around the world (e.g. student wellbeing in the educational policy of mainland China; see Chap. 9). Thus, new areas of research for EER have the potential to emerge as a result of these simultaneous drivers acting in concord. For example, effective education for equity as related to LGBT+ and SEND, plus new theories of educational effectiveness which reflect upon the tensions that can exist between educational equity and educational equality (e.g. Espinoza, 2007; Ford, 2015).

The drive for future investigations of EER that explore new areas/types of educational inequity can also be informed by ILSAs producing research that better documents these inequalities (for examples see Chaps. 7 and 12). EER can then respond to these results with new findings and theories concerning both educational effectiveness and differential educational effectiveness (Mortimore et al., 1988; Nuttall, Goldstein, Prosser, & Rasbash, 1989; Creemers et al., 2010) – in other words, via studies to identify educational policies and practices that can either exacerbate or ameliorate the educational disadvantages, inequities, and inequalities that are revealed from ILSAs. However, while ILSAs may indeed prove useful to EER in this regard, it also needs to be remembered that narratives of educational equity within EER are bound to be limited to the inequities and inequalities of the locations within which such research takes place and also constrained by the various boundaries that limit where (and in what form) EER will take place in the future. For example, in locations that are characterised by high levels of social disadvantage, studying educational effectiveness and inequity can involve differentiating not the *extent* of the exposure to disadvantage, but the *frequency* of this exposure across the

environments that surround the student (e.g. in the home, neighbourhood, and school; e.g. Herrero Romero, Hall, Cluver, Meinck, & Hinde, 2018; Herrero Romero, Hall, Cluver, & Meinck, 2019).

Sixth, to what extent do the themes within this book have unique implications for the developing international perspective in school improvement and School Improvement Research (SIR)? In a direct sense, Chap. 2 notes how EER knowledge and EER researchers have contributed to the development of international school improvement research projects (e.g. the Effective School Improvement project; Reezigt & Creemers, 2005). Similarly, Chap. 3 highlights that the dynamic model is intended to contribute EER insight towards educational improvement, and calls for more international research to evaluate approaches to educational improvement informed by the dynamic model (e.g. the Dynamic Approach to School Improvement; Creemers & Kyriakides, 2012). The Rawlsian philosophical orientation proposed in Chap. 4 is discussed with regard to measuring and monitoring equity, but the same principles could be considered to inform educational improvement and reform efforts intended to promote equity and to underpin the evaluation of such initiatives. Chapter 5 highlights similarities between inter-school networks (or inter-school collaborations) and individual schools in terms of the features found to promote improvement, with a particular emphasis on the important role of processes over structures. The mixed results presented in this chapter underscore the importance of investigating educational improvement initiatives (within and beyond the level of the school) across a diverse range of contexts and approaches. Chapter 6 underscores the importance of context-specificity rather than one-size-fits-all prescriptions of "what works", while international findings from trends at the country level in PISA reported in Chap. 7 suggest some commonality across countries in the association between using assessment data for school self-evaluation and improvement in student outcomes.

Later chapters offer reflections on some of the context-specific conditions of education within particular countries, with implications for educational improvement efforts and research in those areas. Chapter 12 offers a particularly direct insight into international SIR with a focus on the role of the principal, offering specific insights from case studies in Australia and New Zealand but also highlighting some of the common findings across countries from the International Successful School Principalship project with regard to the complex relationships between principals and the contexts in which they work. Chapter 8 contextualises the results of its particular EER study with a detailed account of the circumstances of teachers and schools in the Democratic Republic of the Congo, ranging from fundamental infrastructure problems to teacher training and compensation (i.e. a large proportion of teachers working without pay over long periods of time). These conditions have critical implications for improvement initiatives, calling to mind Mingat's (2007) account of the tensions between different priorities for improving education in low-income countries.

Chapter 9 reflects on three key approaches to optimise educational improvement from a review of the international research evidence base (Reynolds et al., 2016):

1. The need for a renewed focus on the implications of EER approaches and its knowledge base for professional development
2. The need for more ownership and rigorous localised evidence-based research
3. More joined-up public service provision

However, Chap. 9 also highlights a priority within mainland China that is not within these three key approaches and priorities: The need for rigorous systems of educational evaluation and assessment. Thus, we see that there is a need for two areas of simultaneous reflection and awareness: Of key priorities and approaches, *plus* awareness of localised variation that can add, subtract, and/or alter the meaning of these priorities and approaches. In turn, these processes of reflection also require contextual awareness of historical precedence and current paradigms in localised education, educational research, policymaking, and professional practice. For example, the possibility that there may be aspects of teachers' professional learning which may be unique to mainland China (e.g. teacher research groups).

The comprehensive and detailed account of school and classroom factors associated with student outcomes in Chap. 10 may be useful to inform evidence-based improvement initiatives in the context of Belgium and the Netherlands. Chapter 11 similarly provides implications for educational policy and reform in the USA, and Chap. 13 suggests implications of using 'leading indicators' assessed at the start of kindergarten to inform efforts to improve outcomes and reduce inequalities. Educational improvement policies and initiatives based on these findings would benefit from investigation drawing on a school improvement perspective, attending to the processes of implementation as well as the perspectives of stakeholders and the nuances of local context. As Chapman, Ainscow, and Hadfield (2020) note, the relationships between researchers, practitioners and policymakers are essential to such efforts, and "managing these relationships is crucial to the success of attempts to use research knowledge to guide the improvement of policy and practice in the field". The insights from the chapters of this book serve to illustrate the reciprocal relevance of the EER and SIR knowledge bases; across all chapters, the implications drawn from diverse contexts and approaches share an orientation towards educational improvement in the interest of greater quality and equity.

Last but by no means least, the International Congress for School Effectiveness and Improvement (ICSEI; https://www.icsei.net/) recently celebrated its 30th anniversary and over the past 30 years, it has served a leading role in EER's developing international perspective. ICSEI has expanded from a relatively small organisation to one with a large international following as evidenced by the large numbers and wider diversity of participants from different international contexts now attending its annual conferences and other events (as noted in Chap. 2). The conference deliberately meets in countries in different continents to promote its international presence and to foster links between both EER and school improvement research. Its ICSEI global networks (https://www.icsei.net/networks/) cover five areas: Data Use, Early Childhood Education, Educational Leadership, Methods of Researching Educational Effectiveness and Improvement (MOREI), Professional Learning, and Policy makers, Politicians and Practitioners (3P). This growth of ICSEI over time, its

meetings around the world, and its remit that is inclusive of substantive and methodological areas plus inter-profession engagement all speak to the continuing prominence of the organisation for future international perspectives in EER.

14.4 Concluding Thoughts

The contents of this book were written at a time when intense worldwide discussions were ongoing as regards the raising of new walls between people and cultures, both figurative and literal, but all with the aim of forcible separation, control, and the stymying of interaction. There are also increasing concerns about fake news and the misuse of 'big data' to misinform and to promote divisions. Respect for rigorous analysis, well-founded research and the use of evidence seems increasingly low in an era when some powerful policy makers and politicians seem to prefer easy narratives, ideologically-driven and (supposedly) simple 'solutions', and disparage 'experts' whether in relation to education or other areas of public concern including climate change and civil liberties. It is in the context of this uncertain future that there is therefore an *increasing* need to develop, maintain, and communicate an international perspective based on well-founded research and collaborations to promote educational policies and practices that foster more equitable student outcomes for all. As signposted in Chap. 2, the knowledge base that has been developed over the last 50+ years of EER has generated important insights but much more is called for to help study, identify and share findings that will help us to distinguish those policies and practices that are more effective from those and those that are less effective (or worse ineffective), along with under what conditions, towards what outcomes, and for whom. Of course, the marginalised groups in our societies will, by definition, continue to be at-risk of underrepresentation in these perspectives, but with new walls come the potential for new groups to be marginalised and for existing inequities to be exacerbated. Given that the roots of educational effectiveness lie not just in simplistic conceptions of 'what works' but also in the desire for greater educational equity (e.g. Rutter et al., 1979), so our movement towards a more comprehensive and inclusive international perspective in EER will require us to acknowledge and understand the educational contexts and inequities experienced by these emerging marginalised groups. This book and this concluding chapter provide a snapshot of where EER currently stands in its continuing journey towards an increasingly international perspective. It also attempts to identify and illuminate potential future challenges and how these may be overcome as we continue to engage across professional and disciplinary boundaries to foster evidence-based educational policy and practice and promote both effectiveness and equity in diverse global contexts.

References

Ainscow, M. (2016). Collaboration as a strategy for promoting equity in education: Possibilities and barriers. *Journal for Professional Capital and Community, 1*(2), 159–172. https://doi.org/10.1108/JPCC-12-2015-0013

Ainscow, M., Dyson, A., Goldrick, S., & West, M. (2012). Making schools effective for all: Rethinking the task. *School Leadership & Management, 32*(3), 197–213. https://doi.org/10.1080/13632434.2012.669648

Bambanota, G., Van Damme, J., De Fraine, B., Masimango, P. V., Mabela, G. K., Tshite, A. M., ... Legono, J. P. B. (n.d.). Educational effectiveness research in Africa: The case of the Democratic Republic of the Congo (DRC). In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Best, S. (1994). Chapter 2: Foucault, postmodernism, and social theory. In D. R. Dickens & A. Fontana (Eds.), *Postmodernism and social inquiry* (pp. 25–52). New York, NY: Guilford Press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology, vol. 1: Theoretical models of human development* (6th ed., pp. 793–828). New York, NY: Wiley.

Chapman, C., Ainscow, M., & Hadfield, M. (2020). *Changing education systems: A research-based approach*. London, UK: Routledge.

Chapman, C., Lowden, K., Chestnutt, H., Hall, S., McKinney, S., & Friel, N. (2016). *The school improvement partnership Programme: Sustainable collaboration and enquiry to tackle educational inequity* (Final report). Glasgow, UK: University of Glasgow & Education Scotland.

Creemers, B. P. M. (1994). *The effective classroom*. London, UK: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347–366. https://doi.org/10.1080/1743727X.2020.173533410.1080/09243450600697242

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London, UK/New York, NY: Routledge.

Creemers, B. P. M., & Kyriakides, L. (2012). *Improving quality in education: Dynamic approaches to school improvement*. London, UK/New York, NY: Routledge.

Creemers, B. P. M., Kyriakides, L., & Antoniou, P. (2013). A dynamic approach to school improvement: Main features and impact. *School Leadership & Management, 33*(2), 114–132. https://doi.org/10.1080/13632434.2013.773883

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in school effectiveness research*. London, UK: Routledge.

Espinoza, O. (2007). Solving the equity-equality conceptual dilemma: A new model for analysis of the educational process. *Educational Researcher, 49*(4), 343–363. https://doi.org/10.1080/00131880701717198

Ford, D. Y. (2015). Multicultural issues recruiting and retaining Black and Hispanic students in gifted education: Equality vs. equity schools. *Gifted Child Today, 38*, 167–170. https://doi.org/10.1177/1076217515583745

Foucault, M. (1973). *The order of things: An archaeology of the human sciences*. New York, NY: Vintage Books.

Hall, J., Khalil, L., Kelly, A., & Galbraith, D. (2019). *An upstream evaluation of the prototype coaching e-books that will form part of Oxford Reading Buddy: Findings from a mixed methods randomized control trial*. Oxford, UK: Oxford University Press.

Hall, J., & Kreppner, J. (Eds.). (2019). How can education better support the mental health & wellbeing of young people? Contributions from developmental psychopathology & educational effectiveness research. *Lausanne, Switzerland: Frontiers Media*. https://doi.org/10.3389/978-2-88963-146-9

Hall, J., Sammons, P., Sylva, K., Evangelou, M., Eisenstadt, N., Smith, T., & Smith, G. (2016). Disadvantaged families are at greatest risk from austerity cuts to Children's Centres. *British Medical Journal, 352*, i897. https://doi.org/10.1136/bmj.i897

Hall, J., Malmberg, L-E., Lindorff, A., Baumann, N., & Sammons, P. (2020). Airbag moderation: The definition and implementation of a new methodological model. International Journal of Research & Method in Education [online first]. https://doi.org/10.1080/1743727X.2020.1735334

Hall, J., & Sammons, P. (2013). Mediation, moderation, & interaction: Definitions, discrimination & (some) means of testing. In T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 267–286). Rotterdam, The Netherlands: Sense Publishers. https://doi.org/10.1007/978-94-6209-404-8_13

Hall, J., Sylva, K., Sammons, P., Melhuish, E., Siraj-Blatchford, I., & Taggart, B. (2013). Can preschool protect young children's cognitive and social development? Variation by center quality and duration of attendance. *School Effectiveness and School Improvement, 24*(2), 155–176. https://doi.org/10.1080/09243453.2012.749793

Haynes, L., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: Developing public policy with randomized controlled trials*. London, UK: Cabinet Office-Behavioural Insights Team.

Herrero Romero, R., Hall, J., Cluver, L., & Meinck, F. (2019). Exposure to violence, teacher support and school delay amongst adolescents in South Africa. *British Journal of Educational Psychology, 89*(1), 1–21. https://doi.org/10.1111/bjep.12212

Herrero Romero, R., Hall, J., Cluver, L., Meinck, F., & Hinde, E. (2018). How does exposure to violence affect school delay and academic motivation for adolescents living in socioeconomically disadvantaged communities in South Africa? *Journal of Interpersonal Violence [online first]*. https://doi.org/10.1177/0886260518779597

Hopkins, D., Stringfield, S., Harris, A., Stoll, L., & Mackay, T. (2014). School and system improvement: A narrative state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 257–281. https://doi.org/10.1080/09243453.2014.885452

James, M. (2013). *New (or not new) directions in evidence-based practice in education.* Online: www.bera.ac.uk/wp-content/uploads/2014/02/Mary-james-New-or-not-new-directions-in-evidence-based-policy.-Response-to-Ben-Goldacre.pdf. Accessed 12 May 2019.

Jensen, B., Sandoval-Hernández, A., Knoll, S., & Gonzalez, E. J. (2012). *The experience of new teachers: Results from TALIS 2008*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264120952-en

Jude, N., & Kuger, S. (2018). *Questionnaire development and design for international large-scale assessments (ILSA)*. Washington, DC: National Academy of Education.

Kelly, A. (2012). Sen and the art of educational maintenance: Evidencing a capability, as opposed to an effectiveness, approach to schooling. *Cambridge Journal of Education, 42*(3), 283–296. https://doi.org/10.1080/0305764X.2012.706255

Kelly, A. (n.d.). The fifth phase of educational effectiveness research: The philosophy and measurement of equity. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Klieme, E. (n.d.). Policies and practices of assessment: A showcase for the use (and misuse) of international large scale assessments in educational effectiveness research. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Kyriakides, L., Creemers, B., & Charalambous, E. (n.d.). Developing and testing theories of educational effectiveness addressing the dynamic nature of education. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Lindorff, A., Hall, J., & Sammons, P. (2019). Investigating a Singapore-based mathematics textbook and teaching approach in classrooms in England. *Frontiers in STEM Education*. https://doi.org/10.3389/feduc.2019.00037

Lindorff, A., Sammons, P., & Hall, J. (n.d.). International perspectives in educational effectiveness research: A historical overview. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Lopez, A. Y., & Willms, J. D. (n.d.). A national evaluation of kindergarten outcomes: Findings from Uruguay. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Mingat, A. (2007). Cost and financing of education and its impact on coverage and quality of services and efficiency and equity in Sub-Saharan African countries. In T. Townsend (Ed.), *International handbook of school effectiveness and school improvement*. Dordrecht, The Netherlands: Springer.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. London, UK: Open Books.

Muijs. (n.d.). Extending educational effectiveness: The middle tier and network effectiveness. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands, Springer.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperly, H., & Earl, L. (2014). State of the art – Teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231–256. https://doi.org/10.1080/09243453.2014.885451

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 1/2? *Developmental Psychology, 39*(3), 451–469. https://doi.org/10.1037/0012-1649.39.3.451

Norwich, B. (2014). How does the capability approach address current issues in special educational needs, disability and inclusive education field? *Journal of Research in Special Educational Needs, 14*(1), 16–21. https://doi.org/10.1111/1471-3802.12012

Nuttall, D. L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research, 13*(7), 769–776. https://doi.org/10.1016/0883-0355(89)90027-X

Opdenakker, M. -C. (n.d.). Three decades of educational effectiveness research in Belgium and the Netherlands: Key studies, main research topics and findings. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Palardy, G. J. (n.d.). The impact of socioeconomic segregation in U.S. high schools on achievement, behavior, and attainment and the mediating effects of peers and school practices. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Palardy, G. J., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 111–140. https://doi.org/10.3102/0162373708317680

Reezigt, G., & Creemers, B. (2005). A comprehensive framework for effective school improvement. *School Effectiveness and School Improvement, 16*(4), 407–424. https://doi.org/10.1080/09243450500235200

Reynolds, D., Chapman, C., Clarke, D., Muijs, D., Sammons, P., & Teddlie, C. (2016). Chapter 17: Conclusions. The future of educational effectiveness and improvement research, and some suggestions and speculations. In C. Chapman, D. Muijs, D. Reynolds, P. Sammons, & C. Teddlie (Eds.), *The international handbook of educational effectiveness research: Research, policy, and practice* (pp. 408–439). New York, NY: Routledge.

Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (2002). *World class schools: International perspectives on school effectiveness*. London, UK: Routledge.

Reynolds, D., Kelly, A., Harris, A., Jones, M., Adams, D., Miao, Z., & Bokhove, C. (n.d.). Extending educational effectiveness: A critical review of research approaches in international effectiveness research, and proposals to improve them. In J. Hall, A. Lindorff, & P. Sammons

(Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197–230. https://doi.org/10.1080/09243453.2014.885450

Rutter, M. (2006). Implications of resilience concepts for scientific understanding. *Annals of the New York Academy of Sciences, 1094*(1), 1–12. https://doi.org/10.1196/annals.1376.002

Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London, UK: Open Books.

Rutter, M., & Rutter, M. (1992). *Developing minds: Challenge and continuity across the life span*. New York, NY: Penguin Books.

Sammons, P. (1996). Complexities in the judgement of school effectiveness. *Educational Research and Evaluation, 2*(2), 113–149. https://doi.org/10.1080/1380361960020201

Sammons, P. (2010). Chapter 27: The contribution of mixed methods to recent research on educational effectiveness. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods research*. London, UK: Sage.

Sammons, P., & Davis, S. (2017). Chapter 23: Mixed methods approaches and their application in educational research. In D. Wyse, N. Selwyn, E. Smith, & L. Suter (Eds.), *The BERA/SAGE handbook of educational research*. London, UK: BERA/SAGE.

Sammons, P., Davis, S., Day, C., & Gu, Q. (2014). Using mixed methods to investigate school improvement and the role of leadership: An example of a longitudinal study in England. *Journal of Educational Administration, 52*(5), 565–589. https://doi.org/10.1108/JEA-10-2013-0121

Scheerens, J. (2016). *Educational effectiveness and ineffectiveness: A critical review of the knowledge base*. Dordrecht, The Netherlands: Springer.

Schoon, I. (2006). *Risk and resilience: Adaptations in changing times*. Cambridge, UK: Cambridge University Press.

Sen, A. (1992). *Inequality re-examined*. Oxford, UK: Clarendon Press.

Sen, A. (1999). *Development as freedom*. New York, NY: Alfred A. Knopf.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *EPPE technical paper 12: Effective provision of pre-school education, the final Report*. London, UK: DfES/Institute of Education.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. London, UK: Falmer Press.

Teddlie, C., & Sammons, P. (2010). Chapter 7: Applications of mixed methods to the field of educational effectiveness research. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodological advances in educational effectiveness research*. London, UK: Routledge..

Torjesen, I. (2016). Austerity cuts are eroding benefit of sure start children's centres. *British Medical Journal*, 352–335. https://doi.org/10.1136/bmj.i335

Thomas, S. M. (n.d.). School and teacher value added performance and the relationship with teacher professional development in mainland China. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Townsend, T., Berryman, M., Gurr, D., & Drysdale, L. (n.d.). Leadership for Learning in diverse settings: School leaders setting the agenda in Australia and New Zealand. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International perspectives in educational effectiveness research*. Dordrecht, The Netherlands: Springer.

Index

A
Ability, 44, 52, 72, 80–81, 84–85, 94, 192, 210, 213, 258, 261, 363, 368, 377
  cognitive, 242–243, 249, 258, 268, 273
  individual, 258
  intellectual, 80
  interpersonal, 213
  lower, 364
  natural, 80, 81
  sets, 390
Aboriginal Education Workers (AEWs), 340
Absenteeism, pupil, 191
Academic achievement, 4, 135, 233, 260, 265, 290, 293, 298, 308, 309, 316, 319, 365
  academic talent, 84
Academic background, 307
Academic emphasis, 259
Academic engagement, 238–239, 244, 246, 263, 266, 274
Academic orientation, 295, 316
Academic performance, *see* Academic achievement
Academic press, 320
Academic skills, 292, 295, 297, 315
Academies, 95, 102, 105, 110
Academies and Free Schools, 102
Academisation, 150
Academy
  chains, 110
  Federation, 110
Accountability, 76, 86, 128, 151, 154–156, 171, 173, 225, 377
  activities, 167
  context, 233
  internal, 103
  perspective, 152
  policies, 128, 155, 168, 172, 175, 176
  practices, 154
  provision, 292
  purposes, 152
  soft, 168, 171, 174, 177
  strong, 152, 168, 172, 176
Accountability rules, 165
Accountability systems, 133, 154
Achievement, 115, 121, 132, 134, 137, 140, 155, 171, 188, 190, 200, 204
  average, 161, 233, 259, 261, 372
  cognitive level, 193, 254, 257, 273
  country, 35, 101, 136, 155, 167, 176, 400
  growth, 257, 292
  high, 255, 259, 265, 273
  low, 93, 172, 344
  non-educational determinants, 133
  prior, 133, 216, 222
  school level, 168, 171
  tests, 124, 134, 234, 288, 292, 307, 314
Achievement data
  use for accountability, 156, 167
Achievement gaps, 290, 291, 344
  racial, 290
  reducing, 291
  socioeconomic, 290
Achievement growth, country-level, 177
Achievement motivation, 246, 255, 258, 273
Achievement test scores, 134, 288, 292, 307
Achievement tests
  common, 234
  standardized, 314
Across-context 'lists'/across-context lists, 139

J. Hall et al. (eds.), *International Perspectives in Educational Effectiveness Research*,
https://doi.org/10.1007/978-3-030-44810-3

Act Utilitarianism, 77
Action plans, 52–53, 60, 348
  individualized, 347
  strategic, 348
Active learning, 51, 265
Activities, 169
  application, 59
  choice, 254
  collaborative, 104
  concrete, 42
  curriculum-related, 51
  defining, 42
  extra-curricular, 56, 150
  foundational, 341
  implementing core, 171
  instructional, 135
  interest-based, 254
  network, 114
  shared, 340
  structuring, 50
Adaptation, 43, 84, 148
Adaptive behaviour, 265, 267
Adjustment, psychosocial, 244
Administration, 124, 213, 329, 352, 386
Administrators, *see* Administration
Admissions, 295, 308, 315
Adolescents, 187
  delinquent, 210
Advancement
  EER methodology, 14
  social, 84
Advantaged pupils, 81
Advocacy, 76, 332
Africa, French-speaking, 185
African Americans, 290, 293, 300
African countries, 17, 22, 185–204
After-school programmes, 150
Age, 4, 38, 43, 46, 80, 124, 129, 187, 194, 235, 247–248, 293, 344, 361–362, 365–372, 377, 393, 394
  children's/children, 365, 368
  effect, 366, 372
  slopes, 370
  variable, 372
Agency
  professional, 348
  leaders'/leaders, 348, 350
Aggregation, 75–76, 78, 84, 94, 95
Airbag Moderation, 398
Amartya Sen, 85
Ambition, 136, 214
American Educational Research Association (AERA), 14, 234, 291
Analyses
  Aptitude-Treatment-Interaction approach, 188, 234
  comparative, 155
  cost-benefit, 135
  cross-cultural, 122
  exploratory, 202
  logistic regression, 367
  meta, 307
  modelling approaches, 188–189
  qualitative, 107, 111
  quantitative, 107, 110, 119, 175
  regression, 149
  school-level, 170
  three-level multilevel modelling, 215
Analytical modelling approaches, 185, 204
Anchors, 37, 148
Anglo-Saxon societies, 138
Annual growth, 370, 374–377
Anti-social behaviour, 248
APPA, *see* Australian Primary Principals Association (APPA)
Aptitude, 36, 41, 47–48
Archangel, 77–78
Armed conflict, DRC, 186
Asia, 19
Aspirations, 71, 79, 293, 295, 346
Assertiveness, 332
Assessment, 43, 52, 58, 93, 125, 128, 147, 150–157, 164, 171, 174, 176, 194, 212–214, 368, 372, 376, 377
  approaches, 212
  classroom-based, 173
  cross-cultural, 127
  data, 175, 333, 400
  designing
    challenges, 20
  educational, 151
  formative, 155, 168, 176
    effects, 153
  framework, 362
  informal
    processes, 338
  items, 129
  measures, 215
  methods, 212
  mode, 124
  national, 114, 155, 330, 365
  nationwide, 4
  peer, 214
  policies, 3, 151, 154–155, 176, 177
  practices, 3, 339
  pupil, 195

purposes, 165
quantitative, 213
-related questions, 156
reliable, 5, 376
results, 156, 174
public availability, 152
strategies, 173
summative, 153, 172, 176
test scores, 128
time pressure, 124
tool, 367
use, 164, 174
Assessment and evaluation
feedback, 151
Assistant principal/s, 333
At-least-the-trains-run-on-time mentality, 86
Atmosphere, orderly, 233, 255, 261
At-risk students, 9, 269, 275
Attainment, 72–73, 80, 85–87, 91, 101, 104–105, 107, 108, 110, 115–117, 119
analysis, 115
changes, 115
data, 107–108
equity, 91, 95
gap, 74–75
high, 86
improved, 115
prior, 71, 90, 106, 111, 214, 217–219, 223
raising, 110, 119
Attainment Equity Index, 90–93
Attendance, 94, 186, 187, 254, 291, 294, 362, 366, 376
funded compulsory, 209
pre-primary, 365
rates, 292
regular, 314
Attitudes
negative, 259
of school personnel, 295
parental, 131
positive, 259
Australasia, 400
Australasia, Australia and New Zealand, 4, 10, 164–165, 169, 176, 329–331, 335, 338–339, 343, 350, 351
Australian Primary Principals Association (APPA), 335, 338
Autonomy, 211, 254, 264, 265, 274
Autonomy support, 254, 267, 273
Autonomy, leadership and management, 211
Average student performance, 132, *see* Achievement, average
Average Utilitarianism, 75
Average utility, 75
Avoidance orientation, 248
Awareness
ideological, 214
phonological, 336, 363

B
Baby rooms, 394
Background effects, 132
Background factors, 46–47, 215
differentiated effects, 46
social, 132
socio-economic, 133
Background questionnaires, 148, 193
Background variables, 45–47, 111
critical, 48
economic, 44–46
Baseline, 80, 106, 304, 305, 367
model effect size, 313
Basic education, 188
Bayesian modeling, 175
Before-and-after comparisons, 388
Behavior/attitude, 49, 287, 295, 301, 316, 332
Behavioral engagement, 294, 297, 306, 309, 316, 319
Behavioral expectations, 301
Behavioral infractions, 295
Behaviors
observable instructional, 49
on-task, 51
teacher's interpersonal, 51
Beijing (Mainland China), 3, 159, 211, 214
Belgium, 4, 232–237, 239–248, 250, 258–261, 267, 272, 274, 276, 393, 397, 401
educational track offerings, 244
schools, 242
Beliefs, 79, 211, 214, 331–332, 347, 350
self-efficacy, 48
Benefit Principle, 95
Benefit, aggregate, 78, 95
Bentham, 75–77
Between-country variance, 165
Between-school, 86, 90
components, 89–90
performance, 95
Bias, 78, 95, 302
confirmation, 73
conscious, 353
cultural, 130
downward, 303
natural, 77
personal, 77
potential, 303
Birth month, 4
Birth weight, 46

Black box (of school processes, 327
Black people (segregation of), 288
Black slavery, 288
Bolivia, 364
Boundaries, school catchment, 288
British Education Research Association (BERA), 14
Brown v. Board of Education, 288, 293, 318
Bullying, 78

C
Campbell's Law/Campbells Law, 293
Canada, 162, 167, 330, 367
Canada searching, 58
Capabilities, 85–86, 341
  in-school, 347
Capability theory, 85
Capacity, 105, 111, 117, 118, 123, 126, 133, 341, 350
  general cognitive, 124
  individual, 333
  local, 17
  researcher, 16
Capacity building, 25, 331, 335, 352
Case studies of schools, 338
Catchment areas, 124
Categorical outcomes, 303
Catholic school advantage hypothesis, 250
Catholic schools, 249–251
  school leaders, 250
Causal attribution, 134
Causal claims, 173, 175
Causal relationships, 71
Causality, 149, 171, 176
  reversed, 171
Center for Educational Effectiveness and Evaluation, 186, 200
Central Africa, 186, 200–202, 204
Centralized exams, 155
Centre for Educational Effectiveness and Evaluation of Katholieke Universiteit Leuven (Belgium), 186, 193, 200
Certification, 151, 157, 165, 172
Chains of schools, 95
  'no excuses', 103
Challenges
  educational, 24, 187, 289, 290
  research, 15, 19–22, 94, 175, 396, 402
  social, 244
Change interventions/practices, 351
Characteristics, student background, *see* Pupil characteristics, background
Charter Management Organisations (CMOs), 103
Charter schools, 102
  no excuses, 103
Child development, 367, 389
Children
  developing, 292
  disadvantaged, 46, 81, 364
  low-achieving, 254
  supporting reading, 341
Children's achievement, 365
Children's skills, 5, 361, 364, 366, 367, 377
Chile, 17, 20, 168, 365
China, 3–4, 20, 23, 83, 86, 127, 129, 209–212, 214–219, 223–226, 330, 385, 396, 399, 401
  context, 161
  education system, 210
  hukou, 129
  school effectiveness research, 215
  schools, 217, 220
Chinese Ministry of Education (CMOE), 210, 212–215, 225
Chinese National Institute for Education Sciences, 3, 216
Chinese-speaking countries, 161
Citizenship, 214
Citizenship competences, 244
Citizenship, orientations, 258
Civic functions of schools, 292
Civil liberties, 402
Civil rights (USA);civil rights (USA, 289
Civil Rights Act, 288, 289
Civil rights, reform (USA), 318
Civil War (USA), 288
Class
  average cognitive ability level, 268
  mean cognitive ability/achievement, 267
  size, 134, 194, 196, 275
    restricting, 188
  teaching, 211, 224
  time, 265
  variables, 196
Class characteristics, 243, 268
  general, 194
  practice, 267
Class climate, 193, 194, 243, 265, 266
  effects of, 241, 261, 265
Class composition, 196, 199, 242, 243, 249, 258, 267–269, 274, 276
  characteristics, 271
  effects of, 268, 269
Class configurations, 270
Class context, variables, 241

Class effectiveness, 394
Class effects, 190
Class environments, 235
Class factors, 240
Class group, 266, 268, 269, 273
  composition, 268
  high cognitive ability, 273
Class level, 193–194, 232, 246, 248–249, 261–274, 276
Class membership, 2
Class practices, 245, 263, 272
  effects of, 263
  favourable, 253, 272
Class process, 243, 260, 267, 274
  factors, 243
  variables, 267
Class(es), 4
  disadvantaged, 269
  heterogeneous / mixed ability, 222, 267
  high-ability, 267
  high-SES, 273, 311
  learning and relational climate in, 241
  learning climate, 267
  low-SES, 269
  math, 246
  overcrowded, 187
  primary, 245, 246, 263
  secondary school, 246
  secondary school prevocational, 246
  selected, 190
  ungraded, 377
  well-integrated, 275
  within-class slope, 372
Classmates, 234, 275
Classroom
  assessment, 49, 155
  average effect, 374
  behaviour, 72
  climate, 51
  data
    absence of, 131
  differences, 140
  disorder, 51
  disruptions, 51
  disturbances, 302
  effective management, 274
  effectiveness, 353
  effectiveness factors, 131
  effects, 138, 362, 370, 374
  environment research, 51
  environments, 150
  factors, 16, 18, 38, 40, 48, 401
    school level influence, 52
  heterogeneous, 377
  instruction, 233, 367
  learning, 153
  learning environment, 50, 54
  level, 16, 23, 34, 37, 49, 52, 59, 101, 125, 155, 160, 215, 378
  management, 262, 263
  observation, 35, 224
  practice, 5, 78, 156, 160, 163, 172, 339, 347
    effective, 21, 176
  routines, 339
  SES, 4
  studies, 232
  teaching, 72, 152, 153, 171
  variables, 21
  within-classroom average, 372
Classroom Environment Study, 123
Classroom-school interface, 233
Clear communication, 104, 262
Climate, 235, 252, 254, 256, 263, 270, 272, 273, 275, 329
  communitarian, 251
  favourable, 270
  positive, 256, 339
  social, 249
  study-oriented, 275
Climate characteristics, 252, 267
Climate factors, 256, 271
Climate indicators, 273
Climate variables, 273
Closing the Gap program, 335
Clusters (of schools), 105, 394
Clusters (of schools);clusters (of schools, 271
CMOs, *see* Charter Management Organisations
Coefficient of Variation (CoV), 88–89, 93–94
Coercive behaviour, 264
Cognitive load theory, 51
Cognitive outcomes, 237–238, 243, 245, 247, 250, 255, 268, 270
Cognitive self-image, 246
Cognitive skills, 4, 366–374, 376
  average levels of, 372
  development, 372
Coherence, social, 82
Cohesion, 255, 272
Cohort design, 21, 133
Cohort studies, 133
Cohorts, 72, 87, 111–115, 132, 186, 217–218, 220, 224, 225, 350, 362, 365, 372, 374
  representative, 361, 365
Coleman Report, 12–13, 72, 289, 295
Collaboration, 111
  enhancing factors, 104
  international, 14

Collaboration and interaction, 54
Collaboration and networking, 2, 3, 53–54, 102, 108, 110, 111, 194, 211, 216, 217, 256, 272, 402
  differential impact, 104
  school-to-school support, 111
Collaborative arrangements, 105
Collaborative improvement, 117
Collaborative inquiry, 104
Collaborative structures, school-to-school, 2
College, 71, 295, 301, 312, 315, 319
College attendance, 292, 294, 311
  selective (USA), 311
College Choice, 294, 297, 306, 311, 312, 316, 319
College enrollment, 287, 291, 306, 312, 314
College financial aid, 301, 304, 312
College readiness, 292
College, selective, 295, 311
Commitment
  organizational, 253
  student, 252
Common school effect hypothesis, 250
Common steps of effective evaluation, 153
Communities, 54, 75, 104, 127, 139, 224, 250, 293, 330–332, 337, 340–341, 343, 347, 353, 392
  academic, 125
  advantaged, 211
  disadvantaged, 103, 343, 353
  dominant, 352
  local, 152, 352
  low socio-economic status, 138
  nested, 329
  networked, 101
  professional, 214
  school/centre, 332
Community, capacity building, 333
Community-led engagement initiatives, 343
Community members, 341
Community school subsidy, 254
Community support, 341
Comparability, 156, 162, 305
  external, 155
Comparative education (research), 14, 122, 135–139
Comparative research, success, 135
Comparative studies, 22, 23, 55, 59, 135–138, 185, 200
  international, 2, 9, 24, 139
Comparative work, authentic international, 138
Comparators, matched, 108, 111
Comparison groups, 106, 118
Comparison schools, matched, 106, 111, 115
Comparisons, Catholic and independent schools, 331, 335
Comparisons, Catholic and public schools, 249
Comparison, valid, 124
Compensational arrangements, 275
Compensations, 82, 86
Compensatory provision, 83
Competence, 262, 274
  heightened, 342
  social, 352
Competition, 51, 86, 102, 128
  international, 83
Complacency, 214, 399
Complementary cognition, 102
Complexity, 11, 37, 47, 148, 329, 330
Compliance, 213
Comprehension, 336, 363
  linguistic, 363
Comprehensive Doctrines, 79
Comprehensive model of educational effectiveness, 36–37, 39, 60, 233, 271
Compulsory Education Law of China, 209
Computers at home, 175
Conceptual frameworks, 20, 233, 385
Conceptual integrity, 233
Conceptual models (educational effectiveness), 189, 253
Conditions
  quasi-experimental, 173
  supportive, 255, 337
  weather, 201
Confidence, 340–343, 350
Confidence intervals
  overlapping, 134
Confidentiality, maintaining, 332
Configurations and constellations, 277
Confucian culture, 211
Congolese schools, 188, 191, 200–201
  effectiveness of, 188, 200
Congolese teachers, 202
Conscientisation, 347
Consensus among school staff, 255
Constructivism, 37, 39, 50, 264
Constructivist organisational theory, 102
Constructs, unidimensional, 38
Consumer choice, 85
Content knowledge, teachers, mathematics, 203–204
Context factors, 219, 249
Context level, 56
Context specificity, 225

Contexts
affective, 111
culture, 13
diverse, 12, 24, 329, 339, 401
external, 335
influence, 351
local, 401
local country, 137
micro aspects of, 18
national, 3, 18, 155
neighbourhood, 387
political, 11
relational, 347
single, 13
social, 136, 270, 341, 399
sociocultural, 60
socio-economic, 138, 316
Context-sensitivity, 393
Context-specificity, 3, 15, 18, 22, 24, 137, 139, 215, 225, 253, 276
Context-stripping, 386, 395, 397–398
Contextual awareness, 332, 387–389, 398, 401
Contextual complexities, 340
Contextual effects, 140
Contextual factors, 36, 136, 140, 216, 219, 225, 347
Contextual value added (CVA), 2, 387
Contextualised value-added, 2, 219, 220, 222, 224, 225, 387
Continental Europe, 15, 132
Continuing professional development (CPD), 114, 217, 223
Continuity, 43
Contracting external organisations, 111
Control conditions, 367
Control group, 364
Controversies, 301, 307, 317
Convergence, educational effectiveness and improvement, 14, 24
Conversations, professional, 337, 340
Co-operation, 83, 86, 192, 213, 253, 255, 260, 271, 272, 275
Countries
developed, 84, 185
effective, 56
high-income, 17–18
low-income, 15–17, 23, 361, 376, 400
middle-income, 361–362, 366
ranking (or league tables), 10, 20, 132, 150, 163
statistically comparing, 163
Western, 86, 93
Country characteristics, fixed, 175
Country comparisons, 20, 124, 128
Country contexts, 3, 12, 34, 351, 397
Country cultures, 136
Country differences, 123, 125, 131, 135, 163
Country league table rankings, 20, 150
Country's league table positions, 10
Country level effects, 175
Country-level, effectiveness measures, 21
Country level relationships, 176
Country-level trend data, 173
Country scores, 124
Country studies, single, 17, 350
Country variation, 124, 134
Coursework, academic, 295, 307, 314
Creativity, 210, 214, 255, 261, 329
Creemers, B., 231, 233
Criminal conduct, 295
Critical thinking, 77–78, 210, 292
formal, 77
Cross-country comparisons, 35, 154
Cross-country transference, 11
Cross-cultural comparability, 149
Cross-cultural judgments, high-inference, 22
Cross-cultural relevance, 236
Cross-cultural studies, 123, 137–138, 159, 163
Cross-national performance, 136
Cross-national research, 123
Cross-phase federations, 109, 110
Cross-sectional, 149, 175
Cross-sectional data, PISA, 176
Cross-sectional findings, 173
interpreting, 156
Cross-sectional studies, 133
Cultural capital, 269
Cultural change, 332
Cultural contexts, 9, 21, 82, 124, 138, 154
comparative, 137
Cultural dependency, 392–393
Cultural features, 24
Cultural homogeneity, 272
Cultural milieu of the school, 72
Cultural relevance of educational factors, 21
Cultural Revolution (China), 213
Culturally fair, 128
Culture change, 396
Cultures
educational, 250
implicit, 393
inclusive, 14
positive, 332
Curricular arrangements, flexible, 377

Curriculum, 49, 124, 130, 188, 210, 301, 328, 341, 353, 366
age-based, 377
areas, 123
choices, 259
offerings, 271
reform, 20, 209, 212

D
Data
aggregating, 168
attitudinal, 132
big, 138, 402
classroom level, 107
descriptive, 165
district, 174
economic, 140
empirical, 34, 71, 165, 216
evaluative, 54
family-related, 192
missing, 149
pooling, 169
process, 118
qualitative, 117, 216, 338
quantitative, 216–217, 339, 343
relative, 132
school-related, 194
sharing, 151
socio-economic background, 127, 132
test, 173
trend, 173
valid, 35
Data analysis, 128, 200, 233, 365
secondary, 35
Data collection
improved, 342
longitudinal, 133, 234
quantitative, 338
Data collection instruments, 191, 194
Data modelling, 72
Data quality, 159
Data sets, large international, 160, 223
Data use, 156, 272, 401
Database
large international, 169
Datasets
cross-sectional, 134
national, 118
Decentralised system, 55
Decentralization, leadership, 272
Deficit view of groups, 353
Deficit-oriented approaches, 344
Delphi process, 236
Demand side effects, 133
Demand side policies, 134
Democracy, 25, 78, 94
Democratic Republic of the Congo (DRC), 3, 185, 192, 200–202, 204, 400
Democratic societies, 78–79, 244
Demographics, racial, 289
Denomination, 241–242, 249–251, 261
effect of, 249–251
Department for Education (DfE, England), 104–105, 107–108
Deployment, 115
Deprivation, 78, 85, 94, 111, 114, 115
Descriptive analyses, 159
Descriptive findings, 169
Desegregation, 289, 317
forced, 317
Design
cross-sectional, 20, 395
enhanced, 148
experimental, 11
innovative educational effectiveness research, 17
International Large Scale Assessment, 177
longitudinal, 175
quasi-experimental quantitative, 106
repeated series approach, 139
trend, 175
Developing countries, 82, 159, 200, 204, 225
Developing economies, 86
Development
academic, 214, 316
all-round, 213
cognitive, 255, 364
early childhood, 361, 362, 376
educational policy, 216
emotional, 291
individual, 249, 264, 269, 271
institutional, 151
moral, 214
non-cognitive, 255
physical, 210, 367
psychological, 223
social/behavioural, 46
social-emotional, 231
theoretical, 73, 389
Development initiatives, 345
Development orientation, 252
Development programmes, 349
Developmental psychopathology, 388
Developmental skills, children, 362
Developmental trends, 246

Diagnostic data sources, 339
Diagnostic information, 338
Dialogue
  disciplined, 337, 342
  informal, 223
  reflective, 256
Difference-in-difference, 149
Difference Principle, 81–84
Differences
  between societies, 132
  contextual, 23, 248
  cultural, 21–22, 129, 130
  input, 251
  intake, 72
  national, 13, 134
  reducing
    initial achievement, 34
  student recruitment/selection, 244
Differential educational effectiveness, 43–44, 389, 392, 398, 399
Differential effectiveness, 21, 34, 37, 38, 46, 61, 104, 132, 232, 389
Differential effects, 3, 61, 154, 222, 240–243, 250, 251, 254, 256, 259, 263, 274, 276
  long-term, 260
Differential school effectiveness, 47
Differentiation, 38–39, 42–44, 265
Differentiation dimension
  dynamic model of educational effectiveness, 44
Direct effect, 114, 251, 253, 259, 296, 303, 305, 308, 309, 313, 364
Directedness, 364
Directionality, 171, 172
Disability, 86, 388
Disadvantaged areas, rural, 219
Disadvantaged backgrounds, 81, 364
Disadvantaged circumstances, 343
Disadvantaged groups, 73, 86, 88, 378
Disadvantaged positions, 275
Disadvantaged students, 107, 259, 269
Disciplinary boundaries, 402
Discipline, 71, 80, 255, 261, 263, 295, 302, 305
  fair, 302, 319
Discourses
  democratic, 25
  emerging, 347, 353
Discriminant function analysis, 59
Discriminative power, 161
Disparity, 209, 311, 349
  zero, 87
Dissimilarity indices, 294
Distribution of educational opportunities, 150
Distribution, college choice, 88, 310
District, effects, 101
Districts, 102, 106, 128, 130, 139, 173, 212, 216, 317, 328, 329
  effective, 101
  high-SEC, 317
Districts/Local Authorities, 131
Diversity
  ethnic, 258, 266, 269
Dropout, 293
Durkheimian notion of anomie, 103
Dynamic Approach to School Improvement (DASI), 53, 58, 95, 353
Dynamic model of educational effectiveness, 2, 15, 19, 34, 38–53, 61, 76, 78, 94, 400

E

Early childhood, 331, 333, 341
Early Childhood Education and Care (ECEC), 362–365, 376, 378, 401
Early Childhood Longitudinal Study (ECLS), 361, 365
Early childhood, programs, 361–362, 376, 378
Early Grade Reading Assessment (EGRA), 188
Early intervention, 365
Early school-improvement research, 22
Early Years Evaluation, 361–362, 376
East Asia, 11, 22, 159, 168, 224
East Asian approaches to teaching and learning, 11
Eastern Africa, 185
Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), 200
ECLS, *see* Early Childhood Longitudinal Study (ECLS)
Ecological fallacies, 176
Economic and Social Research Council (ESRC), 216
Economic approach, 34–35
Economic capital, 83, 290
Economic growth, 86, 123, 225
Economic migrants, 346
Economic productivity, 123
Economics, 35, 135, 352
Edmonds, 72, 233, 327
Education
  cultural, 255, 261
  equity in, 46, 47, 61, 151, 393

Education (*cont.*)
high-quality, 187, 201
improving, 400
localised, 401
moral, 214, 255
physical, 214, 394–395
post-secondary, 235, 297
pre-primary, 361, 376
pre-school, 122
private, 84
public, 288
quality-oriented, 213, 214
remedial, 80
social, 261
traditional, 254
Education Act (UK), 105
Education administration, 210
Education Endowment Foundation (EEF), 11
Education Law (China), 210, 213
Education Longitudinal Study (ELS), 297, 318
Education policy landscape, 76
Education production functions, 188
Education production models, 34–35
Education systems, 9–10, 15, 17, 19, 82, 95, 119, 136, 153, 209, 210, 216, 226, 344, 345, 353
features, 17, 21
low equity, 344
Educational advantage
low, 331
Educational arrangements, 275, 276
effective, 276
Educational assistance, 54
Educational change, 175, 297
Educational characteristics, 189
cultural, 17
Educational contexts, 40, 288, 402
diverse, 59
Educational effectiveness
differentiated, 4, 273–274
multilevel, 19, 36, 39–40, 57, 95, 106, 188, 248
new research tools, 385
Educational effectiveness research
agenda of, 34, 388
early, 33
early years, 12, 394
levels of, 397
phases of, 33, 72, 80, 383
psychological perspective of, 44–45, 47
sociological perspective of, 45–46
theoretical framework of, 38, 39
theories, 5, 15–16, 383–385
waves of, 389
Educational effectiveness research and improvement of practice, 34, 39, 61
Educational effectiveness research models, 36
Educational Effectiveness Theory, 34–39
Educational effects, new statistical approaches, 24
Educational environment, 56, 102, 160, 329–331, 351, 367, 387, 394, 399
background characteristics, 235
Educational equity, 85, 86, 94, 389, 392, 399
Educational evaluation, 151, 212, 231, 401
Educational governance, 147, 151–152
Educational improvement, 10, 14, 25, 211–212, 400–401
policies and initiatives, 401
research, 10
Educational inequality, 209, 289
Educational innovation, 24
Educational leadership, 233, 252, 253, 401
Educational networks, 16
Educational opportunities, equality of, 288
Educational outcomes, 11, 34, 46–47, 49, 258, 287, 315, 318
disadvantaged pupils, 74
enhancing, 275
Educational performance, 135, 295
comparative, 123
relative, 137
Educational policies, 11, 40, 49, 126, 133, 136–137, 139, 147, 151, 156, 225, 275, 297, 319, 388–389, 395, 398–399, 401–402
appropriately contextualised, 25
effective, 292
evidence-based, 402
national, 56, 128
Educational policy, factors, 136
Educational politics, 147
Educational positions, 238, 240, 245–247, 251, 275
Educational practices, 11, 38, 44, 131, 151, 234, 292, 388
effects of, 148
improving, 38, 39
maladaptive, 44
Educational processes, 24, 133, 398
Educational productivity models, 41, 48
Educational products, 188
Educational Prosperity model, 362, 377
Educational psychology, 393
Educational quality, 10, 24, 212, 224–226, 257
Educational resources, 54

Educational systems, 40, 55, 121–123, 126, 132, 133, 136, 140, 148, 152–154, 185, 186, 189, 190, 231, 387–388, 390–392, 396–398
centralized, 36
comparison, 396
effective, 123
nested, 12
EEF, *see* Education Endowment Foundation (EEF)
EER community, international, 14–15, 234, 397
Effect size, 16, 18–19, 114, 149, 293, 296, 316, 317, 319, 365
Effective education, theory/concept, 392, 399
Effective Pre-School and Primary Education, 46, 365
Effective School Improvement project, 22, 400
Effective teaching, 34, 35, 49–52, 59
Effectiveness criteria, 237–239
Effectiveness factor characteristics
qualitative, 42
quantitative, 59
Effectiveness factors, 13, 21, 38–43, 59, 136, 200, 392–393
Effectiveness metrics, 79
Effectiveness, contextualised, 21
Effectiveness-enhancing characteristics, 276
Effectiveness-for-equity, 74, 94
Efficacy, collective, 253, 259
Efficiency, 71, 134–135, 147
Efficiency perspective, 134–135
Efficiency, index, 134
Elementary and Secondary Education Act (ESEA; USA), 291, 292
Elementary schools, 365
ELS, *see* Education Longitudinal Study (ELS)
Emancipatory discourses, 347
Emerging contexts, 24
Emotional skills, 291
Emotional support, 254
Empathy, 94, 332
Employment earnings, 295
Ender
differences, 187
Engagement
initial, 274
inter-profession, 402
parent, 157, 342
prior, 274
school, 254, 273
England, 2, 11, 20, 46–47, 86, 102, 104, 106, 107, 117, 118, 122, 127, 150, 224, 330
accountability system, 111
English education ministry, 104
English-speaking countries, 102, 154, 162–163, 173, 176
Enjoyment, 343
school, 254, 388
Enrollment rates, 187, 376
Enthusiasm, 332
Entrance Exam to Higher Education (EEHE), *see* Gaokao' (Mainland China)
Entrance Examination to Senior High School (EESHS; China), 216, 218
Environment
cultural, 295
instructional, 38
psychological, 41
school
orderly, 256
Epistemes, 392
Equal opportunities, 44, 81, 95, 259, 378
Equality of opportunity, 96, 293
Equality, educational, 288, 399
Equations, structural, 200
Equipment, 201
lacking, 187
Equitability, 86
Equitable student outcomes, 402
Equitable system, 86
Equity, 1–2, 4, 10–11, 34, 47, 61, 71–72, 75–88, 90–96, 186, 188, 209, 210, 353, 399, 400
defining, 85
low, 344
measurement of, 84–88
operationalisations of, 399
perfect, 88–90
principles of, 84
social, 399
Equity agenda, 86, 389
Equity gaps, 47, 399
Equity metrics, 91, 93
Equity paradigm, 87
Ethnic minority groups, Hispanic, 290, 300
Ethnic origin/nationality/language, 242–243
Ethnic/racial groups
See also race
Ethnic-cultural background, 274
Ethnicity, parental, 251
Ethnocentricity, 13, 18
Europe, 19, 22, 36, 59, 90, 96
European Association for Learning and Instruction (EARLI), 14
European Conference on Educational Research (ECER), 14
European Economic Community, 385
European Economic Community (EEC), 122, 385

European Union, 85
Evaluación Infantil Temprana (EIT), 361–362, 367–368, 370–372, 376
Evaluation, 95, 108, 115, 150–156, 167, 173, 186, 212–214
  effective, 153
  external, 152–153, 156, 376
  improved, 212
  internal, 149, 153, 155, 156
  quality, 210
  school-level, 171, 338
  systematic, 74
Evaluation and accountability, 171, 173
Evaluation for improvement, 153
Evaluation mechanisms, 55, 56, 155
Evaluation methods, value added, 215
Evaluation practices, 153
Evaluation programmes, 152
Evaluation projects, 44
Evaluation responsibilities, 213
Evaluation results, 151–152
Evaluation systems, 153
  regional, 226
Evaluation use, 153
Evidence
  causal, 118
  intervention-based, 398
  qualitative, 119, 217
  quantitative, 104, 341
  syntheses of, 17–19
Evidence base, contextualised, 211
Evidence-based approach, 345
Evidence-based research, localized, 211, 401
Examinations, national, 155, 213
Exclusion, 344
Exclusion rates, 344
Expenditure
  financial, 87–89, 134, 140, 209, 213
  per student, 134
Experimental studies, 11–12, 61, 176, 398
  small-scale, 24
Explanatory power, 138, 165
External brokers, 117–118
External influences, 102

F
Facilities, 388
  equal, 288
Factors
  community, 367, 377
  context-level, 56
  cultural, 18, 23, 44, 47, 131, 135, 140
  economic, 131, 135
  educational, 22–24, 57, 60, 123, 131, 133, 136, 139, 397
  instructional, 243
  intake, 76
  interpersonal, 333
  learning attitude, 217
  learning environment, 261–274
  middle-level, 55
  non-cognitive, 214
  non-educational, 131–132
  non-school, 132
  organizational, 233
  personal, 331
  political, 215
  professional development, 216
  socio-economic, 133–134, 329
  supply side, 134
  system-level, 40, 72
  unidimensional, 39
Failure, pupil, 95, 262
Fair allocation of resources, 90, 94
Fake news, 402
Families, 72, 104, 125, 150, 270, 294, 301, 304, 315, 335, 337–338, 340, 343, 346, 353
  affluent, 290, 304, 317
  disadvantaged, 150, 210, 256, 304
Family engagement, 341
Family income, 300
Family structures, 101
Federations, 105–106, 108–110, 117
  configurations, 118
  impact of, 105, 106, 108–110, 117, 118
Feedback, 35, 41, 44, 150–152, 154, 157, 159–164, 168, 170, 172, 177, 217, 223
  activities, 157
  approaches, 172
  constructive, 332
  corrective, 154
  formative, 164, 172
  individual, 265
Feedback-based approach to school improvement, 152, 167
Feedback cultures, 161
Financial aid, external, 185

Financial support, schools, 56
Finland, 136, 165
First Literacy Study of the IEA, 123
Five-factor model, 233
Fixed effects, 149, 169
Flanders, 231–232, 234–237, 247, 249, 254, 259, 261, 266, 270, 271
Focus
  academic, 301, 314, 319
  curricular, 293
Focus groups, 216
Formative assessment, 52, 152–154, 157, 160, 163–164, 168, 172, 176, 213
Four-year college enrollment (USA), 297, 306, 308, 313, 316, 319
  selective, 310, 315
Free School Meals (FSM), 85, 105–108, 111, 115, 118
Freedom, 78, 81, 83, 85, 332, 397
  political, 86
Friendship, 78, 301
FSM, *see* Free School Meals (FSM)

G
Game theory, 80
Gaokao' (Mainland China), 213
Gap
  attainment/achievement, 47, 96, 107, 115, 153, 335, 344
  poverty, 288
Gender, 4, 35, 44–47, 106–111, 114, 115, 194, 196, 217, 221, 243, 259, 264, 274, 366, 368, 370–372
  differences, 47, 266
General Certificate of Secondary Education; UK (GCSE), 108, 110, 114, 115
Generalisability of effectiveness factors, 20
General-rule Utilitarianism, 77
Generic models, 38, 43
German-speaking countries, 161, 163, 165–167
Germany, 20, 149–150, 153, 161, 163–165, 365
Gini-based metrics, 90–93
Global Competencies, 127
Globalisation, 10, 12, 125
Goal orientation, 37, 171
  mastery, 247
Goal-oriented professional planning, 172
Governance, 102, 105, 127–128, 188, 209, 250, 252, 345
  joint, 105
  structures, 241, 250
Governance by information, 147
Grade Point Average; USA (GPA), 307
Grade retention rates, 308
Graduation rates, 289, 292, 309
Green Indicators (Shanghai, China), 214
Gross national income (GNI), 362
Group composition, 4, 200, 236, 249, 267, 273, 275
  effects of, 196, 232, 267, 275
Group interviews, 330
Grouping procedures, 233
Groupings, 95, 157, 172, 191, 241, 294, 390, 394
  geographical/linguistic, 165
  stereotypical, 47
  structures, 394
Growth Record Report (Mainland China), 213

H
Hare, 77
Head Start, 365, 398
Headteachers, 106, 108, 115, 117, 127, 328
  observation, 224
Health, 213, 390
  mental, 214, 292
  physical, 214, 292
"Hedonic calculus", 76
Hierarchical levels in Educational Effectiveness Research, 391, 392, 394
Hierarchical linear models (HLM), 233, 370, 374
High achievement, influenced, 250
High school graduation (USA), 287, 291, 294, 297, 306, 309, 310, 314, 319
High schools
  high-SEC, 287, 290, 295, 307–312
  low-SEC, 295, 312, 315
  private, 296
  public, 297, 311
Higher education, 237, 247
High-risk students, 260, 263, 264, 269, 274
HISEI index of occupational status in PISA, 171
Historic contextualisation, 387
Home characteristics, 46
Home education level, 368
Home learning environment, 46, 54, 71, 125, 169, 217, 218, 235, 387
Homework, 71, 194, 211, 255, 295, 301
Hours in school'/hours in school, 135
Household average per capita income, 209
Hukou, 129, 217, 218

Human capital, 123, 132
Hungary, 122
Hyperactivity, 248

I
Iceland, 163, 165, 168
ICSEI, *see* International Congress for School Effectiveness and Improvement (ICSEI)
IEEQC, *see* In China (IEEQC)
Illiberal ideologies, 79
ILSAs
  cross-sectional, 151, 175
  data, 20–21, 148–151, 154, 171, 177, 385, 386
  paradigm, 22
  questionnaire design, 148
Immigrant background, 47
Immutability Principle, 95
Improvement
  beyond-school, 211
  continuous, 152
  evidence-based, 39, 401
  sustainable, 119
  sustained, 84
Improvement activities, 14, 152
Improving Educational Evaluation and Quality in China, 3, 212, 216
Improving Teacher Development and Educational Quality in China (ITDEQC), 3, 23, 216
In China" (IEEQC), 3, 212, 216, 224
Incentives
  financial, 102
  perverse, 79, 293
Inclusion, 105, 210, 346
  social, 76
Income Deprivation Affecting Children Index (IDACI), 107–108, 111, 114, 115, 118
Increased spending, 101
Independent schools, 331, 335
Indicators
  leading, 377, 401
  quantitative, 213
  trailing, 377
Indigenous communities, 339–340, 343
Indigenous Leadership Partners (ILPs), 340
Indigenous students, 344
Indirect effects, 40, 43, 101, 252, 259, 268, 303
Ineffectiveness, educational, 16
Inequality, 46, 74, 81–84, 92–96, 127, 212, 288, 318, 329, 385, 399, 401
  reducing, 5, 74, 377–378
Inequities, 92, 150, 209, 390, 399, 402
Information society, 138
Information technology, 125
Infrastructure, 201, 400
Initial teacher education, 118
Input-output studies, 36
Inputs
  financial, 134
  quantified, 327
Inspection, 83, 115, 155, 212–214, 225, 235, 271
Instruction, 3, 37, 41, 43–44, 51, 157, 191, 223, 234–235, 241, 262, 265, 267, 274, 275, 295, 301, 368, 377
  academic, 51
  active, 44
  clear, 263
  differentiated, 150, 257, 264, 274
  direct, 171, 172, 264
  explicit, 263
  structured, 263
  supportive, 262
  well-organized, 264
  whole-class, 254
Instructional conditions, 48
Instructional goals, 5, 378
Instructional practices
  constructivist, 264, 274
  traditional, 264
Instructional support, 261, 267, 272, 273, 275
Instructional time, 262
Instrumentalism, 86
Integrated models of educational effectiveness, 37–40, 46, 48, 59, 73, 233
Integration, 266, 268, 289, 290, 317, 388–389
  multiethnic content, 273
  neighborhood, 317, 319
  racial, 288, 290
  social, 248, 255, 368
  socioeconomic, 288, 317
Intellectual impairment, 127
Intelligence, 48, 80, 255
  high, 255
  numerical, 260
Interaction effects, 37, 48
Interactions
  behavioural, 394
  cross-level, 15, 17, 37, 138, 276
  student-student, 51
  teacher-student, 51
Inter-district redistribution, 317
International Association for the Evaluation of Educational Achievement (IEA),

19–20, 122–125, 127, 130, 137, 147, 163, 169, 385
International Association for the Evaluation of Educational Progress (IAEP), 123–125
International comparative reviews, 154
International comparisons, 72, 82, 147, 385, 387, 392
International Congress for School Effectiveness and Improvement (ICSEI), 10, 13–15, 122, 233, 234, 291, 396, 401
International Large Scale Assessments (ILSA), 2–3, 5, 9–11, 19–20, 23–24, 149, 151, 154, 163, 170, 173, 176, 384–386, 395–399
International research, comparative, 2, 22, 23
International School Effectiveness Research Project (ISERP), 21–22, 24
International School Improvement Project (ISIP), 22
International Successful School Principalship Project (ISSPP), 22, 330, 331, 333, 352, 400
International System for Teacher Observation and Feedback (ISTOF), 14, 22, 35, 236
Internationalization, 10, 121–122, 383, 389, 396
Interventions
  evaluating, 378
  school-based, 389
  school-level, 114
Interventions/practices, 331
Intra-class correlation coefficient, 307
Intuition
  experiential, 77
  professional, 78
Invariance, configural, 160
Ireland, 164, 167, 169, 176
ISERP, *see* International School Effectiveness Research Project (ISERP)
Islamic schools, 251
Israel, 16, 122, 125
ISSPP, *see* International Successful School Principalship Project (ISSPP)
ISTOF, *see* International System for Teacher Observation and Feedback (ISTOF)
Item difficulties, 161
Item Response Theory (IRT), 126, 149, 158, 163, 169
Item-total-correlation, 160

J
Jaap Scheerens, 231
Jan Van Damme, 231, 234
Japan, 20, 122
Job satisfaction, 272
Joint effects, 253, 258, 261, 390
Journal for Research in Educational Effectiveness, 14
Justice
  distributive, 84
  social, 86–87, 94–96, 332, 347, 349
Justice as fairness'/justice as fairness, 75

K
Ka Hikitia, 346, 350
Katholieke Universiteit Leuven (Belgium), 3
Kaupapa Māori theories, 346
Kia Eke Pamuku, 350, 396
Kindergarten, 4, 5, 250, 254, 262, 361–362, 364–368, 376–377, 401
  intervention, 367
  skills at entry, 5
  universal provision, 361–362
Knowledge
  academic subject, 223
  alphabet, 363
  background, 363
  cultural, 346
  evidence-based, 342
  letter, 364
  letter-and-sound, 336
  literacy, 342
  local, 3, 17, 25
  metacognitive, 246, 262, 264
  specialized prior, 38
Korea, 167

L
Labour Party (UK), 96, 335
Language, 107–111, 114, 115, 243–245, 247, 257, 259, 263, 265, 267, 269, 352, 363, 366, 367, 370, 386
  achievement, 237–238, 243, 249, 255, 258, 260, 265, 269, 273
  common, 340, 342
  comprehension, 363
  controlling, 264
  development, 376–377
  national, 352

Language (*cont.*)
oral, 336
outcomes, 237
proficiency, 245, 269
scale, 368
second, 262
skills, 362–363, 368, 374, 376
test, 372
Large scale assessment, international, problems, 124
Large-scale assessment, international, 10, 19, 102
Latent variables, 160, 169
Latin America, 4, 16, 361–363, 365, 376, 377
Leaders, 111, 272, 328–329, 332, 335–338, 340, 342, 348, 350, 352, 353
capability, 345
early childhood, 343
indigenous, 340
middle-level, 106, 335
participative professionality-oriented, 272
philosophy
personal, 332
strategic change, 346
successful, 332, 351
turnover, 340
Leadership, 44, 72–73, 105, 107, 111, 115, 118, 211, 214, 253, 327–331, 334–336, 339–343, 346, 351, 353, 386, 390
"both ways", 340
capacity, 108, 118
connect to learning, 337
critical, 350
deliberate, 104
depth, 340–341
distributed, 333, 335–336
effects of, 252, 253
general dimensions, 351
generic, 343
generic skills, 328
goal-setting, 22
institutional, 390
instructional, 44, 328–330
integrated, 253
and literacy, 343
local authority, 328
middle-level, 330
participative professionality-oriented, 271
pedagogical, 333
positive, 331
practice, 347
qualities, 253
shared, 333, 337, 340, 350
sharing, 342
skills, 350
strategic, 252
styles, 252, 335, 352
successful, 329, 333, 350, 352
support, 111, 115, 272, 342, 346
support for literacy, 342
transformational, 253, 272, 328
Leadership for learning, 328–329, 335–336, 352
Leadership for Learning Blueprint (LfLB), 336–338, 341, 350
Leadership partnerships, 342
Leadership team, 272, 336, 340
strategic change, 348, 350
League tables, 132, 134
international, 83
Learners, 154, 337, 344, 348, 392
Learning
adaptive, 265
additional (outside school), 135
authentic, 263
collaborative, 263
community, 329, 352
contextual, 264
cooperative, 264, 274
discovery, 263, 264, 274
improving, 52
individual, 151, 352
innovative, 263
inquiry-based, 172
meeting children's developmental learning needs, 366
mutual, 24
organisational, 152, 352
out-of-school, 71
prior, 48
rote, 264
term, 54
Learning activities, 50, 169, 264
play-based, 367
Learning aptitudes, 35
Learning atmosphere, 85
Learning climate, 72, 253, 256, 265, 267, 271, 272
average, 270
establishing, 4
negative, 273
optimal, 271
Learning contexts, 347
Learning conversations, 333

Learning difficulties, 80, 83, 188
Learning disposition, 37
Learning environment, 40, 49, 51, 52, 54, 56, 82, 160, 231, 232, 234–236, 241, 246, 248–249, 263, 265, 267, 271
  characteristics, 4, 235, 271, 273
  configurations, 270, 276
  effective, 51
  effects of, 270–271
  effects of configurations of, 270–271
  emotional, 337
  enriched, 295
  establish better, 56
  favourable, 270
  high orderly, 255
  levels, 271
  multiple common, 395
  orderly, 255, 261
  quality of, 275, 276
  school, 82
  structured, 263
  supportive, 275
  unfavourable, 270
Learning factors, 217
Learning goals, 157
Learning growth, 250
Learning improvements, 336
Learning in Families Together (LIFT), 340
Learning inwards, 104
Learning models, 234
Learning opportunities, 52–53, 151, 175, 194, 196, 275
  out-of-class, 150
  out-of-school, 72
Learning outcomes, 34, 45, 51, 56, 58, 60, 117, 231, 333
  ambitious, 103
  improving/increasing, 34, 50, 54, 330
Learning processes, 35, 48, 125, 154, 190
Learning programs, blended leadership, 342
Learning progress, 38
Learning rates, 248
Learning resources, 54
  insufficient, 394
Learning tasks, 45, 49, 244, 246, 248, 255, 268
Learning time, 52
Learning-to-read to reading-to-learn transition, 362
Lessons, laboratory-based science, 395
Letter identification, 363
Levels
  national/regional, 40, 56, 173, 225, 377
  occupation, 118
LfLB, *see* Leadership for Learning Blueprint (LfLB)
LGBT+, 399
Liberty Principle, 81
Limitations
  financial, 312
  methodological, 3, 129
  philosophical, 86
Limitations of ILSA designs, 149, 155
Linear models, 149
  hierarchical, 370
  hierarchical generalized, 305
List of ingredients approach, 137
Literacy, 136, 335–336, 338, 340, 342–343, 347, 352
  improving, 342
  reading, 257
Literacy advisers, 338–339, 342
Literacy blocks, 342
Literacy development, 342, 350
Literacy knowledge, 363
Literacy practices, 342
Literacy test, 130
Literature reviews, 17–18, 362, 366
  commissioned, 122
Load, reading, 129
Local administrations/authorities, 329
Local context issues, 225
Local Education Authorities (LEAs), 55, 101, 105–106, 117, 214, 216–221, 224, 226, 342, 390
Local insight, 23
local inspection systems, 214
Local researchers, 3
Locations, education-supportive, 290
London Challenge, 86
Longitudinal case studies, 343
Longitudinal data, 133, 175, 367
Longitudinal Research in Secondary Education project, 193, 234–237
Longitudinal studies, 133, 276, 363, 365
Long-term effects, 235, 239, 247–248, 362, 364
  addressing, 239
  investigated, 247, 260
  larger, 247
  significant, 247
  studied, 247
Lorenz curve, 90
Lorenz function, 91
Low achievers, 153
Low performers, initially, 273
Low response rates, 127
Low socioeconomic status, 289

Lower ability students, 257
Lower achievers, 44
Lower socio-economic status groups, 132
Low-risk students, 263, 264
Low-SES students, 250, 257, 269, 274, 294

M
Management
  cultural elements, 270
  financial, 213
  formal, 270
Management styles, 270
  effective, 270
Management systems, 138
Management team, 54
Managerial arrangements, 131
Managerial capacities, 252
Manipulations, experimental, 12
Māori, 344–345, 347, 350, 352
Māori Education Strategy, 345
Māori language, 347
Māori learning in mainstream schooling, 350
Māori student achievement, 344, 345
Māori students, 4, 344–345, 350, 351
Māori, student achievement, 346
Marginal utility, 93
Marginalised groups, 346, 353, 402
Mastery learning, 37, 50, 265
Matching, propensity score, 105, 107–108
Materials
  authentic learning, 264
  distributing, 51
Mathematics, 111–114, 128–130, 188, 192, 197, 199, 200, 202–204, 243–244, 246–250, 254, 257, 260, 263, 266, 269, 365
  achievement, 190, 200, 237–238, 245, 247, 249–253, 257, 260, 261, 263, 266, 271, 272, 307
  achievement growth, 250, 260
  curriculum, 301
Mathematics achievement
  gender differences, 268
Matrix-design, 158
MATs, *see* Multi-Academy Trusts
MATs (Multi-Academy Trusts), 102
Maturity effect, 4, 366–367, 370–373, 375
Maximising average utility, 75
Measurement, 5, 17, 35, 38, 73, 78, 85, 86, 92, 95, 96, 122, 128, 129, 152, 169, 176
Measurement equivalence, 128
Measurement error, 134, 275
Measurement frameworks, 38, 42
Measurement invariance, 149, 163–164, 169, 176
Measurement, cross-cultural, 160
Measures
  aggregated, 175
  composite, 307
  developing, 14, 156
  direct, 118
  dispositional, 176
  efficiency, 134–135
  formative, 157
  non-cognitive, 291
  repeated, 149
  "value-added" school effectiveness, 217
Mediated effects, 303, 313
Mediation, 303, 307, 398
Mediation effects, 303, 313
  multiple, 303
Mediators, 303, 316
Mental self-government, theory, 37
Mentoring, 111, 224
Meritocracy, 213
Meta-analyses, 9, 12, 17–19, 34, 49, 51, 53, 55, 58, 287, 363–365
  estimates, 307
  multilevel approach, 18
Meta-analytic approach, 19
Meta-cognition, 263
Metacognitive items, 127
Metacognitive skills, 58, 127–128, 246
Meta-cognitive strategies, 265
Methodological issues, 12, 129, 139, 151, 156
Methodological knowledge bases, 384
Methodological topics, Educational Effectiveness Research, 233
Methods
  context-sensitive, 22
  econometric, 76
  lesson study, 122
  snowball, 232
Methods of Researching Educational Effectiveness and Improvement (MoREI), 14, 401
Metric invariance, 149, 160–162
Metrics, non-traditional, 76
Middle level leaders, direct impact, 335
Migrant students, 129, 148
Minimum educational provision, 83
Minority groups, 257, 269, 275
Minority rights, 332

Minority students, proportion of, 257, 269
Mixed methods, 24, 106, 140, 386, 397, 398
Mixed methods Education Effectiveness Research (MMEER), 397
Mixed methods educational effectiveness research, 397
Mixed Methods International Research Association (MMIRA), 397
Model classes, 223–224
Model fit, 160–161
Model programs, 364
Modelling
  growth curve, 327
  multi-level, 149, 327
Models
  correlational, 317
  econometric, 152
  process-product, 189
  quantitative, 292
  regression-type, 169
Models of educational effectiveness, 33, 37, 38, 235
Moderating effects, 277
Moderation, 398
Modified Delphi technique, 22
Monitoring, 95, 151, 212–214, 252
  evidence-based, 335
  external, 150
  internal teacher, 155
  system-level, 154
Monitoring equity, 96, 400
Moral alignment, 75
Moral imperative, 74
Moral purpose, 86, 336–337, 339, 346, 349, 353
  shared, 342, 350
Motivated actions, 262
Motivation, 35, 36, 47–50, 202, 204, 238–239, 246, 255, 266, 267, 269, 273
  autonomous, 244, 260, 265
  identified, 254
  low-quality, 260
  student
    increasing, 50
  subject, 45, 48
Motivational aspects, 244, 254, 264, 266, 276
Motivational constructs, 246
Multi-Academy Trusts (MATs), 102, 110
Multi-group Confirmatory Factors Analyses, 159, 163, 169
Multilevel growth curve modelling, 108
Multilevel logistic regression, 309
Multilevel mediation, 302, 303
Multilevel modelling, new statistical programs, 233
Multilevel models, dependencies, 302
Multiple perspective case studies, 330

N
Narrative reviews, 18, 19
National Center of Education Statistics (NCES), 300, 365
National cultures, 122, 126, 135–137
National evaluation system, 226
National Institute for Education Sciences (NIES) Beijing, 3, 212, 216
National level policies, 127
National level programmes, 133
National patterns, 165, 176
National performance tables, 134
National policies, 3, 55, 56, 60, 72, 128, 131, 139, 173
  improving, 56
  natural change, 173
National Pupil Database (NPD), 105, 107
National research studies, 21, 36, 57, 122
National reviews, 137
National test programs, 152
Nations
  developed, 287
  industrialised, 126
Natural disasters, 24
Natural distribution of talent, 84
Natural experiments, 388, 396
Naturalistic fallacy, 77
Neighborhoods, 10, 12, 72, 290, 304, 319, 329, 400
Nested data, 302
Nesting, cross-classified, 395
Netherlands, 4, 10, 18–19, 36, 138, 234–237, 239–247, 250–252, 257–258, 261, 266, 267, 271, 274, 276, 397, 401
Networking
  approaches, 118
  impact of, 103–104, 117
  link to pupil outcomes, 103
  theories, 118
Networks, professional, 102
New Social Movements, 103
“New learning” approach, 37
New Zealand, 4, 164–167, 169, 176, 329–333, 344–345, 352, 400
  achievement, 344
  education system, 344
  indigenous Māori students, 344
  PISA description, 344
New Zealand case studies, 4, 331, 350
No Child Left Behind’ Act, 133–134, 291
No-Harm Principle, 95

No-harm test, 95
Non-cognitive outcomes, 193, 234–235, 238, 239, 243–251, 256, 260–274
Non-experimental data sources, 317
Non-linear effects, 305, 318
Non-linear relationships, 48
Non-task-specific strategies, 154
Nordic group of countries, 165
Norm-referenced examination systems, 96
Norms, 165, 350
  cultural, 211
  national, 175
  shared, 160
  system-wide, 174
North America, 13, 15, 22, 162, 330
Norway, 165, 330
Numeracy, 122, 123, 347
  skills, 365, 368

O
Observation
  direct, 20
  instruments, 236
  systematic, 13
OECD, *see* Organisation for Economic Co-operation and Development (OECD)
OECD, criterion for educational success, 86
OECD data, 128, 158
OECD members, 128, 155, 158–159, 163, 174, 257, 288
OECD ranking, 163
Ofsted inspection grades (England), 115
One size fits all, 60, 126, 139, 400
On-task behaviour, 51
On-task time, 52
Operationalization, 122, 137, 235, 255, 399
Opportunity to learn, 49, 124
Optimism, 332
  academic, 259
Organisation for Economic Co-operation and Development (OECD), 20, 74, 76, 122, 126, 128–130, 139–140, 147, 150–151, 153, 157–158, 163, 170, 344
Organization (of schools/education systems), 13–15, 35, 72, 102, 122, 127, 163, 252–254, 276, 329, 333, 349, 366, 385, 402
Organizational arrangements, 241
Organizational issues, 51
Organizational theories, 35–36
Organizations
  direction, 149, 171, 176, 224, 251, 268, 348, 351
  external, 111
  international, 122, 126, 147
  supranational, 10
  transnational, 74, 83, 85
Outcome measures
  multiple, 291
  social, 21
Outcome variables, 244, 297, 370, 374
Outcomes
  academic, 48, 125, 214, 314
  achievement gap, 292
  achievement test score, 307
  affective, 94
  authentic, 352
  behavioral, 94, 318
  behavioral engagement, 309
  commonly-investigated, 239
  early childhood, 376
  employment, 294
  multiple, 288, 291, 293, 313
  non-cognitive, 238
  post-secondary, 287, 301
  proxy, 75
  psychosocial, 364
  reading, 342
  school behavior, 291
  school career, 240
  single, 291, 293
  social, 21–22, 125, 270
  social-behavioural, 248
  socioemotional, 388
  socio-emotional, 1
  student attainment, 110
  subject, 217, 222
Outlier items, removing, 130
Outliers, 87
  extreme, 87
Outputs, 72–73, 75, 78, 87, 134, 189, 327
Output-to-input ratio, 71
Overarching EER theories, 148
Overestimation, 249, 301, 318
  avoiding, 301
Overgeneralizations, 175, 398
Overstatements, 150, 398
  avoiding, 175

P
Pacific Rim, 124
PALL (Principals as Literacy Leaders, 335–336

Palma Index, 87, 92–93
Paradigm shifts, 396
Paradigm-dependency, 392
Paradigms, 75, 125, 176, 392–393, 397
  assessment/evaluation, 152
  utilitarian, 85, 94
Parent support, 196
Parents, 42, 52, 54, 56, 80, 127, 129, 132, 151, 152, 156, 213–214, 258, 259, 267, 268, 330, 338, 340–341, 343
  backgrounds/perceptions, 127
  disengaged, 343
  education, 118, 217, 249, 251
  education level, 300
  informing, 42, 156–157
  involvement, 19, 42, 172, 273
  low educational attainment, 254
  qualification/occupations, 46, 217
  socio-economic status, 118, 169, 301
Parsimonious model, 34
Parsimony principle, 37
Partnership policy, 54
Partnerships
  external, 102, 211
  school-parent, 333
  schools, 105, 106, 111, 158, 186, 335, 338, 340, 346
Pedagogical conceptions, 194, 195
Pedagogical practices, 131, 140, 203, 272
Pedagogical strategies, sharing, 333
Pedagogy, 131, 211, 352
  effective, 103, 348
  relational, 348
Peer achievement, 299, 301, 305
Peer effects, 78, 313, 316
  malevolent, 95
Peer group acceptance, 259
Peer groups, 35, 270
Peer influence, friendship, emotional connections, 301
Peer influences, 4, 293, 296, 301, 306, 309, 313, 316, 319
  socioeconomic-based, 295
Peer parental capital, 299, 305
Peer pressure, 268
Peers, 111, 213, 259, 268, 287, 295, 301, 316, 377
  advanced, 295, 394
Peer-to-peer effects, 82
People, developing, 351
Perceived Feedback (PERFEED), 156, 159–164, 169, 170
PERFEED, *see* Perceived feedback (PERFEED)
Performance approach orientation, 248
Performance avoidance behavior, 244
Performance avoidance orientation, 246
Performance Federation, 109, 110, 117, 118
Performance feedback, 162
Performance measures, 105
Performance orientation, 252
Performance standards, 20
Perseverance, 48
Persistence, 332
Personal Leadership Profile (PLP), 338, 339, 341
Personal prestige, ascribed, 346
Personality, 35, 45, 47, 48
Personality development, 255, 261
Personality traits, 37, 48
Personality type, 44
Person-environment/resource, 394
Perspectives
  comparative, 12, 17
  contextual, 140
  country against country league table ranking, 123
  critical, 126
  ecological, 9
  multi-informant, 235
  pluralistic, 393
  quantitative, 42
  sociological, 35, 46
  theoretical, 10
Phantom effects, 317
Phase Five, educational effectiveness, 74, 78, 94–95
Philosophy
  credible, 94
  formal, 94
PISA
  effectiveness perspective, 134
  response rates, 124, 127, 129
  sampling process, 127
Plausible Values, 158, 168
PLC features, 223
PLCs, *see* Professional learning communities (PLCs)
Plowden Report, 12
PLP, *see* Personal Leadership Profile
Policies
  assessment-related, 171
  enrolment, 209
  federal, 289
  ineffective, 396
  measure, 155
  national/regional, 56
  school-based, 165
  school-level, 172
  socialist, 136

Policies (*cont.*)
system, 156
system-level, 23
Policy
characteristics, 241
macro level, 127–128, 133
Policy context, 24, 71, 80, 118
Policy environment, 329, 351
Policy levers, demand side, 134
Policy makers, 9, 10, 12, 14–15, 20, 25, 48, 56, 76, 78, 85, 86, 90, 92–93, 95, 126, 134, 150
contemporary, 398
educational, 225
local, 216, 392
Policy transfer, 10–11, 21
Politicians, 126, 401–402
Popper, K., 78
Postsecondary, 295, 304, 312, 314
Poverty, concentrated, 287, 317, 318
Poverty rates, 289
Poverty reduction, 225
Power
balance of, 79
coercive, 79
Power imbalances, perceived, 104
Practice, successful, 122, 332
Practices
between-culture, 24
evidence-based, 11
Practitioners, autonomous, 78
Praise, 262
Praxis, transformative, 347
Preschool, 262, 361, 364
Pressures, 126, 138, 252, 293, 388, 397
corruption, 293
external, 48
Primary education, free, 187
Primary goods, 81, 83
Primary schools
effectiveness, 3, 237, 240
high-achieving, 254
long-term effects, 247–248, 250
students
high-risk, 263
Principal-agent-theory, 152
Principal ratings, of teachers, 301
Principals, 147, 150, 158, 164, 167, 172, 174, 176, 191, 194, 195, 201, 202, 328, 330–336, 338–342, 346, 352
effective, 44
personal qualities, 331, 332
personal resiliency, 332
primary, 343
secondary, 4
successful, 22, 330, 331, 351, 352
Principals as Literacy Leaders (PALL), 4, 335, 338–343, 350, 352
Principals as Literacy Leaders in Indigenous Communities (PALLIC), 339–340
Principle of Utility, 76, 77
Problem behaviour, 244, 258
Problem solving, 50, 127, 210
Problems
administrative, 55
disciplinary, 294
localised, 24
methodological, 125, 176
Process factors, 138, 224, 225
Processes, educative, 71–72
Process-product approach, 188
Production functions, 135, 189
Professional activities, 171
Professional development, 3, 54, 58, 59, 103, 108, 122, 133, 157, 201, 211, 215–217, 223–224, 333, 340–342, 349, 401
Professional learning, 177, 216, 224, 335–336, 338, 339, 341, 343, 346, 352, 401
active, 337
effective, 349
well-designed, 335
Professional learning communities (PLCs), 23, 211, 215–217, 223–225
Professional practice, 78, 216, 401
Programme for International Student Assessment (PISA), 20–21, 86, 102, 122–123, 125–138, 154–161, 163–165, 171, 176, 186, 344
Programme for International Student Assessment, *see* PISA
Progress in International Reading Literacy Study (PIRLS), 20, 125, 127, 130, 147, 154, 158, 175, 186, 200
Progress in International Reading Literacy Study, *see* PIRLS
"Proles", 77
Protestant schools, 249, 251
Psychology, 36, 48
Psychometric properties, 55, 376
Public assistance, 295
Public schools, 125, 129, 249–251, 290, 292, 297
Public service provision, joined-up, 211, 401
Pupil characteristics
background, 12, 46, 47, 72, 114, 115, 235, 274
cognitive, 192
individual, 196
non-cognitive, 192

Pupil Level Annual Schools Census; England (PLASC), 105
Pupils
advantaged, 75, 82–84
bright, 81–82
disadvantaged, 72, 74, 82, 95, 115
gifted, 82
high-achieving, 94
troublesome, 83
Pupil-teacher ratio, 88

Q
Qualifications, 194, 293, 345
Qualitative approaches, 106, 397
Qualitative case studies, 106
Quality, 1–2, 34, 43, 61, 114, 138, 150–151, 187, 188, 210–212, 215–216, 241, 261, 264, 272, 275, 335, 364, 377–378
moral, 213–214
structural, 394
Quality assurance, 212
Quality, curriculum, 19
Quantitative approaches, 24, 160
Quantitative research, large-scale, 397
Quantity of teaching, 53, 138
Quasi-Experiment Designs (QEDs), 11, 398
Quasi-experimental, 105, 118, 151, 186, 327, 398
Questionnaire design, 139, 148
Questionnaires, 149, 156, 163, 192, 194, 234, 236, 368, 385
scaling procedures, 163
Questions
closed, 124
student, 171

R
Race
See also Ethnic/racial groups
Randomisation, lack of, 118
Range Ratio, 87–89, 92
Rasch model, 130
Ratio, student-teacher, 35, 224
Rational goals behaviour, 250
Rawls, John, 75, 77–87, 94–96
theory, 2, 78, 96
Rawls's Veil of Ignorance, 80–81
Rawlsian, manifesto for EER, 95
Rawlsianism, 81–84, 86, 400
Readiness skills, 366–367
Reading achievement, 250, 338
Reading comprehension, 192, 196, 243, 245, 258, 260, 263, 264, 266, 269, 274
Reading fluency, 192, 195, 196, 260, 336
Reading habits, 343
Reading improvement, 335, 339, 340
Reading instruction, sustained strategic, 263, 274
Reading proficiency, initial, 233
Reading-to-learn, 362
Reciprocity, 79
Redistribution of students, 296, 316
Redress Principle, 95
Reflective equilibrium, 80
Regular contact, schools and parents, 259
Relational climate, 241, 249, 261, 265, 270, 272, 275
Relationships
student-student, 265
teacher-student, 265, 275
Reliability, 55, 123, 125, 130, 214, 318, 368
Religion, 136
Remediate, 81, 84
Repeaters, 187, 191
Representative samples, 147–148, 210, 260
Research designs
mixed-method, 21
Residuals, 306
school, 306
Resilience, 140, 150
Resource allocation, 94, 152, 337
Resources, 35, 54, 80, 82, 101, 102, 104, 128, 152, 209, 256, 289, 299, 301, 305, 307, 316, 341, 347, 394
financial, 90, 275
minimum, 201
pedagogical, 204
Responsibility
civic, 213
collective, 340
Restricted Utility Principle, 83–84
Right to consume, 82
Rights
basic, 83
equal, 83
liberal democratic, 86
Rule Utilitarianism, 77
Rural/urban, 105

S
Sampling, 20, 129, 158–159, 176, 190
  clustered, 159, 176
  weighted, 190
Scaffolding, 37
Scalar invariance, 93, 149, 161–162, 169
School accountability, increasing, 291
School and class characteristics, process, 270
School and teacher effects, 73, 215, 217, 220
School and year effects, 219
School assessment policies, 157
School autonomy, 102, 104, 128
School capacity, building, 103, 332
School career of students, 231, 234, 235, 238, 243, 246–247, 257, 267, 271, 276
School characteristics, 4, 47, 117, 148, 188, 193, 200–201, 235, 242, 248, 254–256, 273, 276, 304
  changing, 276
  effective, 33, 73
  instructional, 254
  intake, 131
  management, 270
  operationalised, 136
  organizational, 241, 272
  particular, 248
  primary, 248, 252
  secondary, 255
  socio-demographic, 105
  teacher-related, 202
School choice, 152
School/class composition
  characteristics, 242, 257, 260, 268, 273, 275
  cognitive, 242
  effects, 196, 258, 259, 268, 301, 302, 315, 318, 394
  ethnic/racial, 257, 259, 261, 273, 289, 290, 294
  favourable, 249, 259
  heterogeneous ability, 256
  high-SES, 257
  low-risk, 275
  peer, 296, 316
  social, 171
  socioeconomic, 171, 287, 290, 294
  student body, 294
  student intake, 251
  variables, 196
School climate, 35, 149–150, 153, 235, 241, 250, 253, 255, 256, 260
  factors, 235
  indicator, 255, 261
School communities, 56, 250, 332, 340
  diverse, 340
School composition, 242, 249, 251, 253, 261, 270, 273, 294
  characteristics, 258, 260
  effect of, 259
  ethnic, 259
  indicators, 258
  literature, 301
  primary, 260
School conditions, 305
  material, 201
School configurations, 271
School context, 24, 331, 338, 398
  changed, 351
School context/composition, 24, 148, 216, 219, 223, 226, 242, 254, 256, 261, 331, 338, 351, 398
School culture, 217, 272
School denomination, 250, 251
School disorder, 302
School district/local authority level, 101, 106, 292, 317
School Effectiveness and School Improvement (SESI), 14, 233, 276
School effectiveness levers, 211
School effects, 2, 3, 21, 73, 92, 94, 189, 200, 215–216, 219–220, 225, 233, 239, 245–246, 287, 302, 318, 387
  long-term, 248
  short-term, 239, 243–247
School enrolment, 213
School environment, orderly, 256
School equipment, 187, 188, 201
School evaluation, 38, 55, 152–153, 155, 157, 165–168, 173, 176, 212, 215
School factors, resource-based, 20
School failure, 361
School fees, 191
School funding, 209
School improvement, 22–23, 40, 55, 61, 103, 110, 111, 117, 119, 151–152, 156, 167, 177, 328, 332, 386, 389
  evidence-based, 2
  feedback-based, 152
  perspective, 137, 401
  policies, 156
  research, 24, 72–73, 75, 118–119, 215, 327, 383, 400, 401
School improvement capacity, 102
School improvement research, international, 400
School Improvement, Dynamic Approach (DASI), 53, 60, 353, 400
School inputs, 217, 224, 301, 303, 305, 309, 313

School intakes, 71, 76, 106, 118, 131–132, 215, 219, 249, 390
  comprehensive, 93
  disadvantaged, 214
School integration, 289
School leader behaviour, open-systems, 252
School leader characteristics, personal, 4
School leader perceptions, 271
School leaders, 44, 235, 250, 253, 271, 328, 329, 337, 341–343, 349, 353, 392
  content knowledge, 336
  critical self-reflection, 332
  developing relationships, 332
  human relations behaviour, 250–251
  roles of, 4, 22
  successful, 4, 331
School leadership, 4, 241, 253, 260, 272, 327–329, 335
  effect of, 253
  indirect effects of, 252
School learning environment (SLE), 52–53
School-level factors, 38, 40, 72, 136, 271, 275
School management, 49, 51, 53, 105, 115, 127, 211, 214, 252, 333
School networks, 1, 9, 16, 18, 102–104, 106, 117–119, 335, 388, 393, 396, 399, 400
  bottom up, 102
  effectiveness, 103–104
  effects, 387
  establishing clear goals, 104, 118
  national programme, 117
  successful, 114
  support, 118
School organization, 235
  development-oriented, 253, 272
  processes, 261
  research, 231
School performance, 219, 225, 292
School performance feedback systems, 151
School performance, evaluating, 214
School policy, 40, 42–43, 52–55, 59, 149, 156–157, 164–165, 171, 173, 222, 390
  assessment-related, 155
  evaluation/accountability-related, 172
School policy indices, 171
School practice characteristics, 249, 256
School practices, 55, 170, 214, 253, 261, 293, 296, 303, 308, 309, 313, 316, 319
  effective, 316
  experiential, 254
  improving, 43, 104, 296
  socioeconomic-based, 295
School processes, 4, 21, 214, 249, 251, 254, 256, 258, 261, 271, 327
School progress, monitoring, 157
School quality, 211, 213, 214, 269
School-relatedness, 271
School rules, 54
Schools
  average achievement, 293
  coeducational, 259
  community, 254
  comparator, 107–108, 118
  differential effects of, 47, 242
  disadvantaged, 209, 211
  effectiveness status, 59, 138
  fee-paying, 80
  high average ability level, 258
  high performing, 106, 110
  indigenous, 340
  in-take, 105, 118
  low-SEC, 289, 290, 296, 307, 310, 312, 316, 319
  low-SES, 257
  majority-White, 289
  medium-SEC, 307, 311, 312
  migrant, 129
  mixed-SES, 260
  outstanding, 107
  private, 80, 190, 250, 271, 297, 317
  ranking, 152
  selective, 80
  single-sex, 259
  social-constructivist, 254
  special(ist), 129, 331
  successful, 83, 256, 330, 333, 350
  tutors/private tutoring, 135
  underperforming, 94, 106
  unfavorable, 270
  vocational, 256
  within-school average, 372
School selection effects, 398
School self-evaluation, 43, 220, 400
School SES, low average, 257
School size, 35, 105, 194, 196, 251–252
  low, 270
School systems
  equitable, 85
  public, 319
  self-improving, 107
School turnaround, 108
School types, 124, 190, 196, 202, 218, 234, 251, 259, 261, 390
School variation, 16, 140, 370

School-to-school support programme, 106, 112
School-wide practices, 345
Science literacy, 168, 170
Science teaching, 171, 172
Seatwork, 51
Secondary analyses, 20–21, 128, 148, 385
Secondary education, 60, 108–110, 186, 188, 200–201, 209, 216–219, 231, 234, 235, 237–247, 249–250, 255, 259, 263, 268, 271, 273, 345, 346
Segregation, 288, 294, 317
  neighborhood, 288
  racial, 289, 290, 318
  school, 288, 290, 293, 297, 316
  socio-economic, 4
  vertical, 366, 377
Selection biases, school, 244, 297, 301, 304, 307
Selection effects, 304
Selection, academic, 210
Self-concept
  academic, 193, 244–246
  general, 260
  pupil, 190, 192, 248, 388
Self-confidence, 48, 250
Self-determination, 346
Self-determination theory, 262
Self-efficacy, 246, 263, 264, 272
  academic, 263, 265
  general, 272
Self-interest, 84, 94, 95
Self-ratings, teacher, 301
Self-regulated learning, 37
Self-regulation, 48, 50, 154, 248, 263, 265, 266, 364
Sen's capability theory, 85, 86
Senate weight, 159
Sense-making functions of organisations, 102
Shadow coaching, 348
Shadow education, 135
Shanghai, 11, 20, 86, 129, 135, 211, 214
Shared decision making, 333
Shared teacher-student control, 264
SiBO project, 235–237
Sidgwick, Henry, 75
Singapore, 11, 86, 168
  mathematics teaching approaches and textbooks, 11
Single-country studies, 9, 12, 16, 17, 19, 21
Skills
  employment, 292
  letter knowledge, 364
  linguistic comprehension, 363
  math problem-solving, 315
  natural, 81
  new learning, 276
  non-cognitive, 295
  phonological processing, 363
  psychomotor, 38
  reading, 363
  school-entry, 366
  self-regulation, 276
  social, 292
Skills based approach, 126
Skills levels, average, 366, 372
SLE, *see* School learning environment (SLE)
Social backgrounds, 36, 48, 71, 132
Social capital, 261, 301
Social capital theory, 102, 118
Social class, 47
Social development, 46, 210
Social disadvantage, 106, 399
Social goods, 81
Social groups, 47
Social mobility, 81–82, 85, 96
Social networks, 103
Social processes, 293
Society for Research on Educational Effectiveness (SREE), 11–12, 14
Sociocultural background, 260
Socio-economic background, 34, 35, 38, 83, 132, 150, 290, 316
Socioeconomic segregation, 4, 287, 290, 294, 296, 316, 319
Socio-ethnic background, 269
South America, 22, 168
South Asia, 22
Southeast Asia, 22
Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), 200
Southern Hemisphere, 123
Spain, 20, 102
Special Educational Needs, 105–106, 388
Specific-rule Utilitarianism, 77
Stability
  school effects, 82, 220, 224, 233
  social, 79
Staff collaboration, 333
Staff development, 332
Staff turnover, 340
Standardised assessments, formative use, 156–157, 165, 172
Standardized monitoring, 155
Standardized test use, 155–158, 164–167, 172
Standards
  educational, 103, 105, 108, 110, 150, 153, 155, 187, 212, 225
  national, 152, 165, 171
Stand-downs, 344

State-of-the-art articles, 18, 231, 276
Streaming, 80, 94
Strengths-based approach, 341
Strong achievement orientation, 261
Structural arrangements, 341
Structural barriers, 288
Student assignment practices, 317
Student background, 16, 20, 35, 46, 105, 173, 216, 249–251, 264, 267, 387
Student behaviour, maladaptive, 265, 266
Student counselling, 255, 256
Student level factors, 36–37, 44–46, 48–49
Student misbehavior, 56, 316
Student mobility, 292
Student outcomes
  negatives, 16
  procrastination, 265
Student participation, independent, 254
Student-perceived practices, 172, 244
Student portfolios, 213
Student ratings, 302
Student recruitment, 244, 249
Student underperformance, 289
Students
  differences and development, 256, 261
  high-SES, 260, 274
  practice and application opportunities, 51, 59
  pre-literacy skills, 361, 363, 367, 370, 376, 377
Student-teacher ratio
  See also Pupil-teacher ratio
Subject matter acquisition, 255, 261
Subject matter experts, 234
Summative assessment, lower self-esteem, 153
Supply side policies, 133
Suspensions, 294, 308
Sustainability, 58, 119, 331, 349
Sustainable approaches, 107
Sustainable change, 347, 349
Sustainable Development Goals (SDGs), 361, 376, 378
Sweden, 102, 165, 330–331
Switzerland, 122
System effectiveness, 61, 233, 396
System improvement, 18
System leadership, 108, 115, 117–118, 328
System level, 15–16, 38, 40, 47, 56, 61, 86, 148, 149, 155, 175, 387, 395
System reform, 137, 353

T
Taiwan, 122
Task orientation, 263
Task performance, 154
Task processing, 154
Tasmania, 331, 340, 343
Teacher behavior, 4, 13, 48, 50, 58, 131, 140, 232–233, 236, 262, 264, 276, 338
Teacher beliefs, 49
Teacher characteristics, 196, 235, 242, 272, 276
Teacher collaboration, 223, 241, 251, 253, 256, 272
Teacher commitment, 272
Teacher control, 264, 267
Teacher cooperation, 249, 256, 261, 270, 275
Teacher-developed test use, 156, 174
Teacher development, 211, 216, 223–226
Teacher distribution, 209
Teacher effectiveness, 13, 22, 35–36, 47, 49, 136, 149, 215, 216, 232, 234, 394–395, 397
Teacher effects, 2, 17, 18, 73, 130, 215, 217, 220–221, 248, 387
Teacher efficacy, collective, 261
Teacher factors, 16, 49–52, 58, 59, 241, 261
Teacher gender, 196
Teacher interpersonal relationships, 274
Teacher involvement, 267, 274
Teacher leadership, 328
Teacher learning processes, 271
Teacher level, 19, 220, 235
Teacher morale, 301
Teacher neglect, 266
Teacher networks, cross-institution subject, 390
Teacher observation instrument, 236
Teacher participation in decision-making, 253
Teacher professional development, 223
Teacher proximity, 266
Teacher quality, 215–216, 301, 333
Teacher research groups, 211, 224, 401
Teacher salaries, 134
Teacher skills, time management, 51
Teacher training, 164, 196, 201–202, 211, 377, 400
Teacher trust, 266, 275
Teacher unions, 165
Teachers
  authoritarian, 266
  dominant cooperative, 266
  effective, 44, 50–51, 73
  subject, 235
  subject proficiency level, 203
  tolerant/authoritative, 266
  untrained, 187
  value-added effects, 3
Teachers interpersonal behaviours, 266
Teachers interpersonal competences, 49
Teachers' learning, 211, 349

Teachers' practice, 211, 272
Teaching
  direct, 37, 50
  high-quality, 37
  inquiry-based, 171, 172
  mixed-ability, 82
  multicultural, 273
  quality of, 37–39, 53, 59, 114, 115, 149, 154, 172, 319, 337
  structured, 136
Teaching activities, 172
Teaching and Learning International Survey (TALIS), 20, 131, 134, 140, 216, 385
Teaching and learning situation, 40
Teaching approach, active, 37, 39
Teaching Assistants (TAs), 340
Teaching competitions, 223
Teaching effectiveness, 276
Teaching effects, 150
Teaching factors, 19, 58, 138
Teaching learning strategies, 263, 274
Teaching materials, 187, 188, 201, 204
Teaching methods, 130, 223, 255, 261
  student-centred, 210
  whole class interactive, 122
Teaching modelling, 49–50
Teaching practices, 22, 35, 40, 44, 52, 54, 56, 159, 163, 173, 194, 202, 269, 273, 339, 341–342, 348
  effective, 40, 53, 131
Teaching processes, 35, 37
Teaching School Alliances (TSA), 107, 111–114, 117
Teaching skills, generic, 37, 58
Teaching staff cooperation, 255, 261
Teaching styles, 233, 254, 276
  learner-centered, 267, 272
Teaching time, 51, 53, 59
Teaching to the test, 293
Team planning, 340
Test administration
  technology-based, 158
  time, 148
Test anxiety, 246, 264
"Testeritis", 165
Test use, 173
Testing practice, 165
Testing purposes, 157, 158
Textbooks, 188, 192, 201, 202, 204
Theil element, 89
Theory development, 233
Thinking skills, higher-order, 50, 276
Thinking styles, 37, 44, 45, 47–48
Third Millennium Schools (book), 15
'Time in school', 135
Time on task, 49
Time on task, maximizing, 269
Time trends, 217, 219, 225
TIMMS data, 21
Total effect, 303, 307, 309, 313, 316
Total Utilitarianism, 75
Track choice, 245
Tracks (academic, general, vocational), 234, 241, 244, 251, 259, 271
Training activities, 5, 201, 223
Transfers, student, 93–94, 293
Transition, seamless, 110
Transitions, educational, 110, 235, 259, 362
Translation of instruments, 123, 148, 233
Treaty of Waitangi, 346
Trend studies, 149
Trends in International Mathematics and Science, 124, 147
Trends in International Mathematics and Science Study (TIMSS), 20, 35, 76, 124–127, 130, 139, 147, 149, 154, 158, 175, 186, 200
Tribal level, 345
Truancy rates, 294
Trust, 103–104, 111, 275, 289, 340
  relational, 332
Tuition fee, 220
Turkish ethnic/language background, 259
Tutoring
  extra, 80
  private, 71
  supplementary private, 135
21st century skills, 276

U

UK and Continental Europe
  See also Europe and United Kingdom
UKAID, 216
Underachievement, 254, 353
Underestimation, 249
Underperformance, 106, 289, 290, 331
Underrepresentation, 402
Understanding, contextual, 24
Unidimensional requirement, 130
United Kingdom, 122, 165–167, 328, 330
United States, 133, 139, 232, 289, 291, 311, 319, 361

University of Bristol (UK), 3, 216
University of Kisangani (DRC), 3, 186, 190, 200–202, 204
University of Leuven, 235
Uruguay, 4, 361–362, 365, 367–368, 376–377
  teachers, 5, 376
Utilitarianism, 74–80, 82–84, 86, 94–95
Utilitarianism (book);Utilitarianism (book, 76
Utility
  aggregate, 75, 79, 84
  measuring, 75

V
Validity
  cross-cultural, 20
  cross-national, 123
  cultural, 131
  model, 2, 34–37, 39, 57, 61
Values
  community, 352
  competing, 94
  cultural, 272
  intrinsic, 94, 246, 254, 263, 264
  marginal utility, 93
  missing, 176
  political, 332
  positive
    for learning, 56
  professional, 332
  shared, 103, 211
  societal, 40
  universal, 332
Variables
  alterable, 131
  attitude, 217
  class-related, 197
  contextual, 35
  country moderator, 18
  country/area, 19
  finer-grained country, 19
  individual-level, 302
  input, 35
  instructional, 21
  primary selection, 304
  process, 35, 217, 224, 267
  pupil level, 108
  school-level, 170
  school-related, 196
Variance
  between-class, 196
  school-level, 114, 172
  unexplained, 36
  within-classroom, 372
  within-school, 306
Variance components, 374
Variation
  accounting for, 49
  between-classroom, 366
  child-level, 4
  classroom-level, 158
  contextual, 138
  global, 135
  localised, 401
  natural, 10, 12
  observed random, 219
  regional, 3
  school-level, 171
  systemic, 159
  total, 374
  within-region, 23
Variation in annual growth, 374
Variation in student outcomes, 265, 396
Veil of Ignorance Principle, 95
Verbal reasoning, 363
Victoria, 331
Violent conflict, response to, 24
Vision, 261, 350, 353
  building, 339
  common, 332
  overarching, 332
  shared, 256, 335
Vocabulary, 336, 363, 364
  receptive, 365
Vocational education, 245, 251, 256, 262
Vulnerability, 367
  concentration plot, 377
Vulnerable children, 362, 366, 374
  identifying, 5
Vulnerable students, 377–378

W
Wales, 134
Wellbeing, 244, 246, 248, 258, 262, 269, 275, 346, 388, 399
Wellbeing indicators, 265
Wellbeing outcomes, 290
Wellbeing, general professional, 272
Western democracies, 96
Western nations, 16, 86
Western systems, 11
Whānau, 345, 347
What Works Clearinghouse (WWC), 11
White Paper (England), 107
White schools, 290

White-majority schools, 290
Whole-school approach, 342
Willingness to produce, 82
Within-class variance, estimates, 374
Within-country, 169
Within-nation, 138
Within-school, 89
Within-school comparisons, 86
Within-School Imperative, 95
Within-school terms, 90
Word decoding, 192, 363
Word recognition, 363
Work, creative, 213
Working-class life, 82
Works Clearinghouse, 11
World Bank, 10, 122, 126, 201, 385
WWC, *see* What Works Clearinghouse

Y
Year effects, 219