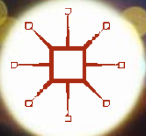


EDITED BY
CHAD R. LOCHMILLER

**COMPLEMENTARY
RESEARCH
METHODS *for*
EDUCATIONAL
LEADERSHIP &
POLICY
STUDIES**



Complementary Research Methods for Educational Leadership and Policy Studies

Chad R. Lochmiller
Editor

Complementary
Research Methods
for Educational
Leadership and Policy
Studies

palgrave
macmillan

Editor

Chad R. Lochmiller
School of Education
Indiana University
Bloomington, IN, USA

ISBN 978-3-319-93538-6 ISBN 978-3-319-93539-3 (eBook)
<https://doi.org/10.1007/978-3-319-93539-3>

Library of Congress Control Number: 2018953698

© The Editor(s) (if applicable) and The Author(s) 2018

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover image © Aciform / Alamy Stock Photo

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

PREFACE

AIMS AND OBJECTIVES OF THE BOOK

Complementary Research Methods for Educational Leadership and Policy Studies is an edited volume that discusses qualitative, quantitative, and mixed methods research approaches. The scholars within this volume present methodologies that have been or are increasingly being used to conduct research related to educational leadership and policy. Each chapter aims to introduce the reader to a specific methodology or perspective, relate the methodology or perspective to the field of leadership or policy, and offer recommendations for scholars who may be unfamiliar with the method discussed. Throughout the book, the contributors draw upon their own research and that of other scholars to illustrate how the method can be used. Collectively, the authors help readers identify new and emerging approaches to conducting research related to leadership and policy.

Although not a handbook, this book is certainly situated among recent comprehensive volumes describing research methods used to study leadership and policy specifically (Bascia et al. 2005; Fuhrman et al. 2007; Ladd and Fiske 2008; Sykes et al. 2009; Young and Crow 2017), and education more generally (Green et al. 2006). Recognizing the increasing connection between policy and leadership, this volume approaches the task of highlighting various research approaches somewhat differently than the extant literature. Instead of viewing leadership and policy as separate domains, the volume instead views these fields as complementary and thus presents methods that allow scholars to envision how shared lines of inquiry might be initiated. One of the primary questions addressed in this

volume is thus: How can we use research methods to expand our collective understanding of the ways in which policy is developed at the legislative level and implemented in specific localities through leadership action?

The research methods presented within this volume demonstrate some of the possible methods that researchers might use to begin working in collaboration. Indeed, while presented as discreet methods, it becomes clear throughout the volume how these methods might support a complementary research agenda that allows us to understand the interaction(s) between policy and practice. Indeed, one of the central goals for this volume was to advocate for expanded dialogue between the fields of educational leadership and policy at a methodological level. In advocating for this expansion, I note that scholars would be well-served by engaging in research that utilizes the strengths of different research methods to explore related research questions situated broadly under the umbrella of educational leadership and policy. The aim is to expand our understanding of policy issues by enlisting multiple, complementary methods to examine the increasingly complex context within which these issues are necessarily situated. This does not suggest that we simply adopt mixed methods, as has been advocated elsewhere, but instead that we utilize research methods in tandem to pull back the veritable ‘layers’ of the issue, topic, or concern.

OVERVIEW OF THE EDITORIAL PROCESS

As the editor of this volume, I enlisted the assistance of blind peer reviewers throughout the editorial process. The reviewers all had expertise in the substantive and methodological areas discussed in each chapter. After each chapter, author(s) proposed an outline for their chapter; I provided an initial round of feedback. This feedback ensured the chapters had a similar structure, depth of content, and offered an accessible introduction for both a novice and an experienced scholar. Following this feedback, chapter authors prepared drafts of their chapters. I sent each draft to a minimum of two peer reviewers. Each reviewer provided feedback on the substance, technical accuracy, and pedagogical quality of the chapter. I then asked the authors to revise their chapters based on the reviewer’s feedback. Depending on the extent of feedback offered, some chapters underwent multiple rounds of review. After approval by each reviewer, I then performed a final review of the chapter and asked authors to make minor adjustments to ensure the volume was well aligned. I was assisted throughout the editorial process by two Ph.D. students who have expertise in leadership and policy studies.

STRUCTURE OF THE VOLUME

The volume consists of three parts. The first part focuses on qualitative and critical approaches to the study of leadership and policy. In this section, the authors describe research methods ranging from the traditional multi-site case study to more emergent approaches such as discourse analysis. Two chapters highlight the value of adopting critical perspectives to unpack existing inequities. The second part focuses on quantitative approaches to the study of leadership and policy. In this section, the authors describe statistical techniques that, taken together, allow scholars to investigate a variety of issues related to leadership and policy. For instance, the volume describes secondary data analysis, matching strategies for causal inference, and value-added models. In addition, authors also discuss the value of social network analysis, which is a methodology that has become increasingly popular in policy research. Finally, in the third section, the authors offer a more applied perspective on leadership and policy research. In doing so, the authors describe mixed methods research, program evaluation, and discuss the possibilities for collaborative, mixed methods research across international contexts.

Bloomington, IN, USA

Chad R. Lochmiller

REFERENCES

- Bascia, N., Cumming, N., Datnow, A., Leithwood, K., & Livingstone, D. (Eds.). (2005). *International handbook of education policy*. Dordrecht: Kluwer.
- Fuhrman, S. H., Cohen, D. K., & Mosher, F. (Eds.). (2007). *The state of education policy research*. Mahwah: Erlbaum.
- Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2006). *Handbook of complementary methods in education research*. New York: Routledge.
- Ladd, H., & Fiske, E. (Eds.). (2006). *Handbook on research in education finance and policy*. Mahwah: Erlbaum.
- Sykes, G., Schneider, B., & Plank, D. N. (Eds.). (2009). *Handbook of education policy research*. Washington, DC: American Educational Research Association.
- Young, M. D., & Crow, G. M. (Eds.). (2017). *Handbook of research on the education of school leaders* (2nd ed.). New York: Routledge.

ACKNOWLEDGEMENTS

I wish to acknowledge the contributions of the chapter authors. Without their willingness to share their expertise, this project would not have been possible. In addition, I acknowledge Palgrave Macmillan, their editorial team, the anonymous peer reviewers who provided helpful feedback on the early conceptualization of this volume. Finally, I wish to thank two of my doctoral students in Educational Leadership and Policy Studies at Indiana University Bloomington, Jennifer Karnopp and Colleen Pawlicki, for their editorial assistance.

This book also benefitted greatly from the expertise of colleagues who served as peer reviewers. In addition to five anonymous reviewers, the reviewers for this book included:

- Aaron Butler, Kentucky Department of Education
- Thomas Crist, University of Bridgeport
- Ana Elfers, University of Washington
- Liz Farley-Ripple, University of Delaware
- Matt Finster, Westat
- Gordon Gates, Washington State University
- Emily Hodge, Montclair State University
- Patricia Kannapel, C.N.A.
- Jennifer Karnopp, Indiana University Bloomington
- Sharon Kruse, Washington State University
- Leslie Locke, University of Iowa
- Christopher Lubienski, Indiana University Bloomington
- Andrew McEachin, RAND Corporation

- Hajiime Mitani, Rowand University
- Lauren, Monet, University of Tennessee Knoxville
- Trena Paulus, University of Georgia
- Mary Piontek, Marzano Research
- Thomas Sugimoto, Center for Evaluation & Education Policy
- Rob Toutkoushian, University of Georgia
- Francesca White, Indiana University Bloomington
- Bethany Wilinski, Michigan State University
- Nedim Yel, Center for Evaluation & Education Policy

Finally, I acknowledge my students in research methods and policy courses who ultimately inspired this volume. Their thought-provoking questions and requests for support helped identify the topics and ideas that compelled me to bring together the scholars in this volume and invite them to share their expertise and knowledge.

CONTENTS

1	Complementary Research Methods: An Introduction and Overview	1
	Chad R. Lochmiller	
Part I	Qualitative and Critical Approaches to the Study of Educational Leadership and Policy	17
2	Qualitative Research in Educational Leadership Studies: Issues in the Design and Conduct of Studies	19
	Jeffrey S. Brooks and Anthony H. Normore	
3	The DIVE Approach: Using Case-Ordered Meta-Matrices and Theory-Based Data Displays to Analyze Multiple Case Study Data	33
	Susan Bush-Mecenas and Julie A. Marsh	
4	Language-Based Methodologies for Policy and Leadership Research	57
	Jessica Nina Lester and Justin Paulsen	
5	Doing Critical Policy Analysis in Education Research: An Emerging Paradigm	79
	Michelle D. Young and Sarah Diem	

6	Critical Research Perspectives in School Leadership: Putting Dignity and Humanity at the Center	99
	Irene H. Yoon	
7	The Potential of (Participatory) Action Research for School Leaders, Local Policy Makers, and University-Based Researchers	121
	Meagan Call-Cummings and Melissa Hauber-Özer	
Part II	Quantitative Approaches to the Study of Educational Leadership and Policy	141
8	Secondary Data Analysis in the Field of Educational Leadership and Policy Studies	143
	Angela Urick	
9	Matching Strategies for Causal Inference with Observational Data in Education	173
	Yongnam Kim, Stanley A. Lubanski, and Peter M. Steiner	
10	Using Quantitative and Qualitative Methods to Study the Content and Effects of Curriculum Materials	193
	Morgan Polikoff, Shauna E. Campbell, and Shira A. Korn	
11	Value-Added and Growth Models in Education Research	213
	Cassandra M. Guarino	
12	Social Network Analysis Methods in Educational Policy Research	231
	Kara S. Finnigan, Daniela E. Luengo-Aravena, and Kim M. Garrison	
13	Essential Steps to Assessing a School System’s Fiscal Health	253
	Joshua R. Zender, Kenneth A. Smith, and John R. Kurpierz	
14	Evaluating Education Policy & Program Costs	281
	Tammy Kolbe and Rachel C. Feldman	

Part III	Mixed, Applied, and Collaborative Approaches to the Study of Educational Leadership and Policy	305
15	Using Mixed Methods to Inform Education Policy Research Colleen E. Chesnut, John H. Hitchcock, and Anthony J. Onwuegbuzie	307
16	Program Evaluation Methods for PK-12 Education Liz Hollingworth	325
17	International Network as Sites for Research on Successful School Leadership Christopher Day and David Gurr	341
18	Taking Stock of Complementary Methods: The Perpetual Quest for Good Research Methods for Educational Leadership and Policy Carolyn J. Riehl	359
	Index	383

NOTES ON CONTRIBUTORS

Jeffrey S. Brooks is Associate Dean for Research and Innovation and Professor of Educational Leadership in the School of Education at the Royal Melbourne Institute of Technology University (RMIT), based in Bundoora, Australia. His research focuses on the way school leaders influence (and are influenced by) socio-cultural dynamics such as racism, globalization, social justice, and school reform.

Susan Bush-Mecenas is a doctoral candidate and provost/final year fellow at the University of Southern California's (USC) Rossier School of Education. Her research interests include organizational learning, district reform, district and school capacity building, and accountability policy.

Meagan Call-Cummings is Assistant Professor of Qualitative Methods at George Mason University. She writes on participatory, feminist, and critical approaches to qualitative inquiry, often focusing on methodological questions of validity, ethics, and reflexivity.

Shauna E. Campbell is a doctoral student in K-12 Education Policy at the USC Rossier School of Education. Her work has focused on the evaluation, selection, and implementation of curriculum materials with particular interest in the role of the Common Core State Standards on these processes.

Colleen E. Chesnut is an assistant research scientist at the Center for Evaluation and Education Policy at Indiana University. She specializes in policy research and evaluation, utilizing both qualitative and mixed

methods approaches. Her areas of focus include education leadership, teacher evaluation, and educational equity for English learners.

Christopher Day is Professor of Education in the School of Education, University of Nottingham; Professor of Educational Leadership in the School of Education and Social Work, University of Sydney; and Chair Professor of Educational Leadership in Beijing Normal University. Recent publications include: *Teachers' Worlds' and Work: Understanding Complexity, Building Quality* (2017); *A Decade of International Research on School Leadership* (2016); and *Successful School Leadership: Lessons from the Field* (2014).

Sarah Diem is an associate professor in the Department of Educational Leadership and Policy Analysis at the University of Missouri. Her research focuses on the socio-political and geographic contexts of education, paying particular attention to how politics, leadership, and implementation of educational policies affect outcomes related to equity, opportunity, and racial diversity within public schools.

Rachel C. Feldman is a postdoctoral fellow at Northwestern University's School of Education and Social Policy. Her research explores schools' organizational responses to federal and local policy initiatives, and how these organizational structures influence teachers' work.

Kara S. Finnigan is an associate professor of Educational Policy at the University of Rochester. She has written extensively about accountability, school and district improvement, and school choice. Finnigan's research blends perspectives in education, sociology, and political science; employs qualitative and quantitative methods, including social network analysis; and focuses on urban contexts.

Kim M. Garrison is a doctoral student and graduate teaching assistant at the University of Rochester. Her research focuses on exploring social, cognitive, and structural work factors that impact teachers' professional identity and well-being. Specifically, she looks at how social and emotional capital theories can contribute to understanding teachers' work lives.

Cassandra M. Guarino is Professor of Education and Public Policy at the University of California Riverside. Her research focuses on teacher quality, teacher labor markets, school choice, and issues in which health and education are linked. Recent work has included several studies related to value-added measures of teacher performance, teacher effectiveness in the early grades, school choice, teacher mobility, and special needs identification.

David Gurr is Associate Professor in Educational Leadership at The University of Melbourne and has a 37-year background in secondary teaching, educational psychology, school supervision, and research in educational leadership.

Melissa Hauber-Özer is a doctoral student in international education at George Mason University. Her research focuses on literacy among adult refugees who have fled Syria and are settling in Turkey and other countries. Her research takes critical, participatory, and feminist forms.

John H. Hitchcock is Associate Professor and Director of the Center for Evaluation and Education Policy at Indiana University. His scholarly area of interest is evaluating interventions for children with special learning needs, as well as refining methods for performing such evaluations.

Liz Hollingworth is Professor of Educational Leadership and Director of the Center for Evaluation and Assessment at the University of Iowa. Her research focuses on issues of leadership and assessment; in particular, how school reform policies affect classroom practice, evaluation, and school administration.

Yongnam Kim is a doctotal candidate in the Department of Educational Psychology at the University of Wisconsin-Madison. His research focus is on causal inference, including quasi-experimental designs and graphical causal models.

Tammy Kolbe is Assistant Professor of Educational Leadership and Policy Studies at the University of Vermont. Her research focuses on the allocation of educational resources (e.g., teacher quality, funding, and learning time) and the cost effectiveness of educational policies and programs.

Shira A. Korn is a doctoral student in K-12 Education Policy at the USC Rossier School of Education and Master's student in Public Policy at the USC Price School of Public Policy. She is interested in the role of federal, state, and local policies in improving educational outcomes for underserved students.

John R. Kurpierz is a doctoral candidate at the Schulich School of Business, York University. His research focuses on the interconnections between politics, psychology, and the financial reality of accounting.

Jessica Nina Lester is Associate Professor of Inquiry Methodology in the School of Education at Indiana University. Much of her research is positioned at the intersection of discourse studies and disability studies. Her work has been most recently published in journals such as *Qualitative Inquiry*, *Qualitative Research*, and *Discourse Studies*.

Chad R. Lochmiller is the Editor of *Complementary Research Methods for Educational Leadership and Policy Studies* and is Assistant Professor of Educational Leadership and Policy in the School of Education at Indiana University Bloomington. His research focuses on issues related to K-12 educational leadership and policy, most notably those related to educational reform, school finance, instructional supervision, and leadership coaching. Additionally, Lochmiller has also written about research methods suitable for research in leadership. His research has appeared in leading journals such as *Educational Administration Quarterly* and the *Journal of Educational Administration*.

Stanley A. Lubanski is a graduate student in the Department of Educational Psychology at the University of Wisconsin-Madison. He is interested in causal inference and quasi-experimental designs. Currently, his research focuses on how adding design elements (like propensity score methods) to an experiment affects the estimation of standard errors.

Daniela E. Luengo-Aravena is a doctoral student at the University of Rochester, holding a BA and MA in Economics from the University of Chile. She uses economic theories to analyze educational policy issues and is studying how the composition of a classroom, for example, gender, ability, and socio-economic status, impacts student outcomes.

Julie A. Marsh is Associate Professor of Education Policy at the University of Southern California's Rossier School of Education and specializes in research on K-12 policy. Her research blends perspectives in education, sociology, and political science. Her research focuses on the implementation and effects of accountability and instructional reform policies, including the roles of central office administrators, intermediary organizations, and community members in educational reform and the use of data to guide decision making.

Anthony H. Normore is Professor Emeritus of Educational Leadership at California State University–Dominguez Hills in Los Angeles. His research focuses on urban leadership growth and development in the context of ethics and social justice. He holds a PhD from the University of Toronto.

Anthony J. Onwuegbuzie is Professor in the Department of Educational Leadership at Sam Houston State University, teaching doctoral-level courses in qualitative, quantitative, and mixed research. Further, he is a distinguished visiting professor at the University of Johannesburg. With a h-index of 83, he has authored/co-authored more than 400 published works.

Justin Paulsen is a doctoral student in Inquiry Methodology in the School of Education at Indiana University. Having received an MPA/MA in Russian Studies and International Development, he uses a variety of methodologies in evaluation projects spanning a range of topics. He is interested in adapting methodologies for use in novel, practical applications.

Morgan Polikoff is Associate Professor of Education at the USC Rossier School of Education, where he researches the design, implementation, and effects of standards, assessment, and accountability policies. His current work on curriculum materials is supported by the National Science Foundation, Bill and Melinda Gates Foundation, and WT Grant Foundation.

Carolyn J. Riehl is Associate Professor of Sociology and Education Policy at Teacher's College, Columbia University. Riehl specializes in organizational dynamics in schools and school systems, leadership and policy, and research methods. Her work has appeared in *Educational Researcher*, *Sociology of Education*, *American Journal of Education*, *Review of Educational Research*, *American Educational Research Journal*, *Educational Administration Quarterly*, and in numerous edited volumes.

Kenneth A. Smith is Associate Professor of Accounting at Central Washington University and a former school district board member. His teaching and research focus on measuring the financial and non-financial performance of government organizations. He has previously published chapters in *Advanced Accounting*, textbooks and several other academic journals.

Peter M. Steiner is an associate professor in the Department of Educational Psychology at the University of Wisconsin-Madison. His research focuses on causal inference, quasi-experimental designs, and vignette experiments.

Angela Urick is an assistant professor at the University of Oklahoma in the Department of Educational Leadership and Policy Studies. She studies leadership for school improvement through school structures, perceptions of leadership, teacher outcomes, and student opportunities to learn.

Irene H. Yoon is an assistant professor in the Department of Educational Leadership and Policy in the College of Education at the University of Utah. Yoon's research explores dynamics of race, class, gender, and ability in classroom discourse, professional learning, and school leadership and school improvement.

Michelle D. Young is Professor of Educational Leadership at the University of Virginia. Young's scholarship focuses on how university programs, educational policies, and school leaders can support equitable and quality experiences for all students and adults who learn and work in schools. Recent publications include: *Handbook of Research on the Education of School Leaders*, and *Leveraging Standards to Promote Program Quality*.

Joshua R. Zender is Assistant Professor of Accounting and MBA program director at Humboldt State University, at California State University campus. He specializes in public financial management practices and has previously served as a finance manager in local government. He is licensed as a Certified Public Accountant (CPA).

LIST OF FIGURES

Fig. 2.1	Research dissemination products for qualitative studies of educational leadership. (Source: Brooks and Normore 2015)	28
Fig. 3.1	The DIVE approach and software tools	39
Fig. 3.2	Illustrative unpublished conceptual framework used during analysis of the Marsh et al. (2017) study	41
Fig. 3.3	Illustrative unpublished case-ordered meta-matrix used during qualitative data analysis for Marsh et al. (2017)	44
Fig. 3.4	Illustrative unpublished case-ordered meta-matrix used during analysis for Marsh et al. (2015)	46
Fig. 3.5	Case visualization from Marsh et al. (2015). Reprinted by permission of SAGE Publications	49
Fig. 3.6	Case visualization used during analysis related to Marsh et al. (2017). (Note: Three public entities are responsible for charter contracting in New Orleans—the RSD, OPSB, and the Louisiana Board of Elementary and Secondary Education (BESE). Several different types of schools operate in New Orleans, including traditional schools overseen by OPSB, charters in a CMO network, and single-site charters)	50
Fig. 9.1	Illustration of PS techniques (a) Initial status. (b) PS-matching (c) PS-stratification. (d) PS-weighting. (Note. “T” denotes treated units, while “C” denotes control units)	179
Fig. 12.1	Basic social structures: dyad and triad	233
Fig. 12.2	Different ways of representing relationships using sociograms	235

Fig. 12.3	Social network of state education agencies and intermediary agencies' connections around common core ELA resources. (Note: SEAs are circles; white are CCSS SEAs, and black are non-CCSS SEAs; grey squares are intermediary organizations, tie thickness represents number of resources used from an agency and node size represents a measure of the agencies' centrality in the network. Source: Hodge et al. (2016, p. 7))	241
Fig. 16.1	Sample logic model for a new reading program	329
Fig. 16.2	Five phases of program evaluation	335

LIST OF TABLES

Table 8.1	Main surveys for educational leadership and policy secondary data analysis	145
Table 8.2	Teachers and principals studied across six main surveys	149
Table 8.3	Students studied across six main surveys	151
Table 8.4	Classrooms and schools studied across six main surveys	152
Table 8.5	Post-secondary transitions studied across six main surveys	154
Table 8.6	National policy issues studied across six main surveys	155
Table 10.1	Illustrative research questions and data sources related to the study of curriculum materials adoptions	199
Table 11.1	Fraction of teachers rated in the bottom 25 percent in the initial estimator who are not rated in the bottom 25 percent in another estimator for random and nonrandom grouping schools	223
Table 11.2	Rank correlations and misclassification error, under different scenarios, using different value-added modeling approaches	224
Table 12.1	Representing relationships using matrices	236
Table 13.1	Horizontal analysis—statement of revenues, expenditures, and changes in fund balance	270
Table 13.2	Vertical or common-size analysis—balance sheet (FY 2012 & 2016)	272
Table 13.3	Financial condition indicators	274
Table 14.1	Comparison between accounting and resource-based cost estimation	285
Table 14.2	Nested cost template	287
Table 14.3	AVID program delivery resource summary	295
Table 14.4	AVID/TOPS program costs	296

Table 14.5	Distribution of costs among stakeholder groups	297
Table 14.6	Comparison of AVID/TOPS and Talent Search program costs and effects	298
Table 15.1	Applications of mixed methodology in education policy research	313
Table 16.1	The differences between research and evaluation	326



Complementary Research Methods: An Introduction and Overview

Chad R. Lochmiller

Interest in educational research has grown substantially both because of federal research investments and a burgeoning literature base about research methods appropriate to the study of educational policy (Bascia et al. 2005; Fuhrman et al. 2007; Sykes et al. 2009). Oddly, there has been relatively limited discussion of the (dis)connections between the fields of educational leadership and policy. Despite their separate treatment as research disciplines, these fields have grown increasingly connected within the context of educational practice. In part, this connection reflects the important role that superintendents, principals, and other formal administrative leaders have come to hold within various policy processes as policy actors, advocates, and implementers (Cranston 2013; Knapp and Feldman 2012; Koyama 2011, 2014). Indeed, as policy actors, these titular leaders not only help shape policymaker's interpretations of educational challenges but are charged with the implementation of policy within the local context of schools. Policy researchers have thus increasingly considered the potential impact of educational practice on policy processes and have even gone so far as to suggest that leaders develop policy from their various administrative practices (Honig 2003). Knapp (2002) observed that

C. R. Lochmiller (✉)
School of Education, Indiana University, Bloomington, IN, USA
e-mail: clochmil@iu.edu

policy influences a variety of day-to-day decisions about teaching and learning. A cursory review of recent policies finds that many policy-driven educational reforms, such as increased pressure to improve the achievement outcomes of the lowest performing schools as codified by the Every Student Succeeds Act (ESSA), have necessitated fundamental changes in the work of educators and reflect increasing awareness by those on the front lines of educational organizations about the shortcomings of previous policy prescriptions.

Despite this convergence, however, a considerable divergence still exists between the fields of educational leadership and policy at both a substantive level (i.e., what we study) and methodological level (i.e., how we study it). Indeed, one need not look further than most doctoral training programs to find that much of the policy curriculum within these programs often focuses on leadership *or* policy, but rarely both. Many students are not mentored to see leadership as being situated within a policy context nor encouraged to explore how the policy context may itself be a product of local educational practices. Instead, students opt into tracks that focus on policy or leadership as discreet research foci and thus reinforce the appearance that these are, in fact, two separate disciplines. Many research-intensive universities are thus intentionally and unintentionally training policy students and future educational researchers to view leadership practice as a by-product of the policy environment rather than as a key influence within it. This training does not invite students to consider how leaders' actions influence, shape, and potentially necessitate particular policy responses. Thus, questions related to the (dis)connections between policy and leadership are left unexamined. Readings focus primarily on issues stemming from the legislative process, the generation and distribution of resources through various taxation mechanisms, economic factors, identification of programmatic responses to student populations, and the effects of large-scale reforms. Students are invited to apply perspectives from economics, political science, psychology, public affairs, sociology, and other traditional academic disciplines. All of this represents a fundamental challenge for doctoral faculty who, like myself, conduct research at the intersection of policy and leadership. Indeed, we often find that our students are prepared as policy scholars who cannot or do not wish to grapple with questions related to the study of leadership. To be fair, there are many doctoral programs that focus on leadership that do not expose their students to questions about policy in a substantive manner. In some cases, leadership students—many of whom are preparing for future work

as superintendents—do not complete coursework related to the education policy process let alone research methods used by policy scholars. At best, their training focuses almost exclusively on managing politics generated by the school board and to a lesser extent the surrounding community.

Few leadership and policy scholars have explicitly described how these fields should inform each other, in what ways their dominant research methods might be brought together to investigate shared lines of inquiry, and precisely what topics leadership and policy scholars might explore collaboratively as part of a shared research agenda. Desimone (2009) observed that “policy research can include the study of policy formation, implementation, effects, and cost-benefit analyses” (p. 163). This description positions specific methods, such as multisite qualitative case studies (Herriott and Firestone 1983) and cost-benefit analyses (Levin et al. 2017), as being central to the study of policy. Yet, for many scholars who study educational leadership, such methods might not support their research foci or raise questions that relate to practice. Conversely, much of what the field of educational leadership knows is based on “a limited set of methodological approaches” (Young and Crow 2017, p. 8). These approaches include descriptive qualitative case studies, surveys, and cross-sectional studies which are often anchored within specific preparation program contexts. Many of these approaches help generate descriptive accounts of local programs but do not contribute as much to the larger discourse around policies impacting leadership preparation nor expand our understanding to help us see how localized leadership interventions can support large-scale improvement. The tendency in both fields has been for scholars to retreat to their respective encampments without considering the rich middle ground between these fields where questions about the intersections and connections between leadership and policy abound (Orland 2009).

ENVISIONING A LEADERSHIP AND POLICY RESEARCH CONTINUUM

Assuming opportunities for integration between these fields exist and are meaningful, particularly at a methodological level, I use this volume to offer a different view of the relationship between leadership and policy research. This view proceeds from the belief that leadership and policy scholars are part of the same academic community. Thus, I view these fields as being situated at two ends of the same research continuum. On one end of this continuum, policy researchers engage in research that aims

to inform policymakers' understanding by describing the design, effects, and impact of their policies. In particular, these scholars seek to describe the broader policy context and various legislative processes that shapes, informs, and ultimately guides leaders' work in schools and districts. I conceive of this end of the continuum as one being focused on traditional policy research or traditional policy analysis, a type of policy research characterized broadly by efforts to "assess information needed to understand, design, plan, problem solve, and implement effective educational policies and practices" (Diem et al. 2014, p. 1071). This end of the continuum is thus principally concerned with studying how policy comes to exist, what it aspires to do within the public sphere, and whether it is effective at doing it.

The other end of the continuum describes what I refer to as applied policy research. This research is principally concerned with understanding how policy influences practice and conversely how practice influences policy. As such, scholars who work at this end of the continuum are intrinsically focused on policy-to-practice connections and thus aim to extrapolate how policy contributes to the work of practitioners (Knapp 2002). This end of the continuum necessarily invites questions about leadership, instruction, and learning. Thus, previously, such research has been traditionally focused on educational leadership. Yet, as scholars nest their studies of leadership within the broader policy environment, it becomes necessary to reevaluate our contemporary understanding of the aims and objectives of policy research. At this end of the continuum, scholars are less concerned with the ways in which policy is designed or assessed than about the ways in which policy is implemented or enacted within particular organizational or institutional contexts. The questions asked by these scholars focus on what policy does under specific conditions and in light of specific leadership activities.

Formulated in this manner, policy and leadership are not the dichotomous fields we have at times assumed. Instead, they are related fields that are part of the same broad research area. Within this area, scholars with different research foci can engage collaboratively to understand the full effect of policy and/or the full influence of leadership. This view has important implications for how scholars think about research methods. First and foremost, instead of viewing methods as being either for policy or leadership, a continuum of this sort suggests that scholars might bring multiple methods to bear on a single issue in order to understand its complexity in total. For instance, a scholar who is interested in the issue of

teacher quality might not only employ sophisticated value-added models to assess teacher's contributions to student learning (as discussed in Chap. 11) but also conduct multi-site case studies to understand how leaders make sense and interpret the results of such models as they design supports for classroom teachers' practice (as discussed in Chap. 3). Within such an agenda, a scholar might extend on this work to examine how leaders talk about performance within the context of preobservation and post-observation conferences using various language-based methodologies (as discussed in Chap. 4). While any one of these lines of inquiry is clearly sufficient, in combination they provide a comprehensive assessment of the design, implementation, and impact of the policy.

Second, this perspective invites policy and leadership scholars to tackle the inherent complexity of contemporary leadership and policy issues and thus see it through multiple methodological approaches, theoretical perspectives, and various analytic levels. As Green et al. (2006) noted in the introduction to the American Educational Research Association's publication, *Handbook of Complementary Methods in Education Research*, "Today, it is virtually impossible for any one approach to be used to address the complex issues being explored through research in education. Further, no longer is it a question of alternative research traditions... but of which approaches are appropriate to the questions under study and which can be productively combined within a program of research" (p. xvi). Indeed, this is the heart of the motivation for complementary methods in educational research and thus compels the production of this volume.

AIM OF THIS VOLUME

As I prepared this volume, began writing this chapter, and ultimately thought about this continuum of policy research, I noted that there are few books that specifically introduce research methodologies and methods in such a complementary fashion. Indeed, there is some indication that the principles which should be used to guide the process of bringing together different research approaches to study leadership and policy issues are somewhat unclear and certainly underdeveloped (Green et al. 2006). Rather, most texts simply identify a suite of methods that can or should be used by these fields. This is not completely surprising. As Riehl and Firestone (2005) and Heck (2004) have both astutely noted, neither the field of educational leadership nor the policy field has widely agreed upon a coherent methodological core. Instead, both fields operate with what

may best be described as a “methodological mosaic” (Shulman 1997, p. 14). That is, a compilation of diverse but relatively compatible methodologies and methods that enable scholars to investigate topics of mutual concern. This compilation has many affordances. Perhaps foremost, it invites innovation and experimentation in the research process that leads to new and potentially valuable interpretations of policy issues. Yet, it becomes quite difficult to determine how scholars who engage in leadership and policy research can work across fields. It also complicates the efforts for students seeking to identify a topic, question, or methodological approach when their interests reside at the intersection of policy and leadership.

A critical question, then, is how research methodologies and methods can be presented in a complementary manner that enables policy and leadership scholars to investigate areas of mutual interest? This volume begins to address this question by first bringing together scholars who specialize in leadership and policy to share their methodological expertise. Indeed, in my previous research, I have noted that the difference between leadership and policy relates not as much to *how* scholars undertake their analyses as it does to *what* scholars study in the first place (Lochmiller and Hedges 2017). Thus, within this volume, the contributors highlight a number of ways in which research methodologies and methods might be employed to study important policy or leadership issues.

SEARCHING FOR BETTER RESEARCH METHODS: SEEING POTENTIAL IN COMPLEMENTARY APPROACHES

To a certain extent, this volume embodies the perpetual quest for better research methodologies and methods to study leadership and policy issues. It bears noting that complementary methods are not simply another call for mixing qualitative and quantitative research approaches. This may be an attractive option for some policy researchers, but it might not produce the kind of detailed understanding of policy issues that the application of multiple methods might derive. Instead, complementary methods represent the application of multiple methodological approaches to the study of related policy and leadership issues. The aim is to derive a more comprehensive and detailed understanding of a policy, its design, implementation, and impact than any single or mixed method can produce alone. As such, complementary methods maintain the integrity of the methods employed but orient research questions in ways that enable

scholars to deepen, expand, or contribute to a larger debate in a more comprehensive way. It is thus partly a way of (re)organizing and (re)orienting our work as leadership and policy scholars to explore what Knapp (2002) has described as a “set of puzzles” that relates to the relationship(s) between policy and practice (p. 5). This approach requires asking, “First, precisely how does policy meet practice – in what forms, under what circumstances, and with what consequences for whom? Second, where should we look, and how, to identify the connections between policy and practice? Third, what framing ideas and lines of investigation would be most helpful in accomplishing this goal?” (Knapp 2002, p. 5). Ultimately, these questions invite policy and leadership scholars to move beyond their respective camps into a middle space that enables them to identify, make sense of, and respond to the uncharted relationship between policy and leadership practice. It invites scholars to consider how their work might become complementary.

This volume presents a variety of different qualitative and quantitative research methodologies and methods that collectively may be uniquely suited to unearthing the salient (dis)connections between policy and practice. It is assumed that the value of these discrete methods increases when the results obtained from their use are viewed in concert with those derived from other methodological approaches. Thus, within this volume, complementary methods are described as discrete research methodologies which scholars can bring together to produce a sophisticated understanding of policy and its relationship to practice. The aim of this volume is to advance an orientation to research design that treats the inherently complex, messy, and nested (inter)relationships between policy and practice as an invitation to employ multiple methodological perspectives or approaches.

Given this focus, the volume invites policy and leadership scholars to move into a more coherent dialogue about their research practices. The vision especially aims to help novice scholars envision new lines of inquiry that use methodologies and methods commonly associated with policy and leadership. To achieve such integration, I have invited leading policy and leadership scholars, and in some cases methodologists, to offer their insights about research methodologies and methods that enable us to examine how policy influences leadership and vice versa. Across the chapters, the volume thus provides readers specific insights about current lines of inquiry and emerging topics that use these methodologies. It provides scholars with specific recommendations about designing research studies

using the methodologies and methods discussed by the chapter authors. And, most importantly, it describes the theoretical perspectives that are often taken up when using these methodologies and methods.

ORGANIZATION OF THE VOLUME

This volume is organized into three parts with each having its own dominant methodological orientation. Part I presents qualitative and critical approaches to the study of educational leadership and policy. In Chap. 2, “Qualitative Research in Educational Leadership Studies: Issues in the Design and Conduct of Studies,” Jeffrey Brooks and Anthony Normore provide an introduction to qualitative research in educational leadership. This chapter, which the authors model after an article in the *International Journal of Educational Management* (Brooks and Normore 2015), offers an introductory conversation that serves to define the field of qualitative research relative to the study of educational leadership. In Chap. 3, “The DIVE Approach: Using Case-Ordered Meta-Matrices and Theory-Based Data Displays to Analyze Multiple Case Study Data,” Susan Bush-Mecenas and Julie Marsh offer an insightful introduction to the case-ordered meta-matrix as an emerging analytic approach. While drawing on established qualitative analysis techniques, this chapter offers a thoughtful adaptation that scholars may find useful when considering data across and within different case study sites. This approach seems especially fruitful when comparing educational leaders’ practices within and across policy contexts. In Chap. 4, “Language-Based Methodologies for Policy and Leadership Research,” Jessica Nina Lester and Justin Paulsen provide an introduction to language-based research methods appropriate for the study of educational leadership and policy. This chapter foregrounds the utility of discourse analysis, conversation analysis, and so on in studying how policy is developed and implemented. As such, the chapter makes an important methodological contribution to the field by highlighting how scholars might employ these relatively novel methods to study leadership and policy issues at a micro level. In Chap. 5, “Doing Critical Policy Analysis in Education Research: An Emerging Paradigm,” Michelle Young and Sarah Diem offer a reconceptualization of policy analysis using critical qualitative perspectives. This chapter makes an important theoretical contribution to the volume and the broader field as it conveys some of the ways in which policy scholars might address issues of (in)equity, (in)justice, power, and oppression found within extant policy structures. Complementing this

chapter, Irene Yoon offers an introduction to critical perspectives that are appropriate to the study of educational leadership in Chap. 6, “Critical Research Perspectives in School Leadership: Putting Dignity and Humanity at the Center.” This chapter provides an invitation to scholars to consider how leadership practices and roles reinforce inequities within particular organizational and community settings. Finally, in Chap. 7, “The Potential of (Participatory) Action Research for School Leaders, Local Policy Makers, and University-Based Researchers,” Meagan Call-Cummings and Melissa Hauber-Özer provide an introduction to action research and participatory action research methods suitable for the study of leadership and policy at the local level. This chapter invites scholars to consider how they might unpack and address inequities at the local, state, and national level. Indeed, these methods may be particularly useful to scholars seeking to change, reform, or otherwise disrupt the inequitable and unjust power structures, which are discussed in detail in Chaps. 5 and 6.

Part II of the volume presents quantitative approaches that can be used to study educational leadership and policy. In Chap. 8, “Secondary Data Analysis in the Field of Educational Leadership and Policy Studies,” Angela Urick provides an introduction to secondary data analysis using longitudinal data. Her chapter provides novice and experienced scholars with a wealth of insights about setting these analyses, obtaining data to use, and offering guidance on the statistical procedures that can be used to carry out these analyses. In Chap. 9, “Matching Strategies for Causal Inference with Observational Data in Education,” Yongnam Kim, Stan Lubanski, and Peter Steiner introduce readers to matching strategies used to establish causal inference with observational data. This chapter contributes to the discussion of quasiexperimental research techniques and thus provides guidance to scholars seeking to establish more robust claims about the impact of interventions on various outcome measures. In Chap. 10, “Using Quantitative and Qualitative Methods to Study the Content and Effects of Curriculum Materials,” Morgan Polikoff, Shauna Campbell, and Shira Korn introduce readers to the study of curriculum materials. As Polikoff and his colleagues adeptly note, scholars have not widely studied curriculum implementation, and thus this represents a fertile area within which scholars might delve meaningfully into the conditions that directly impact teaching and learning. In Chap. 11, “Value-Added and Growth Models in Education Research,” Cassandra Guarino provides an incisive introduction to growth and value-added modeling techniques. These techniques have received increasing attention from

policy scholars, particularly economists, and so understanding how these analytic techniques can be applied to the study of leadership and policy seems timely and needed. In Chap. 12, “Social Network Analysis Methods in Educational Policy Research,” Kara Finnigan, Daniella Luengo-Aravena, and Kim Garrison introduce readers to social network analysis as a compelling analytic technique to study the development, implementation, and impact of educational policy. Finnigan’s groundbreaking work around social networks has helped the field understand how important relational ties are within the context of educational reform implementation. In offering this chapter, she and her colleagues provide an incisive introduction to this valuable methodology. In Chap. 13, “Essential Steps to Assessing a School System’s Fiscal Health,” Joshua Zender, Kenneth Smith, and John Kurpierz provide an introduction to fiscal research in school districts. This chapter provides a straightforward approach to carrying out analyses that are particularly focused on understanding a school district’s overall fiscal condition and health. In Chap. 14, “Evaluating Education Costs,” Tammy Kolbe and Rachel Feldman offer an introduction to cost estimates appropriate to program evaluation and policy analysis. Given changing resource streams in public education, the approaches discussed in both of these chapters seem particularly relevant to the study of policy and leadership.

Finally, Part III of the volume introduces readers to mixed, applied, and collaborative approaches to the study of educational leadership and policy. In Chap. 15, “Using Mixed Methods to Inform Education Policy Research,” Colleen Chesnut, John Hitchcock, and Tony Onwuegbuzie introduce readers to the use of mixed methods research designs when carrying out studies focused on educational leadership and policy. In Chap. 16, “Program Evaluation Methods for PK-12 Education,” Liz Hollingworth offers an introduction to program evaluation as a method for investigating the local impact of policy and leadership decisions. In Chap. 17, “International Network as Sites for Research on Successful School Leadership,” Christopher Day and David Gurr draw upon their experience co-leading the International Successful School Principal Project to illustrate how scholars might undertake collaborative investigations of contemporary leadership and policy issues across national contexts. Collectively, these three chapters offer scholars insights which demonstrate the value of applied and mixed methods approaches to the study of leadership and policy.

In the volume's concluding chapter, "Taking Stock of Complementary Research Methods: The Perpetual Quest for Good Research Methods for Educational Leadership and Policy," Carolyn Riehl offers an insightful consolidation of the chapters presented in the volume and uses the chapter as an opportunity to identify possible future directions for policy and leadership research. Indeed, a central message in her chapter is that this volume and the methods described herein do not constitute a new approach to research but instead reflect evolving approaches to our understanding of a field that continues to seek new, innovative, and potentially more integrated approaches to studying shared topics, ideas, or policy concerns.

HOW SCHOLARS SHOULD USE THIS BOOK

While this book provides the reader with an introduction to various qualitative, quantitative, mixed, and applied research methodologies and methods, it should not be used as a substitute for an in-depth introduction to any particular methodology or method. Rather, scholars should treat this volume as a primer designed to assist them in considering which methods may be useful in conducting a study or program of research related to educational leadership and policy. Thus, as scholars read the chapters, they might pose a series of epistemological questions. As Lochmiller and Lester (2017) noted, "Epistemology refers to the idea of knowledge construction and centers around what we know and how we go about knowing" (p. 12). This involves asking questions, such as:

- How does the approach offered by the scholar suggest a particular way of thinking about and/or understanding of leadership and policy?
- What can be known using this approach, and how does this approach require understanding or making sense of leadership and/or policy issues in particular ways?
- What are the particular affordances of this approach to constructing "new" knowledge relative to the leadership and/or policy issues one might consider?
- How does the approach constrain possible alternative explanations? For instance, is the approach compatible with a critical perspective when it functions from a largely confirmatory or positivist paradigm?

- How should the epistemological considerations related to each approach guide the development of the study design, selection of research methods, identification of potential sources of evidence, or development of the particular analytic approach?

A scholar might also read these chapters and pose a series of methodological questions. These questions articulate the considerations related to the design of the research study, including how the scholar will frame the study design and the particularities of participant selection, data collection strategies, and the overarching analytic approach, as well as consideration of the potential limitations of the study relative to its questions or aims. This volume does not subscribe to the philosophy that questions should drive the selection of research methods as some have suggested (Creswell 2014). Rather, it follows a more emergent line of thinking regarding the design and development of research studies (Knapp 2016; Maxwell 2013). This line of thinking suggests that there may be multiple entry points for a research study and that these can all productively lead to the development of an effectively designed research effort. Thus, as scholars read the chapters, they might ask methodological questions such as:

- What sources and types of evidence matter given the particular research design?
- What data collection strategies should be employed to support a study using this design?
- How do the particular nuances of the analytic approach or perspective influence how a scholar should state or make sense of their claims?
- What are the affordances of the approach relative to other approaches one might take to study leadership and/or policy issues?
- How does the methodology potentially exclude or include particular voices, perspectives, or populations?

Beyond epistemological and methodological questions, this volume might inspire its readers to ask substantive questions, as well. Within each chapter, the authors provide a summary of recent research that has drawn upon the methodology and/or methods featured. This review supports the reader's thinking about topics and/or issues and identifies spaces

wherein readers might conduct research that more fully integrates questions about leadership and policy. As readers consider the topics discussed, they might find it beneficial to ask:

- What are the unexamined, underdeveloped, or insufficiently examined leadership or policy issues within the research presented?
- In what ways might leadership perspectives contribute to the existing understanding of a given policy issue? Conversely, how might policy perspectives contribute to the existing understanding of a given leadership issue?
- What unanswered questions exist in relation to the topic that leadership and policy scholars might endeavor to explore?
- How might leadership and policy scholars define, describe, or conceptualize the issues, topics, or subjects in more robust ways than current research suggests?

These questions should surface further questions for consideration and in doing so create opportunities for scholars to initiate dialogues that support continued exploration of these topics using one or more of the approaches described in the volume.

CONCLUSION

Ultimately, this volume offers novice and experienced scholars a resource to envision new lines of inquiry and new uses of research methodologies and methods to explore issues related to leadership and policy. More particularly, it serves as a resource for researchers who aspire to engage in an expanded dialogue about key policy topics that necessarily involve issues related to policy design, analysis, and implementation. In doing so, it positions leadership research not as a field separate from policy but as an extension of it, which thereby serves particular organizational and institutional contexts. A particular aim of this volume, then, is to suggest that scholars who specialize in leadership and policy see their work not as mutually exclusive but as complementary domains of inquiry situated on a larger research continuum. Such a view has many affordances, not the least of which, is that new understandings and more sophisticated interpretations of the complexity of issues facing educators, schools, and districts can be better understood.

REFERENCES

- Bascia, N., Cumming, N., Datnow, A., Leithwood, K., & Livingstone, D. (Eds.). (2005). *International handbook of education policy*. Dordrecht: Kluwer.
- Brooks, J. S., & Normore, A. N. (2015). Qualitative research and educational leadership: Essential dynamics to consider when designing and conducting studies. *International Journal of Educational Management*, 29(7), 798–806.
- Cranston, N. (2013). School leaders leading: Professional responsibility not accountability as the key focus. *Educational Management Administration & Leadership*, 41(2), 129–142.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed approaches* (4th ed.). Thousand Oaks: SAGE.
- Desimone, L. (2009). Complementary methods for policy research. In D. Plank, G. Sykes, & B. Schneider (Eds.), *AERA handbook on education policy research* (pp. 163–175). Washington, DC: American Educational Research Association.
- Diem, S., Young, M. D., Welton, A. J., Mansfield, K. C., & Lee, P. (2014). The intellectual landscape of critical policy analysis. *International Journal of Qualitative Studies in Education*, 27(9), 1068–1090.
- Fuhrman, S. H., Cohen, D. K., & Mosher, F. (Eds.). (2007). *The state of education policy research*. Mahwah: Earlbaum.
- Green, J. L., Camilli, G., & Elmore, P. B. (Eds.). (2006). *Handbook of complementary methods in education research*. New York: Routledge.
- Heck, R. H. (2004). *Studying educational and social policy: Theoretical concepts and research methods*. Mahwah: Lawrence Earlbaum.
- Herriott, R. E., & Firestone, W. A. (1983). Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher*, 12(2), 14–19.
- Honig, M. I. (2003). Building policy from practice: District central office administrators' roles and capacity for implementing collaborative education policy. *Educational Administration Quarterly*, 39(3), 292–338.
- Knapp, M. S. (2002). *Understanding how policy meets practice: Two takes on local response to a state reform initiative*. Seattle: University of Washington. Retrieved from <http://www.education.uw.edu/ctp/sites/default/files/ctpmail/PDFs/PolicyPractice-MSK-06-2002.pdf>
- Knapp, M. S. (2016). The practice of designing qualitative research on educational leadership: Notes for emerging scholars and practitioner-scholars. *Journal of Research on Leadership Education*, 12(1), 26–50.
- Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external accountability: Challenge for urban school leadership in the United States. *Journal of Educational Administration*, 50(5), 666–694.
- Lochmiller, C. R., & Hedges, S. (2017). Education policy implementation research: A call for new approaches. In J. N. Lester, C. R. Lochmiller, &

- R. Gabriel (Eds.), *Discursive perspectives on education policy and implementation* (pp. 17–40). New York: Palgrave Macmillan.
- Lochmiller, C. R., & Lester, J. N. (2017). Future directions for education policy research and language-based methods. In J. N. Lester, C. R. Lochmiller, & R. Gabriel (Eds.), *Discursive perspectives on education policy and implementation* (pp. 241–252). New York: Palgrave Macmillan.
- Koyama, J. (2011). Principals, power, and policy: Enacting “supplemental educational services”. *Anthropology & Education Quarterly*, 42(1), 20–36.
- Koyama, J. (2014). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly*, 50(2), 279–304.
- Levin, H. M., McEwan, P. J., Belfield, C. R., Bowden, A. B., & Shand, R. D. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*. Thousand Oaks: SAGE.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). Thousand Oaks: SAGE.
- Orland, M. (2009). Separate orbits: The distinctive worlds of educational research and policymaking. In D. Plank, G. Sykes, & B. Schneider (Eds.), *AERA handbook on education policy research* (pp. 113–128). Washington, DC: American Educational Research Association.
- Riehl, C., & Firestone, W. A. (2005). What research methods should be used to study educational leadership. In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research in educational leadership*. New York: Teachers College Press.
- Shulman, L. (1997). Disciplines of inquiry in education: A new overview. In R. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 3–31). Washington, DC: American Educational Research Association.
- Sykes, G., Schneider, B., & Plank, D. N. (Eds.). (2009). *Handbook of education policy research*. Washington, DC: American Educational Research Association.
- Young, M. D., & Crow, G. M. (Eds.). (2017). *Handbook of research on the education of school leaders* (2nd ed.). New York: Routledge.

PART I

Qualitative and Critical Approaches to the Study of Educational Leadership and Policy

Part I of this volume presents research methodologies and methods closely aligned with qualitative and critical approaches to the study of educational leadership and policy. Specifically, this part of the volume addresses the design of qualitative research studies; qualitative analytic strategies, such as the case-ordered meta-matrix; language-based research methods; critical approaches to the study of leadership and policy; and action research.



Qualitative Research in Educational Leadership Studies: Issues in the Design and Conduct of Studies

Jeffrey S. Brooks and Anthony H. Normore

Educational leadership scholars have employed qualitative research methods since the mid twentieth century to explore various dynamics, phenomena, contexts and perceptions (Barnhardt et al. 1979; Wolcott 1970,

An earlier version of this chapter appeared as an article in the International Journal of Educational Management: Brooks, J. S. & Normore, A. N. (2015). Qualitative research and educational leadership: Essential dynamics to consider when designing and conducting studies. *International Journal of Educational Management* 29(7): 798–806. The work is reprinted here with permission of Emerald Publishing.

J. S. Brooks (✉)
Monash University, Melbourne, VIC, Australia

School of Education, RMIT University, Bundoora, VIC, Australia
e-mail: Jeffrey.brooks@rmit.edu.au

A. H. Normore
Graduate Education, California State University-Dominguez Hills,
Carson-Los Angeles, CA, USA
e-mail: anormore@csudh.edu

1973, 1974, 1977). During the first few decades of this era, scholars generally grounded these studies in a specific social science, such as anthropology, sociology or political science (Bogotch, et al. 2008; Brooks and Miles 2010). As time went on, scholars have drawn liberally from these areas without regard for disciplinary allegiance. As a result, they have situated contemporary educational leadership research as an interdisciplinary field of inquiry that applies multiple methods to explore issues salient to the field (Lochmiller and Hedges 2017; Normore and Brooks 2014). Within this plurality, qualitative approaches have emerged as one of many important ways of understanding and contributing to improved practice, theory and research.

Qualitative research in educational leadership has yielded many insightful studies that have enriched and deepened our understanding of how dynamics such as influence, power, communication, collaboration, administration, abuse, equity, management and organizations work in educational organizations and contexts (Capper 1993; Grogan 1999; Theoharis 2009). Certain norms have emerged with respect to the way that qualitative studies are designed, executed and reported (Bogdan and Biklen 1998). However, while there is some broad consensus on what constitutes “good” qualitative research and a proliferation of qualitative studies of educational leadership, there is relatively little methodological literature devoted to describing and considering issues specific to the qualitative study of the field. Scholars seem largely content to borrow from other fields rather than develop bespoke methods (Lochmiller and Hedges 2017). This has limited the methodological development of the field and yielded methodological approaches which may be poorly suited to the special demands of the content or context in which these studies are conducted.

We contend that it is important for qualitative researchers in the field of educational leadership to be critical of the research methods they employ. In particular, we encourage scholars to adapt present approaches or create new ones that might better suit the special needs of this applied field. Although a growing number of printed works on qualitative research methods are available, it is rare to locate a text that examines the use of qualitative research methods in the context of the study of leadership, even though the field has produced a voluminous body of research.

The purpose of this chapter is to draw attention to certain dynamics that scholars should consider when designing and conducting qualitative research on educational leadership. In doing so, we draw on not only the

literature but also on our experience of designing, conducting and publishing qualitative research. While we think of the points we raise in this chapter as a selective rather than exhaustive set of issues, we accept that in this chapter we cover a lot of ground in an attempt to touch on issues we see as critical to the conduct of outstanding work. Some of the dynamics we note here are explored thoroughly by other authors, and in those cases, we include references that readers may want to seek out for additional information. However, we do raise some issues not previously addressed in the literature, and in so doing we invite you to think about the way they might be pertinent to your own work as a scholar. We have organized the chapter around five broad aspects of qualitative inquiry: research design; data collection; data analysis; rigor; and communication.

RESEARCH DESIGN IN QUALITATIVE STUDIES OF EDUCATIONAL LEADERSHIP

When designing qualitative studies of educational leadership, it is important to choose the appropriate research design that will help explore your research questions. In order to choose an appropriate design, scholars must have a clear understanding of (a) what they are studying and (b) which design is most appropriate for that topic, phenomenon, dynamic, person or place. For example, someone seeking to understand how a principal influences school culture might choose ethnography as it is specifically suited to the study of culture (Creswell 1998; Wolcott 1970). A study of a semester, school year, classroom or school might demand a case study design (Merriam 1991; Stake 1995, 2008; Yin 1994). A study of a single leader might employ a narrative, portraiture or oral history design (Horsford 2011). Suffice it to say that choosing the appropriate research design is a critical early decision a scholar has to make when crafting an outstanding study. It helps the scholar think through the *who, what, when, where, how* and *why* issues related to the study and can be thought of as a road map or blueprint for the project (Merriam 1991). Choosing a design also helps locate the work within a methodological tradition that can help inform decisions throughout the research process. One can also choose an emergent design or even combine designs to address a research question that does not fall neatly into one of these traditions. However, while this is possible, we do not advise such an approach for novice scholars—not because they cannot or should not be innovative or creative, but because

mixing traditions invites a high level of complexity and often creates issues with design fidelity (Fielding and Lee 1998).

It is also important to think carefully about epistemology and purpose when designing and conducting qualitative studies of educational leadership. Researchers should consider their beliefs about the nature of knowledge and the kinds of knowledge they intend to generate through their study. Is the purpose to *improve* a person, system or school? Is the purpose to *explore* something we know little about? Do you hope to *discover*, *refute* or *refine* a theory? All of these issues are particularly important with respect to qualitative studies of educational leadership because the nature of the work makes it likely that the research will speak truth to power, and it is a scholar's moral obligation to think through the reasons they have for designing the study in a particular manner (Bryman 1984; Charmaz 2006; Hatch 2002).

One glaring omission in many qualitative research studies of educational leadership is a lack of attention to the relational, power and gatekeeper dynamics that influence the study. As leadership is a relational activity, it is important to be clear about the various relationships related to the study (Eacott 2015)—not only between the researcher and participants, but also in regard to relationships between participants, between the organization and community and any other relationship that may influence the work. For example, if a researcher is going to study the students in an educational leadership program, they must disclose their relationship to the students. If they are the students' instructor or peer, there are obvious reasons to suspect that any data collected may be influenced by the subject's disposition toward the researcher and vice versa. Similarly, in many qualitative studies of educational leadership, the gatekeepers of the contexts are not made clear (Seidman 2013). This is potentially problematic because, for example when conducting a study of teacher leadership in a given context, it may be of great importance whether the researcher is introduced to teachers by a peer teacher or the teacher's superintendent. Put simply, since leadership is at least part concerned with the ways that people influence each other, it is important to consider the various ways that power dynamics may influence the study. Thinking this issue through at the design stage is critical so the scholar can be clear about the role of power in the study.

In summary, choosing an appropriate research design and then adapting it to suit the specific context of the study is one of the most important processes a qualitative researcher will undertake (Janesick 1994). Carefully

thinking through issues related specifically to the ways that leadership practice might influence those decisions is critical to maintaining the fidelity of the study and enables or prevents a scholar from exploring their research questions. Although we touch lightly on the aspect of research design, we note that other scholars have recently attended to this issue in more detail (see Knapp 2017 for a particularly notable example).

DATA COLLECTION IN QUALITATIVE STUDIES OF EDUCATIONAL LEADERSHIP

Over the past 50 years, there have been many exciting developments in terms of the types of qualitative data a researcher might collect. Innovations with visual data, Internet-based data and other sorts of qualitative data have opened up new possibilities for answering complex and interesting research questions (Salmons 2016). That said, there are three basic types of qualitative data that scholars have generated in order to explore their research questions: *interviews, observations and documents*. In the following sections, we consider each (Silverman 2016).

An *interview* (i.e., individual and focus group) is an overarching term used to describe a range of different forms of interviewing most commonly associated with qualitative research. Scholars might structure interviews fully in order to guide the participant in addressing questions in a particular sequence. Alternatively, scholars might allow flexibility within the interview to afford participants the opportunity to share thoughts informally or to allow the participant to direct portions of the interview toward relevant topics or issues. According to Lewis-Beck et al. (2004), the aim of the latter approach is usually to ensure flexibility in how and in what sequence questions are asked, and in whether and how particular areas might be followed up and developed with different interviewees. The composition of a focus group needs great care to get the best quality of discussion (Stewart and Shamdasani 1990). There is no “best” solution to group composition, and group mix will always impact on the data, according to things such as the mix of ages, sexes and social professional statuses of the participants. What is important is that the researcher “gives due consideration to the impact of group mix (e.g., how the group may interact with each other) before the focus group proceeds” (Gill et al. 2008, p. 293).

Often interviews in qualitative studies of educational leadership are a form of elite interview (Harvey 2010; Marshall 1984). Elite interviews are those conducted with those at the top of an organization or social structure. This might include, depending on the context and the topic of study, principals, superintendents, teachers, policymakers, university professors and so on. While these perspectives are certainly interesting, important and critical to many leadership studies, it is important to interrogate the power dynamics of the interview between the interviewer and interviewee, the motivation of the interviewee to speak freely and the various ways that the elite may have privileged information or capacity to influence the organization. As such, it is important to understand that with privilege and power come many temptations or necessities to present information in a particular manner (Aberbach and Rockman 2002). One common issue we have observed in our own work is for school leaders to keep their interview responses at a high level of abstraction. For example, we have had several experiences where an interviewee spoke of missions, visions, dispositions, teacher quality and so on without wanting to further explain or articulate what those concepts actually looked like in practice. This is perhaps motivated by a desire not to disclose potentially sensitive personal data or by a fear that certain things they say might reflect poorly on their performance as a leader. In order to get them to move away from responding with abstract answers, we have used probing follow-up questions that ask for examples or instances to generate richer responses. In any event, it is critical that educational leadership scholars consider the motivations, power and privilege of interviewees when conducting interviews.

Observations can be useful in myriad ways to check for nonverbal cues and expression of feelings (e.g., movement of eyes, head, gestures etc.), determine who interacts with whom, grasp how participants communicate with each other, and check for how much time is spent on various activities. When conducting observations in qualitative studies of educational leadership, it is likewise important to consider the relationship between the observer and observed and the relationship between the subjects. The ways that people manage, inspire or communicate with each other is likely influenced by ways that leadership is practiced in context. That is, it may be a norm in a particular school for teachers to engage with each other in a certain manner in meetings or for students and teachers to interact in a way that is framed by leadership practice (Merriam 1988; Mulhall 2003).

It is important to think carefully about the way that observations might be shaped by the people who exert dominant influence over others.

Moreover, the researcher should think carefully about whether they are making sense of what they observe from an etic (outsider) or emic (insider) perspective. Since the settings in which most qualitative leadership studies are conducted are somewhat familiar, there is a temptation for scholars to feel as though theory can interpret the significance of what they see based on prior experience. However, in many cases the researcher is conducting the study from an etic rather than emic perspective. Confusing the two can lead to erroneous observation data based on the researcher's bias (Peshkin 2000; Wolcott 2002).

Document analysis is a systematic procedure for reviewing or evaluating documents—both printed and electronic material. Bowen (2009) states that “Like other analytical methods in qualitative research, document analysis requires that data be examined and interpreted in order to elicit meaning, gain understanding, and develop empirical knowledge” (p. 27). Documents contain text (words) and images that have been recorded without a researcher's intervention. Researchers refer to documents as “social facts” which are produced, shared and used in socially organized ways (Corbin and Strauss 2008). Documents in qualitative research on educational leadership must be carefully considered in terms of their level of abstraction and formality.

It is common for scholars to use school improvement plans, meeting agendas/minutes, school newsletters, letters home to parents and the like as qualitative documents. This is appropriate for many studies; however, it is important to treat them as what they are—intentionally shaped documents created for the purpose of communicating formal organizational dynamics to a critical audience. Such public and public-private records present an individual perspective as a collective perception and should be approached as such. In a sense, it is important for the researcher to ascertain the degree to which such documents represent an idealized or espoused perspective on the work rather than an actual or critical perspective.

DATA ANALYSIS IN QUALITATIVE STUDIES OF EDUCATIONAL LEADERSHIP

While there are many issues related to analyzing data, we draw educational leadership scholars' attention to one particular matter: transparency of analytic procedures. With respect to the role of theory in analysis,

educational leadership scholars using qualitative methods have a tendency for theory to disappear when discussing how they used constructs to analyze data (Brooks 2016). This means that the field seldom makes transparent the ways in which theory is used to make sense of empirical data. If scholars leave it as “themes emerged” rather than going into more detail about the ways that salient themes presented in findings sections were developed, the field is missing a crucial piece of theory-building and generating information that may help subsequent scholars further refine their analyses (Honig 2006; Wolcott 1982). We are not concerned so much that scholars have used inappropriate analytic techniques, as it would be helpful to follow the lines of thinking and analysis that lead to themes. Laying bare this part of the research process would help us understand how theoretical constructs evolve to a much higher degree than we are currently able to see. This, in turn, will help the field do a better job developing, refining, exploring and discovering new ideas and theories (Fielding and Lee 1998).

RIGOR IN QUALITATIVE STUDIES OF EDUCATIONAL LEADERSHIP

While each research design has a specific approach to establishing rigor, there are a few issues specific to qualitative studies of educational leadership that scholars should consider. It is critical that scholars are aware of the ways that various research designs establish rigor and then even more important that they do not violate these norms and rather meet various thresholds for quality and rigor. For example, case study research uses concepts such as triangulation and member check to establish the reliability and trustworthiness of the work (Merriam 1998). If triangulation is the technique scholars use, then they should actually follow through and explain how they approached this aspect of the study. It is currently common for triangulation or rigor to be given a few sentences in a methodology section, and then for it to remain unclear how or if the technique was actually employed in the study. For example, if a study claims that data were gleaned from an analysis of documents, interviews and observation, all three forms of data should be evident in the findings. If there is an imbalance—say, if there is primarily interview data and only a sprinkling of observation or document data—we should call into question the rigor and quality of the study (Brooks and Normore 2015; Wolcott 1970). Again,

this is not an indictment of scholars in the field, it is a call for greater transparency, care and explanation in the conduct of research.

We also feel strongly that one special approach to establishing rigor bears greater consideration in qualitative studies of education in general—transferability (Malterud 2001). In a field desperate for sharing best practices and learning lessons from around the globe that might be helpful in local contexts, it is curious that researchers have not more carefully considered the transferability of the work beyond the context of a specific study (Krefting 1991). It would be useful, for example to have scholars think beyond quantitative-bound concepts like generalizability when discussing their qualitative work and consider the possibilities of conditions under which the lessons might be applied elsewhere. For example, instead of simply reporting the processes and outcomes of a single-school study, imagine if it were common for educational leadership scholars to openly discuss their perspective on the conditions under which other educational organizations or processes might meet failure or success should they try and implement a similar initiative. This basic issue gets surprisingly short shrift in qualitative studies of educational leadership, which often end just short of taking this final step.

COMMUNICATION AND QUALITATIVE STUDIES OF EDUCATIONAL LEADERSHIP

Qualitative studies of educational leadership should be written in an accessible manner, and we encourage scholars to adopt a writing process that communicates in multiple formats (Ponterotto and Grieger 2007). For example, the writing process for a piece of qualitative research might include all (or even more) of the products listed in Fig. 2.1. Often, scholars produce a few of these products when they conduct qualitative studies. In our experiences, the most common among them are the conference proposal, conference paper or an article. We would add these others, and still more (grant proposals, for example) to the communication of qualitative research. Offering key findings and ideas in multiple formats allows for a scholar to reach the greatest possible audience, and also helps scholars, policymakers, community members and practitioners understand the substance and utility of the work (Tracy 2012).

There are a few further issues we would urge scholars designing and conducting qualitative studies of educational leadership to consider

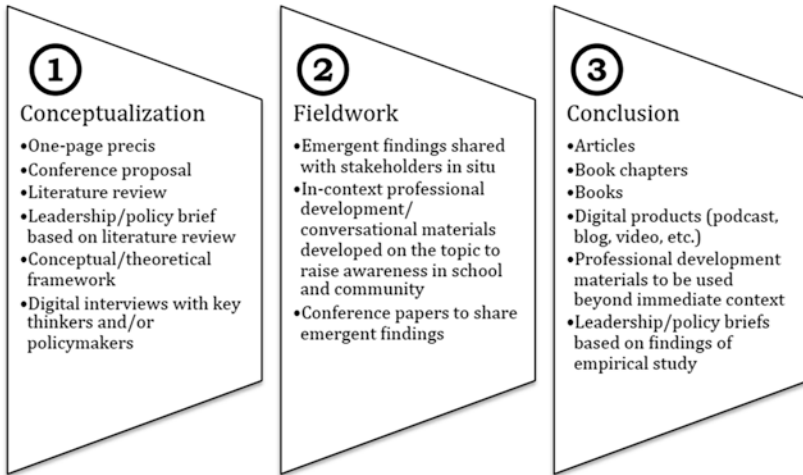


Fig. 2.1 Research dissemination products for qualitative studies of educational leadership. (Source: Brooks and Normore 2015)

(Patton 2005). First, it is important to think through what aspects of the research are handled as a closed or open feedback loop—what products are designed for use in the context you are studying? To whom are the results communicated? Is the timing and manner of reporting useful? Second, it is critical for scholars to succinctly and clearly articulate the “so what” of the study—how should the work help people rethink or approach their work as scholars and practitioners in a new way? Third, can you develop discussion or implementation guides for school leaders? This is a useful way to help communicate complexity in an accessible manner. For example, do the chapter, article or book end with discussion prompts to help people think through how the issue at hand might manifest in other contexts? Do you prompt leaders and those who train or study them to reconsider certain common practices or reinforce those that may be useful (Maxwell 2012)?

CHAPTER SUMMARY

Qualitative research has been a strong and vibrant part of the educational leadership knowledge base for at least 50 years. There are many well-designed and executed studies that have helped refine, deepen and challenge our thinking about the ways that leaders are prepared and the way they practice the art and science of their craft. In this brief chapter, we have tried to think carefully about both basic and idiosyncratic aspects of qualitative studies that educational leadership scholars should consider. We contend that researchers/analysts need to determine not only the existence and accessibility of the qualitative research design and its various data collection strategies for leadership studies (e.g., interviews, observations, documents) but also its authenticity and usefulness, taking into account the original purpose, the context in which it is produced and the intended audience. As Bowen (2009) asserts, the subjective interpreter of data contained in qualitative research should make the process of analysis as rigorous and as transparent as possible. Qualitative inquiry on educational leadership demands no less. While there is a wealth of solid work in this area, there is always room for improvement, and we urge all scholars in this area to take seriously the processes and content of the work they do.

Recommended Readings

Malterud, K. (2001). Qualitative research: standards, challenges, and guidelines. *The Lancet*, 358(9280), 483–488.

Malterud raises several important issues for any scholar in an applied field to consider. In particular, he takes on the issue of transferability, which is underdiscussed in education.

Wolcott, H. F. (2001), *Writing up qualitative research*. Thousand Oaks: Sage.

Wolcott's book raises key issues related to communicating qualitative research in an engaging manner.

REFERENCES

- Aberbach, J. D., & Rockman, B. A. (2002). Conducting and coding elite interviews. *PS: Political Science & Politics*, 35(4), 673–676.
- Barnhardt, R., Chilcott, J. H., & Wolcott, H. F. (1979). *Anthropology and educational administration*. Tucson: Impresora Sahuaro.

- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods*. Boston: Ally and Bacon.
- Bogotch, I., Beachum, F., Blount, J., Brooks, J. S., & English, F. W. (Eds.). (2008). *Radicalizing educational leadership: Toward a theory of social justice*. Rotterdam: Sense Publishers.
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.
- Brooks, J. S. (2016). Everything we know about educational leadership is wrong. In G. Lakomski, C. Evers, & S. Eacott (Eds.), *Questioning leadership: New directions for educational organizations* (pp. 31–44). London: Routledge.
- Brooks, J. S., & Miles, M. T. (2010). The social and cultural dynamics of school leadership: Classic concepts and cutting-edge possibilities. In S. D. Horsford (Ed.), *New perspectives in educational leadership: Exploring social, political, and community contexts and meaning* (pp. 7–28). New York: Peter Lang Publishing.
- Brooks, J. S., & Normore, A. N. (2015). Qualitative research and educational leadership: Essential dynamics to consider when designing and conducting studies. *International Journal of Educational Management*, 29(7), 798–806.
- Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *British Journal of Sociology*, 35(1), 75–92.
- Capper, C. (1993). Educational administration in a pluralistic society: A multiparadigm approach. In C. Capper (Ed.), *Educational administration in a pluralistic society* (pp. 7–35). Albany: State University of New York Press.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks: Sage.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks: Sage.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks: Sage.
- Eacott, S. (2015). *Educational leadership relationally: A theory and methodology for educational leadership, management and administration*. Australia: South Yarra.
- Fielding, N. G., & Lee, R. M. (1998). *Computer analysis and qualitative research*. Thousand Oaks: Sage.
- Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, 204, 291–295.
- Grogan, M. (1999). Equity/equality issues of gender, race, and class. *Educational Administration Quarterly*, 35(4), 518–536.
- Harvey, W. S. (2010). Methodological approaches for interviewing elites. *Geography Compass*, 4(3), 193–205.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. New York: Suny Press.

- Honig, M. I. (2006). *New directions in education policy implementation*. New York: Suny Press.
- Horsford, S. (2011). *Learning in a burning house: Educational inequality, ideology, and (dis)integration*. New York: Teachers College Press.
- Janesick, V. J. (1994). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 209–219). Thousand Oaks: Sage.
- Knapp, M. S. (2017). The practice of designing qualitative research on educational leadership. *Journal of Research on Leadership Education*, 12(1), 26–50.
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45(3), 214–222.
- Lewis-Beck, M. S., Bryman, A., & Liao, T. F. (2004). *The SAGE encyclopedia of social science research methods*. Thousand Oaks: SAGE.
- Lochmiller, C. R., & Hedges, S. L. (2017). Education policy implementation research: A call for new approaches. In J. N. Lester, C. R. Lochmiller, & R. E. Gabriel (Eds.), *Discursive perspectives on education policy and implementation* (pp. 17–40). Cham: Palgrave Macmillan.
- Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet*, 358(9280), 483–488.
- Marshall, C. (1984). Elites, bureaucrats, ostriches, and pussycats: Managing research in policy settings. *Anthropology and Education Quarterly*, 15(3), 235–251.
- Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (Vol. 41). Thousand Oaks: Sage.
- Merriam, S. B. (1991). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Merriam, S. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Mulhall, A. (2003). In the field: Notes on observation in qualitative research. *Journal of Advanced Nursing*, 41(3), 306–313.
- Normore, A. H., & Brooks, J. S. (Eds.). (2014). *Educational leadership for ethics and social justice: Views from the social sciences*. Charlotte: Information Age.
- Patton, M. Q. (2005). *Qualitative research*. New York: Wiley.
- Peshkin, A. (2000). The nature of interpretation in qualitative research. *Educational Researcher*, 29(9), 5–9.
- Ponterotto, J. G., & Grieger, I. (2007). Effectively communicating qualitative research. *The Counseling Psychologist*, 35(3), 404–430.
- Salmons, J. (2016). *Doing qualitative research online*. Los Angeles: Sage.
- Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.
- Silverman, D. (Ed.). (2016). *Qualitative research*. Sage: Thousand Oaks, CA.
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage.

- Stake, R. E. (2008). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 119–149). Los Angeles: Sage.
- Stewart, D. W., & Shamdasani, P. M. (1990). *Focus groups: Theory and practice*. London: Sage.
- Theoharis, G. (2009). *The school leaders our children deserve: Seven keys to equity, social justice, and school reform*. New York: Teachers College Press.
- Tracy, S. J. (2012). *Qualitative research methods: Collecting evidence, crafting analysis, communicating impact*. New York: Wiley.
- Wolcott, H. F. (1970). An ethnographic approach to the study of school administrators. *Human Organization*, 29(2), 115–122.
- Wolcott, H. F. (1973). *The man in the principal's office: An ethnography*. New York: Holt, Rinehart & Winston.
- Wolcott, H. F. (1974). The elementary school principal: Notes from a field study. In G. D. Spindler (Ed.), *Education and cultural process: Toward an anthropology of education* (pp. 176–201). New York: Holt, Rinehart & Winston.
- Wolcott, H. F. (1977). *Teachers versus technocrats: An educational innovation in anthropological perspective*. Eugene: University of Oregon, Center for Educational Policy and Management.
- Wolcott, H. F. (1982). The anthropology of learning. *Anthropology and Education Quarterly*, 8(2), 83–108.
- Wolcott, H. F. (2001). *Writing up qualitative research*. Thousand Oaks: Sage.
- Wolcott, H. F. (2002). *Sneaky kid and its aftermath: Ethics and intimacy in fieldwork*. Walnut Creek: Rowman Altamira.
- Yin, R. K. (1994). *Case study research: Design and methods*. Thousand Oaks: Sage.



The DIVE Approach: Using Case-Ordered Meta-Matrices and Theory-Based Data Displays to Analyze Multiple Case Study Data

Susan Bush-Mecenas and Julie A. Marsh

Case study methods are widely used and well suited to help researchers understand questions of *how* and *why* in the study of educational policy and leadership. Defined as “a contemporary phenomenon within its real-life context, especially when the boundaries between a phenomenon and context are not clear” (Yin 2013, p. 13), case study considers not only the object or process of interest but also its environment and surrounding conditions. Much like other qualitative methods, case study can help address questions of “process rather than outcomes” (Merriam 1988, p. xii). Uniquely, case studies may allow researchers to understand policy mechanisms more deeply and suggest potential causal links (according to local causality, or “the actual events and processes that led to specific outcomes”), particularly when a phenomenon is not under the researcher’s control and/or results in multiple outcomes (Maxwell 2012, p. 132).

S. Bush-Mecenas (✉) • J. A. Marsh
Rossier School of Education, University of Southern California,
Los Angeles, CA, USA
e-mail: sbush@usc.edu; julie.marsh@rossier.usc.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_3

Compared to single case study, multiple case study has the potential to produce compelling interpretations with enhanced external validity (Merriam 1988). “By looking at a range of similar and contrasting cases,” Miles and Huberman (1994) argue, “we can understand a single-case finding, grounding it by specifying *how* and *where* and, if possible, *why* it carries on as it does. We can strengthen the precision, the validity and the stability of findings” (p. 29). Thus, multiple case study involves analyzing case processes and outcomes with attention to how they are qualified by local context (Merriam 1988; Miles and Huberman 1994). From this point, researchers may “build a general explanation that fits each of the individual cases, even though the cases will vary in their details” (Yin 2013, p. 112). Of course, drawing such conclusions across multiple cases requires rigorous and complex analysis, with attention to the tradeoffs involved in deeply considering each single case in context while also observing themes across cases.

Multiple case study is particularly well suited to educational policy research because of the emphasis on examining the process of policy implementation, including *what*, *how*, *where*, and *why* the policy played out as it did. While case study is a common technique in educational policy research, analytic methods vary widely among researchers. These studies typically draw upon large quantities of data such as interviews with multiple individuals in several case study schools, districts, colleges or universities, systems, or states; observations of multiple events; documents; and even survey data. While most use coding and memoing to organize and interpret data, the sheer volume of data can make it difficult to analyze across multiple case studies. Moreover, many of the existing analytic strategies are designed to reach either evaluative aims (e.g., understanding program outcomes and effectiveness and how the program is experienced and implemented; Greene 2000; Patton 2002) or purely theoretical aims (e.g., using the cases to generate generalizable principles and assertions about broader phenomena; Yin 2013). Yet educational policy and leadership researchers often wish to address both goals concurrently—to generate implications for policy and practice and contribute to theory, allowing for implications that extend beyond the particular policy under study (Stake 2005). Together, these concerns necessitate a systematic approach that at once embraces qualitative research traditions and lends rigor to multiple case study analysis.

In this chapter, we describe one rigorous method for analyzing multiple case study data that advances both evaluative and theoretical goals. For simplicity, we refer to this method as DIVE because it involves taking a

deep dive into multiple case data through four stages of analysis: describe, integrate, visualize, and expand. Specifically, this approach involves cyclically analyzing case-ordered meta-matrix displays—tables with cases as rows and theoretical and empirically driven constructs as columns (Miles and Huberman 1994; Miles et al. 2013; Averill 2002; LeCompte and Schensul 1999)—and overlaying empirical cases on theoretically informed conceptual diagrams to understand theoretical propositions and uncover patterns among variables. We do not purport to have invented these methods or claim that we are the first to combine them. Rather, we explicate a structured way to use these methods for novice and developing researchers. Further, we do not intend to imply that the procedure should be followed lockstep; rather, we draw upon two of our previous studies to demonstrate variation in how to use these methods to serve the research questions and data at hand.

To use this method, we begin with traditional qualitative methods: first-pass coding and within-case analysis, moving to cross-case analysis and the development of second-pass codes. Next, all coded data are entered into a matrix, with cases as rows and coding categories/variables as columns, and data are synthesized to surface patterns and relationships among cases and constructs. Then, empirical cases are superimposed upon theoretically informed diagrams to examine case data along multiple variables concurrently. Finally, we explore the case narratives, cross-case outlines, matrices, and diagrams together through memoing and writing. These steps form the basis of inquiry into both the descriptive (representing *how* the policy is implemented) and explanatory (addressing *why* or *what* conditions relate to how implementation proceeded) dimensions. We suggest that this analytic strategy can be used to strengthen and deepen multiple case studies and better understand the policy implementation. When used rigorously, the DIVE approach helps to enhance validity, promote theoretical interpretation beyond simple categories or themes, and illuminate relationships across themes.

To describe the DIVE approach, we draw on examples from two empirical studies of educational policy implementation. In each study, we analyzed a range of data (including in-depth interview, observation, survey, and document data) across a sample of K-12 public schools. In what follows, we discuss why the DIVE approach is well suited for multiple case study analysis. Then, we describe the four-stage procedure in detail, specifying our use of software to organize data. Finally, we recommend ways to address validity concerns and identify limitations to this approach.

REVIEW OF RELEVANT LITERATURE

Why DIVE? There are several analytic techniques that are commonly applied to multiple case study. Most fall within the general category of thematic analysis (Braun and Clarke 2006), which, at a minimum, involves organizing and coding data to illuminate major categories or themes. Through coding, “the undigested complexity of reality” (Patton 2002, p. 463) in field notes and interview transcripts is indexed for later retrieval (Coffey and Atkinson 1996; Miles and Huberman 1994). Codes can be thought of as abstract representations of ideas, objects, or phenomena (Strauss and Corbin 1990) and can be categorized by their analytical level. For example, codes may be classified as descriptive (e.g., what is happening), inferential (e.g., under what conditions it is happening), or explanatory or pattern coding (e.g., why it is happening) (Miles and Huberman 1994). At once, coding makes data more manageable and makes it easier to discern pattern and themes, thus helping to ensure validity by linking ideas back to supporting data (Richards 2014; Richards and Morse 2012).

Multiple case studies, however, often include large quantities of varied data, gathered through interviews, observations, documents, and surveys. The process of comparing within and across cases is complex, particularly if the goal is to understand relationships among various constructs and contextual variables. Organizing and synthesizing such large volumes of data necessitates additional strategies to enhance the validity of findings.

In response to these challenges, some have recommended structuring multiple case data through the use of Qualitative Comparative Analysis (QCA). QCA involves assigning stringent, dichotomous categorizations to qualitative data (e.g., classifying a case as having strong leadership or not having strong leadership) and utilizing Boolean algebra and logical inference to find commonalities between different cases with the same outcome (Rihoux and Ragin 2008). This method allows the researcher to synthesize data over a range of 5–55 cases (usually around 22 cases and about six variables), which would otherwise be extremely difficult to analyze using traditional methods (Ragin 2008; Mello 2012). Along with this powerful analysis, however, come severe limitations to the nuance and complexity of understanding that typically emerge from qualitative analysis. QCA draws the relationship between a constellation of factors associated with a particular outcome (all categorized as presence or absence) and is, thus, narrower in scope than traditional qualitative meth-

ods that allow for identifying patterns or relationships among various constructs (which can be categorized in nondichotomous categories).

Thus, analytic approaches on each end of the spectrum come with challenges. On the one hand, thematic analysis may be perceived as overly reliant on researcher's subjectivity and/or may raise questions about rigor and validity (Clarke and Braun 2013). These challenges are intensified when researchers deal with the large quantity of data associated with multiple case study. Further, in the educational policy context, traditional concepts of scientific rigor associated with quantitative statistical and causal analyses are generally perceived as more robust and valid than classical qualitative methods. On the other hand, QCA involves quantifying qualitative data in order to objectively identify patterns across cases. This strategy, however, sacrifices the nuance and contextual understanding afforded by qualitative data and analysis.

As a more moderate approach, several researchers have advocated the use of displays—"a visual format that presents information systematically"—to help organize and condense multiple case data, illuminate patterns, and understand conceptual relationships (Miles and Huberman 1994; Miles et al. 2013, p. 108). As Miles et al. (2013) argue,

Credible and trustworthy analysis requires, and is driven by, displays that are focused enough to permit a viewing of a full data set in the same location and are arranged systematically to answer the research questions at hand. (p. 108)

There are myriad displays used to analyze data, including network diagrams linking constructs, process diagrams or logic models, and cognitive maps. Utilized across the fields of health, business, and social policy, one promising type of display involves creating tables (also referred to as matrices, case-ordered meta-matrices, or frameworks) in which rows represent cases and columns represent variables (e.g., Averill 2002; Gale et al. 2013; Ritchie and Spencer 2002). Rather than a staid approach, matrix construction "is a creative yet systematic task that furthers your understanding of the substance and meaning of your database, even before you begin entering information" (Miles et al. 2013, p. 113). One benefit of matrices is that they approximate a more scientific approach, attractive to policymakers and leaders accustomed to quantitative statistical and causal analyses, while retaining the complexity and richness of qualitative data.

Thus, we propose a four-stage approach to multiple case analysis with matrix displays at its core. In DIVE, we combine traditional methods of within- and cross-case analysis (such as coding, memoing, and

case narratives) with case-ordered meta-matrices and unique visual displays superimposing empirical cases on relevant theoretical diagrams. We seek to understand the nuance of cases and identified emergent findings, test conceptual relationships and illuminate patterns among cases, and bring together the empirical and theoretical in conversation. Thus, DIVE represents an approach to balancing rigor and nuance, objectivity and subjectivity, theoretical and empirical, and qualitative and quantitative. Next, we describe this analytic procedure in detail.

USING THE DIVE APPROACH

In this section, we describe in detail our use of case-ordered meta-matrix and our approach to visually displaying results based on theory, using examples from two published studies: one utilized democratic theory to frame an analysis of parent engagement in a district-wide school reform policy (Marsh et al. 2015) and the other utilized accountability and organizational theories to examine teacher and administrator responses to a state-wide teacher evaluation policy (Marsh et al. 2017).

The first study examined parent engagement in a school improvement effort in the Los Angeles Unified School District, using democratic theory to frame our understanding of the quality and nature of engagement across cases. In this effort, the district identified a set of low-performing and newly built schools and solicited “bids” from internal and external operators to take over operation. The selection process was intended to promote and rely upon parent input. Drawing on data over three years from 16 cases (including 60 interviews, 16 focus groups with 112 people, 138 hours of observation, administrative data, documents, and survey data), we explored the following research questions: (1) How did the design of parent engagement mechanisms in this improvement effort change over time? (2) To what extent did the improvement effort play out as intended in local communities? and (3) What factors shaped the evolution of parent engagement policy and practice over time?

The second study examined the implementation of a state teacher evaluation policy in New Orleans schools, using both accountability and organizational theories to guide our understanding of teacher and administrator responses to evaluation. Under this state policy, schools were expected to measure teacher effectiveness via observations and student performance data to inform staffing decisions and instructional improvement. Drawing on one year of data in eight cases (including administrative

The DIVE Approach to Multiple Case Study Analysis

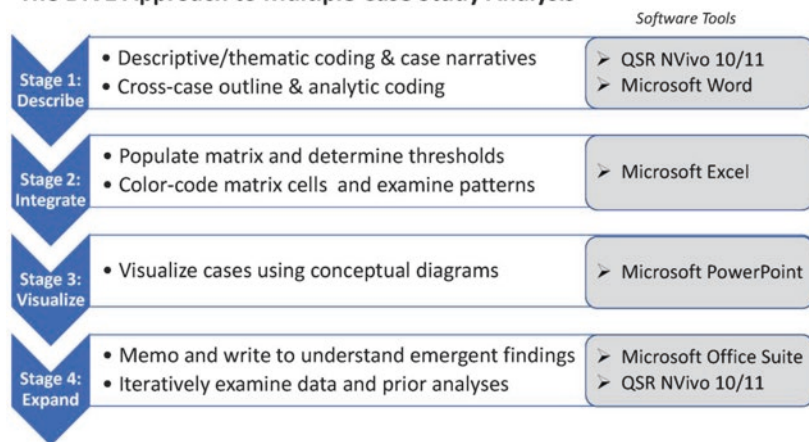


Fig. 3.1 The DIVE approach and software tools

data, documents, and 56 interviews), we explored the following research questions: (1) How and to what extent does the design and implementation of state-driven teacher evaluation policy vary across school settings? and (2) What organizational factors are associated with variation in school implementation?

Both studies shared a common purpose that lent itself to the use of case-ordered meta-matrix analyses: to examine policy implementation across multiple cases while expanding on theory and informing policy and practice. Throughout the following discussion, we will utilize examples from these two studies. Next, we demonstrate how we analyzed data in these two cases through the four stages of our DIVE approach: (1) describing and interpreting the cases using within- and cross-case analysis, (2) integrating case data using matrix displays, (3) visualizing cases over conceptual diagrams, and (4) expanding analyses iteratively through memoing, review, and writing (see Fig. 3.1 for an overview of the DIVE approach).

Stage 1: Describe the Cases and Interpret Emergent Themes Using Theory In each study, we began by organizing all transcribed interviews/focus groups and detailed observation fieldnotes in QSR NVivo 10/11, qualitative data analysis software (QDAS). We then coded our data (in NVivo) using a set of codes defined based on our initial research questions and

basic categories of implementation (e.g., perceived purpose, training, communication). Next, we organized coded data into case write-ups that traced the narrative of each case (using Microsoft Word software). This allowed us to understand policy implementation within the specific context of each case. Specifically, we developed a standardized outline listing case characteristics and basic categories of implementation. Under each outline item, we mined data from interview transcripts and observation notes, wrote summarizing statements, documented evidence (e.g., “five of six teachers endorsed teacher evaluation program”), and included one or two illustrative quotes. Through this initial first-phase coding and write-up process, we traced the narrative of each individual case embedded within its unique context.

We next completed a cross-case analysis (using Microsoft Word), aggregating evidence across cases on each broad category (e.g., code and write-up section) in outline format. Initial coding, case write-ups, and cross-case analysis allowed us to understand general themes in implementation, supporting our evaluative aims. Specifically, we were able to examine how each policy was messaged, rolled out, and implemented across case sites, and the early perceived outcomes of the policy. For example, in our parent engagement paper, we read our cross-case outline and compared the number of parent meetings and their level of attendance across cases, the content of discussion during those meetings, and parent perceptions of their role in the advisory voting process. These simple categories allowed us to examine the fidelity of implementation and variation across case contexts. Notably, in both of the projects we describe, our initial focus was on understanding policy implementation, and theoretical connections were subsequently drawn inductively from the early stages of analysis.

From the initial case analyses, several themes emerged that led us to review the extant theoretical literature. In our teacher evaluation study, for example, we noticed that teachers, administrators, and schools (as organizations) appeared to behave in very different ways in response to this state policy. We drew on accountability and organizational theories to frame the different ways schools might respond to teacher evaluation that were emerging from the qualitative data. This informed a theoretical framework to guide analysis and a secondary set of codes to delve deeper into three categories of school responses to teacher evaluation: reflective, compliant, and distortive responses (see Fig. 3.2). Reflective responding schools engaged in reportedly meaningful reflection (i.e., producing and using

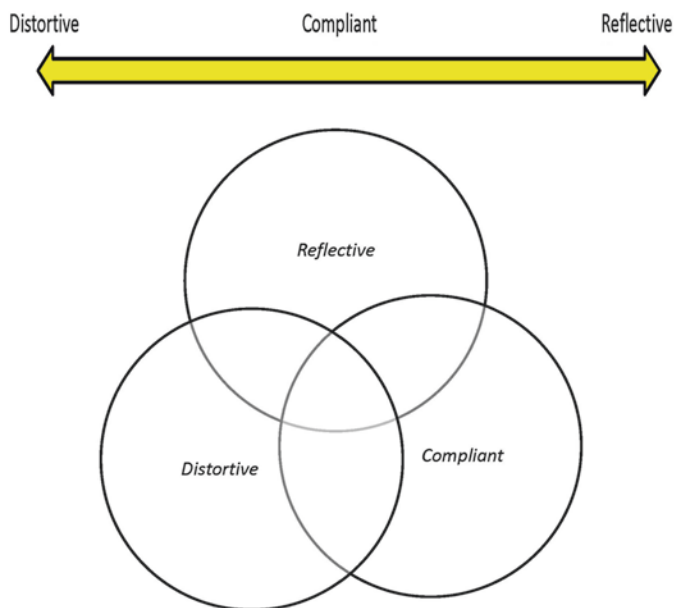


Fig. 3.2 Illustrative unpublished conceptual framework used during analysis of the Marsh et al. (2017) study

evaluation data to think about instruction and ways to improve it) and, in some cases, taking actions to bring about improvement. Compliant schools followed technical requirements but did not embrace the policy as a means to reflect on or change practice. Distortive school responses precluded reflection altogether, by taking actions that led to actual or perceived invalid measures of teaching practice (e.g., “gaming” strategies). Using these categories, we again coded our data in NVivo.

In this initial theoretical framework, we assumed that overall response types would fall along a continuum, with individual events falling into one response type (reflective, compliant, or distortive). Through coding, however, we discovered that individual examples from our data sometimes demonstrated multiple response types and that cases often exemplified hybrid response types. It became clear that schools might demonstrate responses that simultaneously reflect multiple categories. As such, we revised our initial framing to conceptualize these categories as overlapping.

In this way, we iteratively considered theory and emergent themes within each stage of analysis.

In another example, our parent engagement study utilized the principles of deliberative democratic practice to understand patterns of participation. During our stage one cross-case analysis, we noticed variation in the number of participants in parent meetings and the content of their conversation. Based on this observed theme and our review of the literature, we further coded and organized our data in NVivo along two main dimensions. One dimension examined *who* was involved, ranging from representative (involving a limited set of individuals representing parent interests) to participatory (involving all parents likely to be affected by the outcome). The second dimension examined *how* and *for what purpose* parents were involved, ranging from interest-based (participants seek to advance individual interests via private voting) to deliberative (participants seek to advance the communal good via public discussion) democratic engagement.

In developing these codes for theoretical analysis and visualization during the second stage, it was critical to consider the directionality of the concept and thresholds for different qualities or levels of a construct. In the parent engagement paper, this took the form of some inferential and explanatory coding and the categorization of existing variables. For example, we believed that the turnout for a parent voting process would be one indicator of the representative-participatory continuum describing “who” is engaged, so we drew upon attendance data and counts of those who voted. To address the “what” on the deliberative-interest-based continuum, we paid special attention to observed discussion from parent meetings and focus groups.

Clearly, the way we operationalized these theoretically informed categories varied across these two studies. In the parent engagement study, we drew upon multiple sources of evidence, setting thresholds to define categories after coding. In the teacher evaluation study, we coded along what we initially assumed were mutually exclusive categories (later aggregating the data to fit a more nuanced conception of the categories). This second-phase coding is crucial in the DIVE approach, as these empirical and theoretical considerations will form the basis for constructing the case matrix. Specifically, descriptive, inferential, and explanatory variables used in matrix analysis may represent conditions, processes, behaviors, and outcomes involved in policy implementation but should be part of a coherent framework (e.g., theory of action, logic model, typology, or taxonomy)

rather than an assortment of random themes. This approach facilitates comparison across cases and the construction of a cohesive theory of action. We note that, while this approach was well suited to our studies, more theoretically informed studies might begin with a deductively determined theoretical framework or set of constructs to guide analysis from the beginning. In such cases, researchers may use this process of understanding cases and cross-case themes in order to confirm their theoretical framework or may skip forward to the second stage of analysis.

Stage 2: Integrate and Synthesize Data Using Matrices Once within and cross-case analyses were complete, we created and populated our matrix. In this step, we brought together coded material and other evidence (e.g., survey and administrative data) in conversation, designing a matrix that included all data types and relevant variables, themes, or codes. We began by defining the matrix architecture: rows represented cases, using case pseudonyms or numbers; the columns represented codes, theoretical constructs and categories, and other data.

We purposefully made use of Microsoft Excel software to organize findings into matrix cells because of the flexibility it affords. NVivo does have the capacity to group data in frameworks or matrices that show numerical data in cells, linked to coded textual data. We discuss later how these quantified representations of data can be useful in specific contexts. For our matrix analysis, however, we found that Excel allowed us greater flexibility to view textual data, write summary statements, add related quantitative data points (such as survey data), and track decision-rules regarding data coding and categorization. We were also able to color-code cells to reflect the strength or quality of cases on specific variables (e.g., green for high instructional leadership, yellow for medium, red for low), which allowed us to examine patterns across variables and cases more easily.

Specifically, in our two studies, we grouped the columns as follows (see Fig. 3.3). The first several columns contained descriptive coding regarding the case characteristics: school size, student achievement results, governance, and so on. The next set of columns were key, theoretically informed codes pertaining to the overall implementation outcome: the school's overall evaluation response type or type of democratic engagement enacted. The final set of columns included coded data regarding contextual and organizational conditions, like leadership quality and school climate.

	Author	Governance Model	Level	Achievement (report card grade)	Student Demographics	School Size	Evaluation System	Response Type	Leadership Type	Leadership Style	Structures for Collaboration	Content, Frequency of Collaboration	Professional Culture	School Climate/ Student Expectations	Mission, Goals	Individual Capacity	Organizational Capacity	Buy-in	History	Autonomy	Organizational Networks	Use of Technology	
Gull																							
Nighthawk																							
Plover																							
Heron																							
Pelican																							
Shearwater																							
Longspur																							
Finch																							

Fig. 3.3 Illustrative unpublished case-ordered meta-matrix used during qualitative data analysis for Marsh et al. (2017)

To populate the matrix, we began by copying coded data from our case narratives and NVivo coding into appropriate cells (e.g., data from the “Gull” case school coded for “leadership style” were copied into the cell intersecting the Gull row and leadership column). Next, we went about synthesizing and condensing the data to make it manageable for analysis by summarizing the data in each cell. This summary often included the character of comments and details regarding the number of respondents or source of data and was typically accompanied by a selection of illustrative quotes. For example, the cell summary for our Gull case on “leadership style” might read:

There are few administrators and no formal teacher leadership. Four of five teachers report that principal provides “some freedom” then “tightens up”. In cases of ineffective teacher rating, teachers receive “advice” following observations, which most describe as not useful (“Do I think the administrators sometimes think they’re helping? They do, I think, but not all the time.” Teacher 4)

By condensing the data in this way, we systematically analyzed the data for each coding category within each case, lending discipline to the analysis and providing the basis for comparison across cases.

Although we now had condensed summaries in each cell, we needed further synthesis to make comparison across multiple cases and variables possible. To accomplish this goal, we set thresholds categorizing the strength or quality of cases on specific variables to allow for comparison across cases (see Fig. 3.4). These thresholds were set using three kinds of criteria: *numerical*, *prevalence*, and *quality*. No matter what method of data synthesis is used, it is critically important to record the criteria and decision-rules used to determine thresholds for each column.

For *numerical* data like surveys, we determined thresholds using distribution cut-offs and existing metrics. For example, for one survey item, we used quartiles to compare cases against one another or the total population (e.g., case school A was in the lowest 25% of cases in number of parents attending meetings of the full population schools involved in the reform). For another, we utilized the turnout rate for local elections (at 11%) as the standard for “high” engagement used to compare local school parent voting turnout rates. While this level of turnout might seem low, we believed that research on local elections helped us to set a realistic target for expected election turnout.

Dimensions of Parent Engagement														
Who					What									
	Vote count/ (bare count)	% Vote or Rubric Count/Total Parent Population	% Meeting Attendance/ Total Parent Population	Non-LAUSD outreach to parents	Communicated Purpose of Meetings ^{***}	Level of Parent Confusion Regarding Purpose	Method of Appeal to Parents	Adequacy of Parent Feedback ^{****}	Public Nature of Decisions Reported Level of Understanding	Language & Transition Issues	One-way versus two-way conversation	Personal interest, small group interest, community interest ^{****}	Parent perception of accountability	DTI perception of accountability
PSC 2.0														
1	3	3	3	3	A, C	2	2	3.5	2	3	1	6, 1, 0	1	1
2	1	1	2	2	A, C	1	1	2	2	1	1	4, 0, 3	3	1
3	1	1	1	1	A, C	1	1	2	1	3	1	7, 2, 4	1	1
4	3	3	3	3	A, B, D	1	2	3	1	1	1	15, 4, 8	3	1
5	2	1	2	3	B, D	3	1	2.5	1	3	2	6, 3, 4	1	1
PSC 3.0														
6	2	3	3	1	B, C, D	1	3	2	1	1	2	1, 2, 1	1	1
7	1	1	1	1	D, E	1	1	2	1	1	3	7, 1, 0	3	1
8	1	1	1	2	B	1	1	3	1	1	3	17, 3, 4	3	1
9	1	1	1	3	E	2	3	1	2	2	1	21, 3, 5	3	1
10	1	1	1	1	B, C, E	2	1	4	3	3	3	5, 2, 3	1	1
11	1	1	2	1	A, B, C	1	1	3	2	2	3	10, 2, 14	1	1

* Ratings as follows: 1=low, 2=medium, 3=high

** As gather information regarding educational programs, B= parent learning/capacity building, C= evaluate plans, D= provide feedback to writing teams, E= encourage parent advocacy

*** Mean survey responses from design team leaders at cases. Stem: "Our team received adequate feedback from parents and the community on our school plan." Responses: 1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree

**** Number of parent questions and/or comments in meetings coded as reflecting personal interests, small group interests, and community interests

Fig. 3.4 Illustrative unpublished case-ordered meta-matrix used during analysis for Marsh et al. (2015)

For other categories, we considered the *prevalence* of responses. That is, what proportion of respondents expressed a particular perspective? For example, in the teacher evaluation study, we wanted to understand overall school response to teacher evaluation. As such, we thought it was important to understand whether reflective or compliant responses were expressed by all interviewed teachers and administrators or just a subset. We also considered whether it was only administrators and not teachers who reported a particular response type. Depending on the nature of coding, frequency of coding by words or references (i.e., coded passages) might help to illuminate patterns. For example, we also drew upon NVivo-generated framework matrices demonstrating the percentage of words coded at each response type, as well as the percentage of both teacher and administrator cases in which such responses were found, as an indicator of the prevalence of reflective versus compliant versus distortive responses in a particular case.

Of course, the *quality* of data cannot necessarily be ascertained using *numerical* or *prevalence* criteria. Therefore, we utilized *numerical* and *prevalence* thresholds around coding with caution. Specifically, measures of response coding (i.e., percentage of words coded, proportion of teachers and administrators coded at a certain response type) were only used in combination with careful assessment of the quality of response data (e.g., an interview quote about a specific instance when a teacher removed a student from the classroom to improve evaluation ratings might be more indicative of a distortive response, compared to a general statement that a teacher modifies instruction to improve their rating) to determine the overall case categorization. Thus, *numerical* and *prevalence* thresholds of coding helped to confirm our assessments of the *quality* of responses.

In fact, most variables relied upon thresholds set based on the *quality* of data. For example, for the code “shared instructional leadership,” we defined what low, medium, or high levels of this construct might look like. A case defined as “high” on shared instructional leadership might have many teachers in leadership roles, teachers reporting strong support for the leadership, reciprocity among teachers and administrators, and consistent administrator and teacher involvement in instructional improvement. In this way, we sought to retain a nuanced understanding of how our cases represented and demonstrated key inferential and explanatory constructs.

Finally, we ordered (and reordered) our cases according to key constructs or implementation outcomes. In our teacher evaluation study, for example, we ordered cases from most reflective to most distortive. We

then color-coded cells according to the thresholds described above. For example, high levels of shared instructional leadership were colored green, medium levels colored yellow, and low colored red. Shading cells in this way allowed us to look across the matrix to examine patterns crucial to our *what* questions: What design features appeared to be related to different responses to teacher evaluation? What organizational contextual factors (e.g., leadership, history with evaluation, governance, autonomy) seem to be related to reflective responses to teacher evaluation? Here, color-coding according to each construct or variable thresholds made it easier to discern patterns. For example, once cases were arranged from green (reflective) to yellow (compliant) to red (distortive) on their response type, it was easy to examine the color array to see if responses appeared to be related to variables such as teacher collaboration or school climate. We tested additional patterns by rearranging cases, for example by their thresholds on shared instructional leadership, to illustrate the interrelations among variables or constructs.

Stage 3: Visualize Cases Using Conceptual Diagrams Matrix analyses, so far, allowed us to answer some of our *what* questions and unearth patterns among variables. Next, we sought to answer our *how* questions (e.g., Overall, how did schools across the district respond to evaluation? Overall, how did schools across the district enact parent engagement?) and examine the relationship of several variables or constructs across a third dimension, such as time, using conceptual diagrams. By visualizing cases on the theoretically informed diagrams (created using Microsoft PowerPoint), we were able to illustrate the overall implementation patterns and variation.

To situate the cases on these diagrams, we utilized the construct thresholds (high/medium/low) from our matrix analysis (see Fig. 3.5). For example, for our parent engagement cases, our cross-case analysis indicated that all cases fell in the more participatory range, quadrants 2 and 3. We then divided these quadrants into thirds and situated our cases according to the matrix ratings on three categories along the participatory-representative continuum and three categories along the interest-based-deliberative continuum.

Examining these displays in our parent engagement case, we found that Los Angeles case schools implemented parent engagement in ways that were primarily interest based and participatory, particularly in the early



Fig. 3.5 Case visualization from Marsh et al. (2015). Reprinted by permission of SAGE Publications

years. We also found that in later years of the reform, following a key shift in policy, schools implemented substantially more deliberative (though comparatively participatory) forms of engagement relative to those in the first year. Thus, situating empirical cases on conceptual diagrams can help to illuminate patterns over time as well as across theoretical constructs.

For our teacher evaluation study, we positioned our cases on the Venn diagram according to overall response type, determined using a combination of the quality of responses, and NVivo-generated numerical measures of the proportion of respondents reporting a particular response, and the frequency of words coded for each response type (see Fig. 3.6). Next, we added other variables to examine patterns among case characteristics and key constructs. For example, we considered whether school governance (e.g., charter versus traditional school type and authorizer) might be related to responses by adding visual indicators of each case's governance model.

We showed that not all New Orleans case schools responded to evaluation in the state-intended reflective ways and that school governance

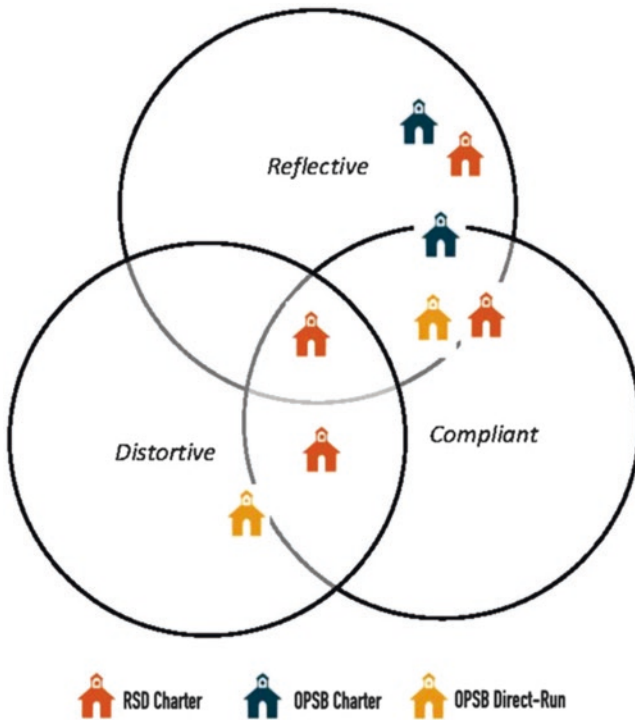


Fig. 3.6 Case visualization used during analysis related to Marsh et al. (2017). (Note: Three public entities are responsible for charter contracting in New Orleans—the RSD, OPSB, and the Louisiana Board of Elementary and Secondary Education (BESE). Several different types of schools operate in New Orleans, including traditional schools overseen by OPSB, charters in a CMO network, and single-site charters)

model had little to do with their response type. In this way, these displays concurrently lent rigor to our analysis and provided the flexibility and space to question, develop, and build on our theoretical framework.

Stage 4: Expand Analyses Through Iterative Memoing, Review, and Writing Finally, we found it essential to return to the data and all three prior stages of analysis to examine our *why* questions. For example, as we sought to understand why we saw varying responses to teacher evaluation,

we reflected on our matrices and data displays and returned to our case narratives to further understand how other issues might be at play. By reflecting on specific case narratives, we began to question how an individual's or school's goal orientation might be related to response. We asked:

Could rigorous teacher evaluation policies eventually modify organizational context and/or goal orientations? One might hypothesize that the experience of “putting on a show” could serve as a mastery experience, facilitating teacher learning. As such, even distortive responses might lead to organizational and individual learning, albeit at a slower pace. (Marsh et al. 2017, p. 26)

This insight formed an important point for discussion: reflecting on our individual cases after systematically comparing across cases generate a new line for consideration and future study.

In each case, the DIVE approach allowed us to move beyond our major policy and implementation findings, surfacing key theoretical learnings. For example, in our parent engagement study, we discovered the crucial importance of trust to democratic practice. Trust emerged inductively during second-phase coding, and we observed patterns of trust relating to more deliberative models in our matrix and data displays. Thus, we affirmed and offered direction for further theory building around that relationship.

We also sought feedback from colleagues on our work as we moved into the writing process and reanalyzed data in response to questions and queries that arose. For example, in the teacher evaluation study, our analyses were initially completed according to school-level cases. During peer review, we were asked if individual characteristics might be thought to influence response types: Were less experienced, younger teachers more reflective and veteran teachers more compliant? Because all data were already coded in NVivo, we were able to create individual respondent (administrator and teacher) cases and apply relevant characteristics (years of experience, gender, race, subject taught, position, etc.). Using NVivo's framework matrices function arranged by these cases, we were able to readily examine the responses to evaluation on an individual level. Had these analyses unearthed new patterns (they did not), we would have moved back through a revised matrix analysis.

One final tool for strengthening findings is attention to craft rivals (alternative explanations, such as the null hypothesis, research bias, or

threats to validity) and real-world rivals (like alternative theoretical explanations, societal shifts, implementation issues, etc.) to explanatory findings (Yin 2013). Essentially, this involves testing for robustness by questioning the underlying logics and implications of findings and returning to earlier stages of analysis to question the data. For example, in our teacher evaluation study, we found that many schools responded to state teacher evaluation policy in compliant or distortive ways. Rather than assume that such responses were indicative of only school conditions or staff mind-sets, we considered the possibility that, despite the policy's intent, staff development and improvement were accomplished through other school mechanisms (such as mentor teachers, peer observation, or informal classroom walkthroughs). We also considered whether other compliant or distortive responses to policy could, in fact, be an adaptive response to a "bad" policy. Returning to the data to question our findings improved our confidence in the validity and rigor of our analyses.

As these examples illustrate, all four stages of DIVE can and should be used iteratively. We think of this process as repeatedly varying the focal length; we zoom in to understand case context, we zoom out (by simplifying data) to understand patterns and surface themes across cases, and we zoom back in to understand why these patterns arose. Memoing was a valuable tool for adjusting the analytic focal length, and recording each inquiry and findings, throughout each stage of the DIVE approach. Indeed, the writing process itself served as one final analytic tool, as we brought together within- and cross-case, narrative and thematic, matrix and diagram, and empirical and theoretical.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

The DIVE approach sits between methodological paradigms. While some findings may have a more scientific, quantitative feel (e.g., discussion of testing hypotheses and examining independent and dependent variables, setting thresholds, and examining numerical representations of qualitative data), the underlying logics are firmly rooted in qualitative traditions. Given this tension, we recommend that novice and emerging scholars pay careful attention to issues of sampling, reporting, and generalizability in advance of completing analyses and as they interpret studies using the DIVE approach.

For example, while external validity remains an important concern, the aim in case study research is analytic generalizability. That is, case study

findings are generalizable to theoretical propositions but not to specific populations (as in statistical generalizability). Likewise, sampling logic is rooted in a statistical conceptualization of generalizability and causality, which assumes that a random sample of a population will demonstrate statistically similar outcomes to the study population as a whole (within a limited range of variability). The limited sample sizes and differing purpose of qualitative methods preclude such comparisons (Maxwell 2012). Instead, Yin (2013) suggests that researchers consider replication logic when conducting multiple case studies. That is, case studies may be selected based on either a prediction of similar results (literal replication) or contrasting results due to predictable reasons (theoretical replication). This sampling perspective bounds the kinds of generalized findings we may take from multiple case analysis.

These considerations are critically important, not just while defining theoretical precepts or designing data collection but throughout analysis and writing. Integrating and synthesizing data in matrices, in particular, call upon the researchers to carefully understand where and when numerical simplification and representation is appropriate. Noting that five of ten respondents reported school practices encouraging shared leadership may be a useful way to triangulate evidence of shared leadership. Having six of ten respondents report such practices, however, does not indicate a higher level of shared leadership for a particular case. It is also important to consider missing data in any cell, as qualitative data collection like interviewing may result in some gaps in data which cannot be interpreted. Missing data may challenge comparison. To deal with missing data, consider documenting decision-rules. What percentage of respondents need to be represented to consider data complete? Where case data are missing, should the construct be dismissed? What other data sources can be used to triangulate findings?

We suggest that there are several ways that researchers can best communicate these limitations and readers can evaluate research that uses the DIVE (or similar) approach. First, in our final publications, we carefully stipulate the limitations of our research, particularly regarding generalizability and how quantified representations of qualitative data can be understood. Second, we include detailed information on our sampling and recruitment procedures, data collection methods, and how and when theoretical frameworks made their way into the study. Third, we specify the analytic methods, often using appendices. This includes code books that list codes along with their definitions and examples of coded material,

as well as unpopulated matrices (to protect anonymity of cases), and even NVivo framework matrix displays showing the percentage of words coded at particular codes for each case. While this kind of information can seem exhaustive, it allows the reader to clearly understand how data were collected and analyzed and, therefore, how the findings can be understood. Documenting and communicating the methods used is essential to bolstering the rigor and robustness of the DIVE approach.

CHAPTER SUMMARY

Combining traditional qualitative methods (coding, case narrative, cross-case comparison, memoing) with matrix and diagrammatic displays, the DIVE approach is a useful tool for researchers to synthesize large quantities of complex multiple case data to uncover patterns and emergent themes. As discussed, the DIVE approach is made up of four stages of analysis: (1) describing and interpreting the cases using case narrative and cross-case comparison, (2) integrating and synthesizing case data into matrix displays, (3) visualizing cases over conceptual diagrams, and (4) expanding analyses iteratively through memoing, review, and writing. Making visible the patterns and relationships within and across both cases and conceptual constructs helps to answer *what*, *how*, and *why* questions. By strengthening and deepening multiple case studies, the DIVE approach helps to enhance validity, promote theoretical interpretation beyond simple categories or themes, and illuminate relationships across themes to generate findings that are evaluative, policy relevant, or theoretically focused.

Recommended Readings

Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks: Sage.

Miles, Huberman, and Saldaña provide a detailed description of qualitative analytic methods from start to finish, with a special emphasis on using displays to organize and analyze data. An excellent text for novice and experienced researchers alike, the authors detail each analytic strategy step-by-step.

Ritchie, J., & Spencer, L. (2002). Qualitative data analysis for applied policy research. *The Qualitative Researcher's Companion*, 573(2002), 305–329.

Ritchie and Spencer detail the use of frameworks (similar to matrices discussed in this chapter) to examine qualitative data. With its focus on applied policy research and its detailed instructions from coding through setting up displays, this is an ideal text for novice researchers in educational policy and leadership.

REFERENCES

- Averill, J. B. (2002). Matrix analysis as a complementary analytic strategy in qualitative inquiry. *Qualitative Health Research*, 12(6), 855–866.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist*, 26(2), 120–123.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies*. Thousand Oaks: Sage.
- Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13(1), 117.
- Greene, J. C. (2000). Understanding social programs through evaluation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Qualitative research* (3rd ed., pp. 981–999). Thousand Oaks: Sage.
- Le Compte, M. D., & Schensul, J. J. (1999). *Analyzing & interpreting ethnographic data*. Walnut Creek: Altamira Press.
- Marsh, J., Strunk, K., Bush-Mecenas, S. C., & Huguet, A. (2015). Democratic engagement in district reform: The evolving role of parents in the Los Angeles Public School Choice Initiative. *Educational Policy*, 29(1), 51–84.
- Marsh, J., Bush-Mecenas, S. C., Strunk, K. S., Huguet, A., & Lincove, J. (2017). Evaluating teachers in the big easy: How organizational context shapes policy responses in New Orleans. *Educational Evaluation and Policy Analysis*, 39, 539–570.
- Maxwell, J. A. (2012). *Qualitative research design: An interactive approach: An interactive approach*. Thousand Oaks: Sage.
- Mello, P. A. (2012). Parliamentary peace or partisan politics? Democracies' participation in the Iraq War. *Journal of International Relations and Development*, 15(3), 420–453.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.

- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks: Sage.
- Patton, M. Q. (2002). Designing qualitative studies. *Qualitative Research and Evaluation Methods*, 3, 230–246.
- Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond* (Vol. 240). Chicago: University of Chicago Press.
- Richards, L. (2014). *Handling qualitative data: A practical guide*. Thousand Oaks: Sage.
- Richards, L., & Morse, J. M. (2012). *Readme first for a user's guide to qualitative methods*. Thousand Oaks: Sage.
- Rihoux, B., & Ragin, C. C. (2008). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques* (Vol. 51). Thousand Oaks: Sage.
- Ritchie, J., & Spencer, L. (2002). Qualitative data analysis for applied policy research. *The Qualitative Researcher's Companion*, 573(2002), 305–329.
- Stake, R. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Qualitative research* (3rd ed., pp. 433–466). Thousand Oaks: Sage.
- Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research* (Vol. 15). Newbury Park: Sage.
- Yin, R. K. (2013). *Case study research: Design and methods*. Thousand Oaks: Sage.



Language-Based Methodologies for Policy and Leadership Research

Jessica Nina Lester and Justin Paulsen

This chapter introduces language-based methodologies and methods, which we argue are particularly relevant to education policy and leadership studies. More specifically, in this chapter we discuss critical discourse analysis (CDA), discursive psychology (DP), and conversation analysis (CA) (including membership categorization analysis). Within the landscape of language-based methodologies, discourse and conversation analysts have long provided perspectives on how one might go about studying talk and texts (Lester 2014). In this chapter, we first provide a general discussion of language-based methodologies, giving particular attention to the field of discourse analysis and its relevance to policy and leadership research. Here, we also note some of the underlying theories and concepts that influence many discourse analytic approaches. Second, we present CDA, DP, and CA and highlight the following for each methodological approach: (1) the primary research focus; (2) key features or characteristics; (3) the relevant education policy and/or leadership literature which has employed CDA, DP, and/or CA; and (4) primary resources for those new to such

J. N. Lester (✉) • J. Paulsen
School of Education, Indiana University, Bloomington, IN, USA
e-mail: jnlester@indiana.edu; jumpauls@indiana.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_4

methodologies and methods. Third, we provide three illustrative examples, drawing upon relevant policy and leadership scholarship. We conclude with suggestions for policy and leadership scholars interested in the study of language writ large, pointing explicitly to the practical applications and implications for those who might employ CDA, DP, and/or CA.

AN ABBREVIATED OVERVIEW OF LANGUAGE-BASED METHODOLOGIES

Broadly, language-based methodologies, which conversation analysis (Sacks 1992) and discourse analytic perspectives (Wood and Kroger 2000) are included within, allow scholars to engage in a close analysis of language at varying levels and for varying purposes. We focus our discussion in this section on discourse analysis (DA) writ large. DA has been conceptualized as an umbrella term that includes within it varying theories and analytic approaches, broadly focused on the study of talk and text produced in everyday social life (Potter 2004). Across discourse analytic perspectives, it is assumed that language is performative; that is, it is assumed to always *do* something. For instance, through language people go about ascribing blame, building accounts, negotiating peace, and so on. Relatedly, discourse analysts generally assume that it is through language that the social world is built, ordered, sustained, reframed, and so on. In other words, a social constructionist position is typically taken up, which assumes that language produces the social world(s) (Berger and Luckmann 1967). Finally, it is important to note that in varying ways discourse analytic perspectives can take up a critical perspective, as they function to question taken-for-granted knowledge and practices.

Indeed, there are a multitude of distinct approaches to DA (Jorgensen and Phillips 2002), with some focused on more macro oriented perspectives to language (e.g., discourses of human rights) and others focused more on micro, everyday interactions (e.g., the way in which ‘question formulations’ mark or make visible how policymakers position some stakeholders as knowledgeable and others as un-knowledgeable). These approaches afford researchers the analytic flexibility to study a variety of relevant topics, including the ways in which the talk of policymakers shapes what and who comes to be positioned as good teachers (Gabriel and Lester 2013) as well as media representations of key education policy issues and the consequences of such representations (Piazza 2014).

More particularly, within the fields of policy and educational leadership, there are several approaches that have been positioned as particularly productive in answering pressing and relevant questions and offering critical insights about social life. While DA encompasses a variety of approaches, theories, and perspectives, conversation analysis is a single qualitative methodology that has often been positioned as distinct from DA. Thus, we use the term ‘language-based methodologies’ throughout this chapter. Given the vast differences across these approaches, it is not possible to generate a step-wise discussion of how analysis *should* proceed or even how to design a DA study. Rather, we assume that if you are to carry out a DA study, it is important to go deep in familiarizing yourself with the assumptions of a given approach—noting that concepts such as ‘language’, ‘discourse’, ‘data’, and even ‘analysis’ are often theorized in distinct ways.

AN INTRODUCTION TO CDA, DP, AND CA

It is notable that relatively little scholarship in education leadership has used language-based methodologies. In fact, in our search for relevant literature within some of the top education leadership journals (*Educational Administration Quarterly*, *Journal of Research on School Leadership*, *Journal of Educational Administration*, *Leadership & Policy in Schools*, *Journal of School Leadership*, *Educational Leadership*, *Educational Management Administration & Leadership*, *International Journal of Educational Management*), we found few relevant articles. Thus, in this chapter, we primarily draw upon examples from the education policy literature, where language-based approaches have been more commonly used, particularly CDA. Further, within the field of education policy, there have been several recent discussions regarding the use of CDA and other language-based approaches (e.g., Foucauldian-informed DA, DP, etc.) for the study of education policy. A recent special issue in *Education Policy Analysis Archives* (EPAA) focused entirely on the use of CDA for the study of education policy and included six empirical examples of the varying ways that CDA might be drawn upon to unearth critical insights related to language use and policy (Lester et al. 2016). As a follow-up to the 2016 special issue published in EPAA, a second special issue was published illustrating how diverse approaches to the study of language might further the field’s understanding of language use and education policy (Lester et al. 2017b). The articles included within the second special issue drew up a variety of language-based methodologies, including discursive

psychology (Hurst 2017), Foucauldian DA (Burman et al. 2017), and Bakhtinian DA (Wilinski 2017), among others. Collectively, these two special issues point to the relevance of language-based methodologies for policy scholars, and in one case, also highlight the potential for the field of education leadership (e.g., Hurst 2017). Further, a recent edited volume highlighted the utility of discourse analytic perspectives for the field of education policy, and included multiple empirical examples of diverse and varied methodological applications (Lester et al. 2017c).

While scholars have fairly described CDA as the most common language-based approach used in education policy scholarship (Lester et al. 2017a), we argue here that there is great potential for the use of other discourse analytic approaches. Hence, in the next subsections, we introduce first that which is perhaps most familiar to readers—CDA. Then, we move to describe DP and conclude with a discussion of CA and membership categorization analysis (MCA).

OVERVIEW OF CDA

CDA is multidisciplinary and encompasses a broad range of approaches, theories, and methods focused on the study of language and more particularly sociopolitical problems. As its name indicates, CDA foregrounds a commitment to ‘criticality’ and has historically attended to power, inequality, and/or dominance in political realms as produced in and through talk and text (van Dijk 2001). Notably, while CDA cannot be traced back to a single origin, CDA approaches are generally informed by critical theories and Foucault’s notion of power (Foucault 1990). Critical social science questions social life more broadly than just issues of efficiency and effectiveness (Fairclough 2003). Thus, most approaches to CDA pose questions regarding the ways in which discourse, ideologies, and structures (re)produce social dominance (van Dijk 2015). CDA understands ideologies to include the role of power in the “positions, attitudes, beliefs, perspectives, etc. of social groups” (Fairclough 2003, p. 9). Thus, ideologies in CDA should be shown to “contribute to establishing, maintaining, and changing social relations of power, domination, and exploitation” (Fairclough 2003, p. 9). van Dijk suggests a set of notions that are often featured in CDA, including: “‘power’, ‘dominance’, ‘hegemony’, ‘ideology’, ‘class’, ‘gender’, ‘race’, ‘discrimination’, ‘interests’, ‘reproduction’, ‘institutions’, ‘social structure’ or ‘social order’ (van Dijk 2015, p. 354). CDA is thus a means of “understand[ing], expos[ing], and ultimately challeng[ing] social inequality” (van Dijk 2015, p. 352).

Key Features of CDA As we have noted above, CDA encompasses a wide range of approaches, including Fairclough’s approach (Fairclough 1992), a discourse-historical approach (Wodak 2001), and sociocognitive studies (van Dijk 2001), among others (Fairclough et al. 2011). In addition, there are several common considerations related to CDA—some of which we describe next.

First, CDA attempts to “bridge the gap” between the societal-level discourses or ideologies and the micro-level language seen in specific conversations or texts. Fairclough described this as “interdiscursive analysis” that includes linguistic analysis as well as “seeing texts in terms of the different discourses, genres, and styles they draw upon and articulate together” (Fairclough 2003, p. 3). Thus, CDA assumes that everyday interactions explicitly and implicitly draw upon ideological norms, thereby reinforcing or challenging the discourses of power. CDA scholars analyze language to identify how ideologies work in everyday interactions.

Second, it is important to note that the very way that discourse is defined and theorized varies across discourse analytic perspectives. Within CDA, discourse is generally defined in three ways (at least): (1) as an analytical category that includes resources drawn upon for meaning-making; (2) as including words, pictures, gestures, and so on, and (3) as social practice (Fairclough 2011).

Third, CDA takes a broad view of what counts as data. Fairclough (2003) outlined the possible realm of data as including “any actual instance of language in use,” while also noting that this very description is a bit limiting as “visual images and sound effects” might be included (p. 3). In other words, there are a wide variety of data sources that can serve as possibilities for CDA scholars, including policy documents, policy conversations, and so on.

Review of the Policy Literature Employing CDA Relative to other language-based methodologies, CDA has been more commonly used, particularly in education policy scholarship. Our literature review identified well over 40 different studies using CDA to investigate education policy issues. The literature includes at least 30 different topics of focus among the studies. These included a focus on such things as specific policies (e.g., no promotion of homosexuality statutes, zero tolerance policies), specific injustices (e.g., racism against Chicana students in California, use of textbooks to constitute neocolonialisms), or trends in education (e.g., discursive shifts in defining

Norwegian school leadership roles, free compulsory basic education) (Abrahamsen and Aas 2016; Barrett and Bound 2015; Bazzul 2014; Huber 2011; Kennedy-Lewis 2014; Nudzor 2013). Additionally, the studies pursued policy or leadership topics in 20 different country contexts. The most commonly used data sources were government policy documents. However, many CDA studies analyzed news articles, policy speeches, and to a lesser extent interview and focus group transcripts.

There was great variability in how scholars took up CDA. For instance, some scholars employed CDA by conducting more traditional content or thematic analyses to identify critical themes of power, ideology, or discrimination. Therefore, they drew upon the theories that undergird CDA in combination with other more traditional forms of qualitative analysis. Other studies did not explicitly describe or evidence a methodological approach and launched into a critical evaluation of the policy, citing key CDA scholarship. Furthermore, there were some studies that engaged in traditional DA methods by examining lexical choices (e.g., nominalization), grammatical constructions (e.g., passivization, modal verbs) and stylistic choices (e.g., intertextuality, genres) yet positioned the broader goals of the work in alignment with CDA.

Key Resource for the Study of CDA As you begin your study of CDA, we recommend consulting one of Fairclough's texts, *Analyzing discourse: Textual analysis for social discourse* (2003), to develop your understanding of CDA.¹ Fairclough's texts are central in conceptualizing CDA among the studies in our literature review. This particular book is especially useful to the novice researcher as it is designed for a broad audience and is thus especially accessible. Additionally, it takes a pragmatic approach by focusing on the doing of CDA with step-by-step examples of conducting CDA from real texts.

OVERVIEW OF DP

DP has historically focused on the ways in which psychological matters are made 'real' in talk and text (Potter 2012). Such 'matters' include things like beliefs, identities, cognition, emotions, and so on, with language

¹Note for those working substantively in education: Rebecca Roger's (2011) writing around CDA in education may be particularly useful.

positioned as the medium of social action (Potter and Hepburn 2008). Traditionally, talk was conceptualized as representing inner thoughts; yet, DP scholars shift this perspective and position talk and text as producers of mind, emotions, identities, and so on. Since the earliest writings (see Edwards and Potter 1992; Potter and Wetherell 1987), DP has become an interdisciplinary perspective, spanning many fields. Further, it has been described as encompassing at least two strands of work: one aligned with CA and another positioned as more ‘critical’ and generally referred to as critical discursive psychology (Tileagă and Stokoe 2015). Potter (2012) described three historical moments in DP, which include the production of work focused on: (1) studying interpretative repertoires and engaging in more critically oriented analyses situated within Potter and Wetherell’s (1987) early work; (2) respecifying psychological matters as bound and produced in and through language-in-use; and (3) attending to the sequential features of talk and therefore informed by CA.

Key Features of DP We highlight here just three key features associated with DP, noting that indeed others exist. First, DP orients to the very notion of discourse in a particular way. Specifically, DP scholars define discourse as (1) action oriented, (2) constructed and constructive of the world, and (3) situated (at multiple levels). Notably, DP attends specifically to language-in-use (i.e., language as it is being used in actual interactions) and thereby takes a more micro-oriented approach to analysis. However, critical DP scholars often move between micro and macro claims, drawing upon poststructural perspectives in making these claims.

Second, DP has been heavily influenced by ethnomethodology and CA. As such, many DP scholars draw extensively upon CA principles, with even their data sources often reflective of CA’s commitment to naturally occurring data sources (e.g., classroom talk rather than interviews with teachers).

Third, like CA, analysts drawing upon DP typically take up what is often referred to as ‘unmotivated looking’ in which they aim to foreground the participants’ orientations within a given interaction (Sacks 1992). As such, it is not unusual for research questions to be formulated after an analyst has engaged in at least one pass of unmotivated looking, with the data itself driving the research focus. In this way, the researcher’s pet theory or agenda is assumed to be checked by the data itself.

Review of the Policy Literature Employing DP In our review of the literature, we found relatively few studies in education policy and leadership that employed DP. For example, a search of *EPAA* using the search term “discursive psychology”, resulted in only two articles (Gabriel and Lester 2013; Hurst 2017). Rachael Gabriel, a reading policy scholar, has perhaps produced the majority of DP-related education policy studies, which include both journal articles (e.g., Gabriel and Lester 2013; Gabriel and Paulus 2015) and book chapters (e.g., Gabriel 2017). Beyond this, Hurst (2017) has most recently published an article focused on the superintendent discourse as evidenced on Twitter. In this article, Hurst draws upon DP and importantly works at the intersection of policy and leadership. Given this particular article is one of the few located squarely within policy, we discuss it in detail in the section below focused on applications to policy and leadership. While DP has been rarely used in education policy and leadership, we argue that this particular discourse analytic approach offers an incisive way by which to attend to language-in-use and that which is produced in and through its use.

Key Resource for Those New to DP Sally Wiggins’ most recent text, *Discursive psychology: Theory, method, and application* (2017), is one of the few published books that offers a theoretically grounded yet practical discussion of how to conceptualize, design, and carry out a DP study. For those who are interested in exploring the possibility of carrying out a DP study, this particular text is a useful starting point. It is both accessible and provides the needed starting points for understanding the undergirding theories that shape a DP study.

OVERVIEW OF CA AND MCA

CA focuses on the study of talk-in-interaction in everyday and institutionalized contexts (ten Have 2007). More particularly, CA has historically focused on attending to the sequential nature of talk and how participants within an interaction make sense of and orient to the nuances of the interaction. Patterns of interaction, then, are typically studied by analysts closely attending to the organization of talk, with conversational structures such as turn design, repair, turn-taking, among others, being considered (McCabe 2006).

More broadly, CA scholars approach social research from an intensely emic perspective. A key assumption in CA is that social theories and

ideologies are extraneous to the interaction unless explicitly oriented to by the interactants (ten Have 2007). Talk, from this perspective, is not assumed to be reflective of internal mental processes or external concepts, but rather is constitutive and constructive in its own right. Thus, individuals co-create order in their speech, which allows researchers to identify the processes that make up this creative process. CA contends that the examination of the sequence of speech uncovers the machines that generate the course of interactions in particular contexts.

A related approach is MCA, which was developed concurrently with CA and similarly seeks to explicate commonplace theories held by individuals. MCA accomplishes this by focusing primarily on how individuals categorize the things around them. MCA uses discursive data to identify the activities, characteristics, and relationships that make up categories (e.g., teachers, students, administrators). Understanding this provides insights into the boundaries a particular group or culture places on different categories.

Key Features of CA and MCA There are indeed a multitude of features we could highlight here. As such, what we chose to share next is partial and serves as a starting point for readers.

First, in CA, each turn in the conversation is viewed as having “order at every point”, with individuals understood to be accomplishing some act with their language (Sacks 1984, p. 22). Thus, one of the fundamental aspects of analysis in CA is referred to as “next-turn proof” (Hutchby and Wooffitt 2008, p. 13). This conveys that the meaning in the sequence is understood in the way one speaker’s turn links to the turn before it; that is, how a speaker orients to a previous turn reveals the function of the initial turn. Thus, the meaning is grounded in the interactants’ interpretation rather than the analysts’ alone, so to speak. Interactions, then, are assumed to have an organization, which could be shaped by the institutional setting (e.g., classroom interactions with a teacher question-student answer-teacher evaluation pattern) in which aberrations from these norms offer other possible interaction forms.

Second, conversation analysts collect naturally occurring data; thus, analysts generally record (either audio or video) naturally occurring interactions, with some scholars also focusing on naturally occurring text-based data sources. Within CA, there is also a growing focus on computer-mediated communication data, including text-based data (Giles et al. 2015). In CA, recordings are important because researchers return

often to the source that captured the interaction (Hutchby and Wooffitt 2008). Naturally occurring interaction refers to data that exists regardless of the presence or agenda of a researcher. CA brackets other data or information that traditional research would want to account for (e.g. demographic or background characteristics of the speakers, member checking, etc.) and focuses exclusively on the talk/text itself.

Third, one key aspect of carrying out a CA study is the detailed transcription process. Conversation analysts generally use the Jefferson method (Jefferson 2004) of transcription, an approach to transcription that reflects CA's commitment to attend to both *what* is said and *how* it is said. Conversation analysts approach transcription in a way in which they aim to "get as much as possible of the actual sound and sequential positioning of talk onto the page" (Atkinson and Heritage 1984, p. 12), and thus paralinguistic features are often included.

Finally, MCA specifically involves identifying the categories of interest followed by examining the surrounding text to identify the way the categories are discursively constructed. Some introductory analytical features include identifying (a) membership categorial devices that link multiple categories together in meaningful relationships (categories = mother, daughter, son; categorial device = family); (b) category-bound activities (i.e., activities assigned to particular categories); (c) category-bound predicates (i.e., descriptors and characteristics assigned to categories); and (d) standardized relational pairs (i.e., a stated or implied relationship between categories denoting obligations and duties, that is, teacher and student) (Stokoe 2012).

Review of the Policy Literature Employing CA and/or MCA In our review of the literature, we identified nine studies that have used CA or MCA to address questions relevant to educational policy. Four of these studies addressed some element of language policies in bilingual or multilingual classrooms (e.g., immersion or immigrant reception) (Bonacina-Pugh 2012; Dooly and Unamuno 2009; Musk 2006; Mori 2004). The others focused on teacher categorization and global education reform (Paulsen 2017), policies designed to engage parents in education (Thomas et al. 2015), discipline policies and student reputation (Maclure et al. 2012), teacher engagement strategies (Houen et al. 2016), and student self-evaluation and fabrications (Tanner and Prieto 2014). A key aspect of these studies is that the authors argued that policies are enacted in interac-

tion or that policies crystallize categorizations common in the culture. The data included traditional CA/MCA type data, such as in-class interactions, while also including somewhat less traditional sources, such as interviews, focus groups, policy documents, and a classroom artifact.

Key Resource for Those New to CA We recommend using Paul ten Have's *Doing conversation analysis: A practical guide* for advancing your study of CA. This resource was frequently cited among the studies in our literature review and is commonly used as a course textbook in CA courses. It covers a breadth of topics, such as CA's genesis and theoretical background; more pragmatic issues, such as collecting, creating, and analyzing data sets; and examples of effective CA studies. The text also provides exercises to begin developing your capacity as a conversation analyst and serves as a reference for key studies and theoretical pieces, as each chapter includes a list of references to explore the subject matter further.

APPLICATIONS TO THE STUDY OF POLICY AND LEADERSHIP

In this section, we provide three detailed examples of exemplar studies that have employed the methodologies discussed in the sections above.

Example of a CDA Study In their study, "Exposing ideology within university policies: A critical discourse analysis of faculty hiring, promotion and remuneration practices", Uzuner-Smith and Englander (2015) offered a useful example of how CDA can interrogate the taken-for-grantedness of educational policy. The authors began by establishing neoliberalism as reimagining the university as a key economic entity given its manufacturing role in the knowledge economy. This ideological turn has produced discourses focused on strategic plans, targets, and competition. These discourses then are operationalized in the corporatization of university faculty, as "targets, indicators, and evaluations have been put in place to determine the value of faculty" (p. 63). This study examined the extent to which these ideologies and discourses have penetrated hiring, promotion, and remuneration policies in Turkish and Mexican universities. To accomplish this, they used public documents, including a description of a university's point system for faculty human resource decisions and a manual for a university's merit pay program.

The analysis begins by explicitly documenting the key aspects of the hiring and promotion process (e.g., the points toward hiring or promotion allocated for each kind of scholarly activity). Thus, readers are able to see the policy in the value-free, amoral context traditionally attributed to policy. Having established this, Uzuner-Smith and Englander followed Fairclough (2012) by clearly defining the social wrong involved with this policy: such a rewards policy controls the activities of faculty and turns faculty into “pieceworkers” paid by the piece (p. 71). The authors conducted CDA on both the structure of the point system, critiquing the implicit ideologies on how points are allocated, and the textual aspect of the policy, identifying the discourses making up the policy. The textual analysis focused on linguistic descriptors associated with teachers: use of present tense, inclusive versus exclusive pronouns, modal verbs, passivization, and nominalization. These tools create a sense of matter-of-factness of such a policy without alternatives. Passivization and nominalization help in this construction as it eliminates any actor doing or creating this policy; it simply is and has been. It further establishes the powerlessness of the faculty to act outside certain prescribed activities. Thus, the neoliberal discourse and construct of the knowledge-based economy is readily present in the policy documents’ establishment of the faculty as widget makers pushing the economy forward.

Uzuner-Smith and Englander then followed Fairclough’s (2012) framework by seeking to empower the oppressed party to address the social wrong. They argued that because every element of the networked system (e.g., universities, education ministries, national governments) are all operating under the same ideology, opposing the policy will appear “nonsensical” (p. 79). Given the neoliberal commonsense of global economic competition, the system demands productive university faculty to aid the nation’s competitive position. Thus, the authors recommended taking steps toward increasing faculty autonomy within this kind of performance-based system. One suggestion was allowing faculty to reassign points within this framework to reflect discipline priorities. Another suggestion was to include qualitative aspects to the point system to more holistically account for the faculty work.

Example of a DP Study In his study, “The discursive construction of superintendent statesmanship on Twitter”, Hurst (2017) studied how school superintendents “interact with and discuss policy and macro-political issues on Twitter” (p. 3). Notably, this work is positioned at the

intersection of education policy and leadership, with the findings pointing to insights for policy and leadership practice. He noted that recent scholarship has highlighted how school leaders are using social media, like Twitter, to display particular aspects of their identities as leaders. As Hurst noted, platforms such as Twitter can be particularly challenging for superintendents to navigate, as in these spaces they often make visible aspects of their work which are “politically charged” (p. 3). Yet, he noted that scholars have pointed to how engaging in social media is not necessarily simply optional for superintendents, as it is becoming an increasingly common aspect of engaging with the public.

Drawing upon DP, Hurst initially included 16,658 tweets, ultimately analyzing only 1619 tweets, which were produced by superintendents (drawn initially from a database of 570 superintendents). He found that there were two broad categories of tweets within the database: those that were retweets and those that were crafted by the individual. He focused his analysis on those tweets which were constructed by one of the superintendents.

Hurst reported on two themes generated from his analysis: representation of engagement and activism. The first theme, representation of engagement, focused on those tweets in which superintendents actively engaged in conversation around presumably complex policy issues and topics. Hurst noted how tweets were constructed in a way that their relationship with ‘powerful others’ or policy actors was made explicit, notably positioning the superintendent in “the same room/group as the powerful” (p. 13). The second theme, engagement in activism, highlighted how superintendents constructed tweets that pointed to their advocacy for particular policy choices. Hurst noted that historically superintendents have often been positioned as politically neutral or as needing to be cautious in revealing their political perspective. Yet, Hurst illustrated how in his data set superintendents crafted tweets and deployed particular discursive strategies to make their positions explicit.

Hurst concluded that superintendents use Twitter to craft “images of themselves as political insiders and advocates” (p. 23), at times even fighting for particular policy choices at the expense of others. Further, he suggested that through Twitter superintendents are taking up their role of “statesman” in an online context, displaying a strategic image of self and commitments.

Example of a CA Study Bonacina-Pugh's (2012) study, "Researching 'practiced language policies': Insights from conversation analysis", demonstrated the usefulness of CA in policy analysis. Her study focused on language policy as enacted in French induction classes for newly arrived immigrants. Language policy analysis has traditionally been confined to examining the policy as text or conducting CDA type studies on the discourses at play in the policy documents. Bonacina-Pugh suggested a broader perspective that "emphasizes the fact that a language policy can be interactionally constructed in practice" (p. 217). Thus, interactional sequences can be identified that illustrate how language policy is actually constructed in everyday interactions. Conceptualizing language policy in terms of everyday conversation, regularities, and patterns brings to the fore the need for CA to uncover the implicit rules of practiced policy.

Bonacina-Pugh suggested the need for both sequential (CA) and categorizational analysis (MCA) to understand practiced language policy. Because she assumed language practices are regular, that is in "order at every point" (Sacks 1984, p. 22), sequential analysis is critical to discovering the underlying rationale. Bonacina-Pugh indicated that categorizational work in conversation is important because "social actors organize their social world into 'categories', to which they associate a set of activities" (p. 219).

French education policy establishes induction classes set aside specifically for newly arrived immigrants. This creates a multilingual environment; for example, the class used in this study had 12 students with eight different native languages represented, although with some commonly shared first and second languages. The official language policy for French education is monolingual French. Bonacina-Pugh used CA to demonstrate that the practiced language policy significantly differs from this. She demonstrated the subtleties of interactional rules using a number of different extracts from the data set. Doing so, she identified a key rule: students categorize others based on their preferred language and interact based on this categorization. When a preferred language is not shared, students resort to using French. This contrast with the policy as written gives policy analysts a more accurate sense of how their policy is being enacted.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

We believe it is important to be explicit about our position regarding the very process of making methodological choices. Specifically, we argue that the choices you make around methodology are inextricably informed by (1) your own positioning and perspective about knowledge and reality, (2) the theories that inform your work, (3) the discipline from which you work, and (4) the kinds of questions or inquiries you are most interested in posing, among other considerations. Indeed, if you determine to take up a language-based approach, you need to assure that there is coherence between your overarching research design, the methodological approach, and the methods (including data sources and analytic process) that you use. Further, it is important to note that within the area of DA, it is not unusual to take up a synthetic approach (Jørgensen and Phillips 2002), wherein varying approaches and discourse theories are drawn upon within the context of one study. Some scholars have indeed argued for taking up a synthetic approach—one wherein you “remain methodological bricoleurs and refrain from developing an all-purpose technique for discourse analysis” as a “totalizing master methodology would serve only to repress new and alternative forms of analysis” (Torfing 1999, p. 292). While in some of our own work we have taken up more synthetic approaches (see, Lester and Paulus 2014, for an example), we see great value in also working within a single methodological perspective. Moreover, starting with a single methodological perspective is often a useful beginning point for scholars new to language-based approaches.

Practical Considerations Related to CDA As you think about whether CDA might be a useful approach in your research, here are some areas to consider.

- CDA research will necessarily invoke a critical research perspective. This suggests that your study should be less concerned with objectivity (as traditionally defined) and more concerned with addressing social wrongs, while empowering marginalized populations.
- Studying how scholars like Foucault and Bourdieu conceptualize power may be helpful in developing a critical approach.
- CDA requires an ability to link macro-level discourses and ideologies to micro-level linguistic data. Thus, a researcher should understand influential, discursive trends in national and global affairs (e.g., neo-liberalism or nationalism).

- Our review of the CDA literature in educational policy and educational leadership indicated that many researchers stop at the macro-level and do not engage in the micro-level linguistic analysis. There are certainly exceptions, however (see, for instance, Souto-Manning 2014). As described and conducted by leading CDA researchers, this methodology is bound to intricate analyses of lexical and grammatical choices in policy texts. Thus, you should be prepared to move between macro- and micro-level analyses, noting that some scholars attend to meso-level discourses and practices as well (see Anderson 2015, for an example).
- If you take up a CDA approach, it will be important to spend time gaining experience in linguistic analysis, which may require attending courses or workshops and/or (at a minimum) engaging in additional reading/study.

Practical Considerations Related to DP

- DP emerged from social psychology and thus it can be difficult to make sense of the potential connection between DP and education. Yet, there are many possibilities. And, thus, an important starting point is to familiarize yourself with the history of DP and then move to consider the research in education writ large that has drawn upon DP.
- As noted above, contemporary DP has been greatly influenced by CA and therefore it is important to become intimately familiar with CA prior to carrying out a DP study.
- Like all of the approaches we have discussed in this chapter, DP brings with it a particular conception of discourse. It is critical to spend time familiarizing yourself with how DP scholars define and theorize notions of discourse prior to engaging in analysis, as such assumptions shape how analysis proceeds.
- It is common within DP to participate in group analysis sessions. Thus, it is helpful to join data analysis groups to learn more about how to *do* DP and to also acquire feedback on your own data.
- It is useful to participate in courses, workshops, or conferences that highlight the theoretical basis and analytic process of DP. While the majority of such gatherings take place in the European context, some US-based universities offer courses (e.g., Indiana University's School of Education) focused specifically on DP.

Practical Considerations Related to CA and/or MCA As you consider whether CA and/or MCA might be a fruitful approach in your research, here are some areas to consider.

- CA/MCA research are typically predicated on the use of naturally occurring, interactional data. Thus educational policy or leadership research needs to consider whether the subject of the research would occur in such settings, like Bonacina-Pugh's above "practiced language policy" (2012, p. 213).
- CA and MCA research focus on identifying the "machinery" that brings "order in every point" to everyday interactions (Sacks 1984, p. 22). These patterns help researchers understand the limits or constraints on how individuals understand categories or act in certain situations. Researchers thus need to consider whether their policy or leadership interest would benefit from discerning such patterns.
- CA or MCA research requires detailed transcription of interactions. This is an intricate analytical process where researchers decide what details are attended to in the transcribed version of the data. Researchers committed to learning this methodological perspective should begin studying Gail Jefferson's transcription system (Jefferson 2004).
- Generally, CA requires extensive training and thus if you take up a CA approach we strongly recommend that you participate in CA-related coursework and/or workshops (e.g., Rutgers' School of Communication holds regular CA workshops). Further, it is often beneficial to work closely with scholars who have been trained in CA in order to develop your skills further.
- It is critical to consider how you will assure quality in your DA or CA study. A notable article has been written about common "pitfalls" when engaging in discourse analytic work, which has some useful application to CA research as well (see Antaki et al. 2003). Importantly, this article also points to, at least implicitly, some of the critiques that may be made of discourse analytic work that does not consider these common "pitfalls".

CHAPTER SUMMARY

In this chapter, we introduced readers to language-based methodologies. We focused our discussion on CDA, DP, and CA. We noted that CDA has been most often used in policy scholarship in comparison to other

language-based methodologies. We argued, however, that other approaches also provide important analytic possibilities. Notably, in our discussion, there is much we did not say and indeed many discourse analytic approaches were left out. For instance, of importance to policy scholars, political DA (van Dijk 1997) is an approach that may be particularly relevant. Thus, we caution the reader in assuming that what we have shared in this chapter is comprehensive and rather position our contribution as situated and partial.

Recommended Readings

Jørgensen, M. W., & Phillips, L. J. (2002). *Discourse analysis as theory and method*. London: SAGE.

Jørgensen and Phillips' text is a fairly accessible general overview to DA broadly. Further, it provides brief overviews of Laclau and Mouffe's discourse theory, CDA, and DP. One of the key features of this book is its focus on introducing the core theoretical positions within DA writ large, as well as the individual approaches discussed. As such, this book serves as a useful starting point for those interested in learning about some of the key features of DA.

Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction*. London: SAGE.

Wooffitt's text provides a general overview to both DA and CA. In contrast to Jørgensen and Phillips, this text has a greater focus on CA and thus offers both critical insights related to DA generally and CA more particularly. Further, Wooffitt points to key insights related to the history of DA, as well as foundational features of DP, CDA, and CA.

REFERENCES

- Abrahamsen, H., & Aas, M. (2016). School leadership for the future: Heroic or distributed? Translating international discourses in Norwegian policy documents. *Journal of Educational Administration and History*, 48(1), 68–88.
- Anderson, K. T. (2015). The discursive construction of lower-tracked students: Ideologies of meritocracy and the politics of education. *Education Policy Analysis Archives*, 23(110). <https://doi.org/10.14507/epaa.v23.2141>.
- Antaki, C., Billing, M. G., Edwards, D., & Potter, J. A. (2003). Discourse analysis means doing analysis: A critique of six analytic shortcomings. *Discourse Analysis Online*, 1. <https://extra.shu.ac.uk/daol/articles/open/2002/002/antaki2002002-paper.html>

- Atkinson, J. M., & Heritage, J. (1984). *Structures of social action*. Cambridge: Cambridge University Press.
- Barrett, B., & Bound, A. M. (2015). A critical discourse analysis of no promo homo policies in US schools. *Educational Studies*, 51(4), 267–283.
- Bazzul, J. (2014). Critical discourse analysis and science education texts: Employing Foucauldian notions of discourse and subjectivity. *Review of Education, Pedagogy, and Cultural Studies*, 36(5), 422–437.
- Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. New York: Garden City.
- Bonacina-Pugh, F. (2012). Researching ‘practiced language policies’: Insights from conversation analysis. *Language Policy*, 11(3), 213–234.
- Burman, E., Greenstein, A., Bragg, J., Hanley, T., Kalambouka, A., Lupton, R., & Winter, L. (2017). Subjects of, or subject to, policy reform? A Foucauldian discourse analysis of regulation and resistance in UK narratives of educational impacts of welfare cuts: The case of the ‘bedroom tax’. *Education Policy Analysis Archives*, 25(26). <https://doi.org/10.14507/epaa.25.2320>.
- Dooly, M., & Unamuno, V. (2009). Multiple languages in one society: Categorisations of language and social cohesion in policy and practice. *Journal of Education Policy*, 24(3), 217–236.
- Edwards, D., & Potter, J. (1992). *Discursive psychology*. London, UK: Sage.
- Fairclough, N. (1992). Intertextuality in critical discourse analysis. *Linguistics and Education*, 4(3–4), 269–293.
- Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. London: Routledge.
- Fairclough, N. (2011). Semiotic aspects of social transformation and learning. In R. Rogers (Ed.), *An introduction to critical discourse analysis in education* (pp. 147–155). New York: Routledge.
- Fairclough, N. (2012). Critical discourse analysis. In J. P. Gee & M. Handford (Eds.), *The Routledge handbook of discourse analysis* (pp. 9–20). New York: Routledge.
- Fairclough, N., Mulderrig, J., & Wodak, R. (2011). Critical discourse analysis. In T. Van Dijk (Ed.), *Discourse studies: A multidisciplinary introduction* (pp. 357–378). London: Sage.
- Foucault, M. (1990). *The history of sexuality—Volume 1: An introduction*. New York: Vintage.
- Gabriel, R. (2017). Constructing teacher effectiveness in policymaking conversations. In J. N. Lester, C. R. Lochmiller, & R. Gabriel (Eds.), *Discursive perspectives on education policy and implementation* (pp. 219–240). New York: Palgrave Macmillan.
- Gabriel, R., & Lester, J. N. (2013). Sentinels guarding the grail: Value-added measurement and the quest for education reform. *Educational Policy Analysis Archives*, 20(9). Retrieved from: <http://epaa.asu.edu/ojs/article/view/1165>

- Gabriel, R., & Paulus, T. (2015). Committees and controversy consultants in the construction of education policy. *Educational Policy*, 29(7), 984–1011. <https://doi.org/10.1177/0895904814531650>.
- Giles, D., Stommel, W., Paulus, T., Lester, J., & Reed, D. (2015). The micro-analysis online data: Methodological developments. *Discourse, Context & Media*, 7, 45–51.
- Houen, S., Danby, S., Farrell, A., & Thorpe, K. (2016). ‘I wonder what you know...’ teachers designing requests for factual information. *Teaching and Teacher Education*, 59, 68–78.
- Huber, L. P. (2011). Discourses of racist nativism in California public education: English dominance as racist nativist microaggressions. *Educational Studies*, 47(4), 379–401.
- Hurst, T. M. (2017). The discursive construction of superintendent statesmanship on Twitter. *Education Policy Analysis Archives*, 25(29).
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Cambridge: Polity Press.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation*. Philadelphia: John Benjamins Publishing Company.
- Jørgensen, M. W., & Phillips, L. J. (2002). *Discourse analysis as theory and method*. London: Sage.
- Kennedy-Lewis, B. L. (2014). Using critical policy analysis to examine competing discourses in zero tolerance legislation: Do we really want to leave no child behind? *Journal of Education Policy*, 29(2), 165–194.
- Lester, J. N. (2014). Discursive psychology: Methodology and application. *Qualitative Psychology*, 1(2), 141–143.
- Lester, J. N., & Paulus, T. M. (2014). “That teacher takes everything badly”: Discursively reframing non-normative behaviors in therapy sessions. *International Journal of Qualitative Studies in Education*, 27(5), 641–666.
- Lester, J. N., Lochmiller, C. R., & Gabriel, R. (2016). Locating and applying critical discourse analysis within education policy: An introduction. *Education Policy Analysis Archives*, 24(102). <https://doi.org/10.14507/epaa.24.2768>.
- Lester, J. N., White, F. A., & Lochmiller, C. R. (2017a). Language-based approaches to the study of education policy. In J. N. Lester, C. R. Lochmiller, & R. Gabriel (Eds.), *Discursive perspectives on education policy and implementation*. New York: Palgrave Macmillan.
- Lester, J. N., Lochmiller, C. R., & Gabriel, R. (2017b). Exploring the intersection of education policy and discourse analysis: An introduction. *Education Policy Analysis Archives*, 25(25). <https://doi.org/10.14507/epaa.24.2971>.
- Lester, J. N., Lochmiller, C. R., & Gabriel, R. (Eds.). (2017c). *Discursive perspectives on education policy and implementation*. New York: Palgrave Macmillan.

- MacLure, M., Jones, L., Holmes, R., & MacRae, C. (2012). Becoming a problem: Behaviour and reputation in the early years classroom. *British Educational Research Journal*, 38(3), 447–471.
- McCabe, R. (2006). Conversation analysis. In M. Slade & S. Priebe (Eds.), *Choosing methods in mental health research: Mental health research from theory to practice* (pp. 24–46). Hove: Routledge.
- Mori, J. (2004). Negotiating sequential boundaries and learning opportunities: A case from a Japanese language classroom. *The Modern Language Journal*, 88(4), 536–550.
- Musk, N. J. (2006). *Performing bilingualism in Wales with the spotlight on Welsh: A study of language policy and the language practices of young people in bilingual education*. Doctoral dissertation, Linköping University Electronic Press.
- Nudzor, H. P. (2013). Unearthing the discursive shift in the ‘fCUBE’ policy implementation in Ghana: Using critical discourse analysis. *International Journal of Research & Method in Education*, 36(2), 179–201.
- Paulsen, J. (2017). Membership categorization analysis in education policy research. In J. Lester, C. Lochmiller, & R. Gabriel (Eds.), *Discursive perspectives on education policy and implementation*. New York: Palgrave Macmillan.
- Piazza, P. (2014). The media got it wrong! A critical discourse analysis of changes to the educational policy making arena. *Education Policy Analysis Archives*, 22(36). <https://doi.org/10.14507/epaa.v22n36.2014>.
- Potter, J. (2004). Discourse analysis as a way of analysing naturally occurring talk. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (2nd ed., pp. 200–221). London: Sage.
- Potter, J. (2012). Discourse analysis and discursive psychology. In H. Cooper (Ed.), *APA handbook of research methods in psychology: Vol. 2 quantitative, qualitative, neuropsychological, and biological* (pp. 111–130). Washington, DC: American Psychological Association Press.
- Potter, J., & Hepburn, A. (2008). Discursive constructionism. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (pp. 275–293). New York: Guildford.
- Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*. London: Sage.
- Rogers, R. (Ed.). (2011). *An introduction to critical discourse analysis in education*. New York: Routledge.
- Sacks, H. (1984). Notes on methodology. In J. Maxwell Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 21–27). Cambridge, UK: Cambridge University Press.
- Sacks, H. (1992). *Lectures on conversation*. Oxford: Blackwell.
- Souto-Manning, M. (2014). Critical narrative analysis: The interplay of critical discourse and narrative analyses. *International Journal of Qualitative Studies in Education*, 27(2), 159–180.

- Stokoe, E. (2012). Moving forward with membership categorization analysis: Methods for systematic analysis. *Discourse Studies*, 14(3), 277–303.
- Tanner, M., & Prieto, H. P. (2014). In between self-knowledge and school demands: Policy enacted in the Swedish middle-year classroom. *Discourse: Studies in the Cultural Politics of Education*, 35(4), 554–569.
- ten Have, P. (2007). *Doing conversation analysis*. London: Sage.
- Thomas, S., Keogh, J., & Hay, S. (2015). Discourses of the good parent in attributing school success. *Discourse: Studies in the Cultural Politics of Education*, 36(3), 452–463.
- Tileagă, C., & Stokoe, E. (2015). *Discursive psychology: Classic and contemporary issues*. New York: Routledge.
- Torfin, J. (1999). *New theories of discourses: Laclau, Mouffe and Zizek*. Oxford, UK: Blackwell Publishers Ltd.
- Uzuner-Smith, S., & Englander, K. (2015). Exposing ideology within university policies: A critical discourse analysis of faculty hiring, promotion and remuneration practices. *Journal of Education Policy*, 30(1), 62–85.
- van Dijk, T. A. (1997). What is political discourse analysis. *Belgian Journal of Linguistics*, 11(1), 11–52.
- van Dijk, T. A. (2001). Critical discourse analysis. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 349–371). Malden: Blackwell Publishers.
- van Dijk, T. (2015). Critical discourse analysis. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis*. Oxford, UK: Blackwell Publishers Ltd.
- Wiggins, S. (2017). *Discursive psychology: Theory, methods, and application*. London: Sage.
- Wilinski, B. (2017). Knowing and interpreting prekindergarten policy: A Bakhtinian analysis. *Education Policy Analysis Archives*, 25(27). <https://doi.org/10.14507/epaa.25.2211>.
- Wodak, R. (2001). The discourse-historical approach. In R. Wodak & M. Myers (Eds.), *Methods of critical discourse analysis* (pp. 63–94). London: Sage.
- Wood, L. A., & Kroger, R. O. (2000). *Doing discourse analysis: Methods for studying action in talk and text*. Thousand Oaks: Sage.



Doing Critical Policy Analysis in Education Research: An Emerging Paradigm

Michelle D. Young and Sarah Diem

Critical policy analysis (CPA) encompasses a number of different perspectives and developments that aim to critique and offer alternative strategies for examining educational policy issues. Policy is viewed as something to be critiqued or troubled rather than accepted at face value (Bacchi 2012). “Critical policy researchers engage in critique, interrogate the policy process, and the epistemological roots of policy work, examine the players involved in the policy process, reveal policy constructions,” and consider how policies and the problems they address might appear if reframed from a different perspective (Diem and Young 2015, p. 841).

The study of educational policy through a critical frame allows for a nuanced, holistic understanding of the complexities associated with education

M. D. Young (✉)
Educational Leadership and Foundations, University of Virginia,
Charlottesville, VA, USA
e-mail: mdy8n@virginia.edu

S. Diem
Educational Leadership and Policy Analysis, University of Missouri,
Columbia, MO, USA
e-mail: diems@missouri.edu

policy, from problem finding and framing to policy development, implementation, and evaluation. In this chapter, we discuss the practice of critical policy analysis within the field of education. We begin with a discussion of the theoretical and methodological approaches used by critical policy researchers, outlining key methodological and theoretical features, and then provide specific examples of critical policy studies. Subsequently, we provide guidance for novice scholars who are interested in engaging in CPA and identify a set of recommended readings. However, we wish to be clear: it is not our intent to prescribe how critical theoretical frameworks should be used in the design, conduct, and analysis of educational policy research. There are many possible ways to work with theory, and in this chapter, we articulate a number of principles, practices, and possibilities drawn from the scholarship of contemporary critical policy analysts.

TRADITIONAL AND CRITICAL POLICY ANALYSIS

When policy analysis first emerged as a field of inquiry in the 1970s, it was a “thoroughly technicised” (Ball 1995, p. 259) field where policy scientists used a specific set of methods to determine the best manner in which to implement a policy decision (Fay 1975). Policy research was viewed as a logical and scientific approach for examining policy processes and impact. As part of this field, educational policy research has generally functioned within a positivist paradigm, reflecting traditional epistemological, ontological, and methodological approaches to inquiry (Young 1999). In this chapter, we refer to this approach as traditional policy analysis.

Traditional positivist approaches to educational policy analysis have long dominated the field (Lather 2001; Levinson et al. 2009; Nagel 1984; Young 1999; Young and Diem 2014). The widespread adoption of this approach to policy research resulted in a set of norms for the field regarding the appropriate “value-neutral” way to carry out educational policy research (Marshall 1997; Scheurich 1994; Stanfield 1993; Young 1999). Briefly, traditional policy analysis (TPA) approaches in education tend to include the following four key tenets:

1. TPA focuses considerable energy on planning, adoption, implementation, examination, and/or evaluating educational changes or reforms. Change or reform is typically viewed as a linear process that can be planned and managed.
2. TPA frames research behavior as goal driven where rational individuals weigh the costs, benefits, and subsequent outcomes of a given action or strategy.

3. TPA is based on the beliefs that researchers are capable of obtaining, accumulating, and understanding the knowledge necessary for identifying and deciding between policy solutions and planning for implementation and evaluation and that this information can be expressed to others.
4. TPA is based on the assumption that researchers can effectively evaluate policies, policy alternatives, and practices and then based on these evaluations are able to identify and ameliorate problems (Boyd 2000; Diem et al. 2014; Honan 2015; Scheurich 1994; Young and Diem 2017).

Although traditional approaches to educational policy analysis can vary in design and application, they generally reinforce the perspective that “empirical research can access the information needed to understand, design, plan, problem solve, and implement effective educational policies and practices” (Diem et al. 2014, p. 1071). While the above four tenets may not reflect the work of all traditional policy researchers, they do highlight elements that distinguish more traditional policy analysis and, as such, provide a helpful archetype for the purpose of comparison.

Dissatisfied with theoretically narrow and rationally driven approaches to policy analysis, a number of researchers have questioned the nature of policy, how it is created, and its impact and started using critical frameworks to problematize traditional approaches to analysis (McDonnell 2009; Young and Diem 2017). Influential critical policy scholars like Apple (1982), Ball (1991, 1993, 1994), and Popkewitz (1997, 2000) comprised the first generation of critical policy researchers (Levinson et al. 2009), critiquing the rational-technical approach of traditional educational policy research and emphasizing the role of power and ideology in the policy process. To illustrate, in his seminal text, Apple (1982) rejected the notion of a linear policy process of inputs and outputs and, instead, critiqued and theorized about how educational policies and policy institutions produce unequal power relations. In his curriculum policy work, he interrogated the dynamics associated with race, socioeconomic status, gender, and student resistance. His work demonstrated how policy can reshape the lived world of different student populations, restructure opportunity, and restrain the capacity to act (see also Forester 1993 and Pillow 1997). In his early critical policy work, Ball (1991, 1993, 1994) paid particular attention to “how issues are defined, how data are gathered, how concepts are tied to metaphors that circumscribe understanding,

and how findings are characterized” (Diem and Young 2015, p. 840). In later work, he used critical frameworks to explore shifting forms of governance, policy networks, and the diffusion of values (Ball 2008).

Our research on the roots and landscape of critical policy work reveals that scholars engaged in CPA take little at face value; rather, critique serves as a foundation for much CPA work (Diem et al. 2014). As such, CPA can assist in uncovering structures of oppression and inequality (Lugg and Murphy 2014), which is why many scholars whose research has a social justice bent are attracted to it. For example, Fernández and López (2017) utilized CPA as an analytical tool to complicate the power dynamics of Latin@ parents organizing in a Midwest urban school, Welton, Harris, Altamirano, and Williams (2017) used CPA to uncover the politics of student voice as it relates to power and school policy within a high school class focused on social justice education, and Milani and Winton (2017) analyzed Ontario, Canada’s school fundraising policy through a critical democratic lens to explore the ways in which democratic values were supported or subverted.

In addition to the above similarities, scholars whose policy work draws from critical frameworks have a number of other commonalities. Specifically, scholars who employ this approach to their work tend to incorporate one or more of the following critical practices:

1. CPA interrogates the roots and development of educational policy.
2. CPA probes the difference between policy rhetoric and practiced reality.
3. CPA examines the distribution of power, resources, and knowledge and the creation of “winners” and “losers.”
4. CPA scrutinizes the complex systems and environments in which policy is made and implemented.
5. CPA explores social stratification and the impact of policy on relationships of privilege and inequality.
6. CPA is interested in the nature of resistance to or engagement in policy by members of historically underrepresented groups (Diem et al. 2014; Young and Diem 2017).

Methodologically, there are similarities as well. Critical policy researchers, as will be discussed in further detail below, pay particular attention to how their theoretical perspectives inform their research, design, and methodological choices. CPA scholars reflect what Anfara and Mertz (2015)

refer to as the “Theory as More” perspective, wherein “theory affects every aspect of the study, from determining how to frame the purpose and problem, to deciding what to look at and for resolving how to make sense of the data collected” (p. 11). Furthermore, given the nature of their interests and the policy questions they raise, it may be unsurprising to learn that critical policy scholars are more likely to use qualitative research approaches (Young and Diem 2017). These commonalities, particularly those pertaining to methodology, will be taken up in greater depth in the following section.

REVIEW OF RELEVANT LITERATURE CONCERNING THE THEORY AND METHOD OF CPA

One of the key distinguishing features of CPA is the relationship between theory and method. In CPA, methodology and theoretical perspectives work hand in glove (Diem et al. 2014; Young 1999). CPA involves a lens for looking as well as a way of looking. You cannot have one without the other. A comparable example is critical discourse analysis (CDA). CDA includes both theory, ranging from macro- and micro-sociological theories to linguistic theories, as well as a method. It is a research approach that bundles theory and method to interpret the social phenomenon under analysis (Meyer 2001). Fairclough (2001) describes CDA as being “in a dialogical relationship with” social theories and methods (p. 121).

To be clear, in CPA theory and method are “inextricable,” but they are not equal (Young and Diem 2017, p. 5). That is, theory leads method. Young (1999), like many other CPA scholars, subscribes to the understanding that the “research frame one uses dictates, to a large extent, the way one identifies and describes policy problems, the way one researches these problems, the policy options one considers, the approach one takes to policy implementation, and the approach taken for policy evaluation” (p. 681). In the following two subsections, we take up the role that theory and method often play in critical policy analysis.

Theoretical Considerations for CPA Theory plays a significant role within critical policy analysis, and CPA scholars have drawn on a variety of critical perspectives and methods in their exploration of policies, policy contexts, policy processes, policy communities, and policy impact (Young and Diem 2014). Some of this work has been informed by poststructural

frameworks, critical theories, feminist theories, queer theories, and critical race perspectives, among others, drawn from multiple disciplines across the humanities and social sciences. According to those who engage in CPA, the ability to analyze educational policies through one or more critical theoretical framework, as opposed to a rational, traditional approach, results in policy analyses that have more depth and breadth (Diem et al. 2014; Ulmer 2016).

One's choice of theory is greatly influenced by the paradigm or interpretive framework from which one works, as paradigmatic frameworks they provide a "basic set of beliefs that guides action" (Guba 1990, p. 17). They encompass a researcher's epistemological, methodological, and ontological premises, which reflect three different aspects of knowledge or knowing (Anfara and Mertz 2015; Denzin and Lincoln 1994). Epistemology is concerned with how we know the world around us, methodology is concerned with how we take in knowledge about the world, and ontology is concerned with the nature of reality itself (Denzin and Lincoln 1994). In the field of education, positivist and postpositivist paradigms, which undergird TPA, "provide the backdrop against which other paradigms and perspectives operate" (Denzin and Lincoln 1994, p. 99).

Though not deterministic, CPA scholars understand that the theory they employ to examine a given research problem has significant implications for the way they will think about the problem; they understand how theory influences the questions that they ask about the problem, their research design, and the methods they choose to use in their investigation of those questions (Diem et al. 2014; Diem and Young 2015; Morrow and Brown 1994; Young 1999). "Theory impacts the identification of the research topic or problem, it impacts the way the researcher thinks about the problem, and it impacts the questions that s/he asks about the issue" (Diem and Young 2015, p. 844). As Gulson, Clarke, and Petersen (2015) write, "theory is a provocation in education policy studies, engendering new ways of conceiving and doing" policy analysis (p. 7). We discuss the methodological considerations of CPA below.

Methodological Considerations for CPA CPA is not possible through mechanistic, atheoretical coding, and reducing data to themes. CPA requires the application of a critical theoretical framework in the treatment of data (i.e., collection, analysis, interpretation, etc). However, there is no "one best way" of doing CPA. CPA scholars resemble what Denzin and

Lincoln (1994) described in their first handbook of qualitative research as “bricoleurs.” Bricoleurs are well versed in a variety of research designs, have facility with multiple data collection strategies, and are “knowledgeable about many interpretive paradigms...that can be brought to any particular problem” (p. 2). As CPA scholars bring their paradigmatic and theoretical frameworks to bear on policy issues, like other researchers, they make research design decisions that connect the study’s focal issue to their development of research questions and the identification of data collection methods and analytic procedures (Young and Reynolds 2017).

Given the nature of their policy questions and perspectives, as introduced earlier in this chapter (e.g., interrogating the roots and development of educational policy, exploring social stratification and the impact of policy on relationships of privilege and inequality, etc.), CPA scholars are more likely to use qualitative research approaches, which provide an opportunity for deeper engagement with their research subject (e.g., persons, policies, discourses) (deLeon and Vogenback 2007; Denzin and Lincoln 2005; Morrow and Brown 1994; Young and Diem 2017). In this chapter, we limit our discussion to qualitative approaches to CPA.

Design issues, including the framing of research questions, the location of data sources, and decisions about timelines and research strategies, are shaped by a researcher’s critical framework, and, as a result, research designs will vary from one CPA to another. For example, O’Malley and Long (2017) use Queer Theory to “understand and theorize the processes enabling the adoption of a domestic partnership benefits policy, inclusive of same sex couples, in a fairly conservative Texas school district and, more significantly, to make visible and problematize the socio-political constructions that frequently render the design of queer inclusive educational policy for K-12 schools a Sisyphean task” (O’Malley and Long 2017, p. 66). Thus, their research design was informed both by the tenets of Queer Theory as well as their interest in interpreting and deconstructing a policy event (the adoption of a policy) as well as the context, participants, meaning making, and events leading up to the policy adoption.

Informed by a principle of CPA that recognizes the importance of historically and politically contextualizing policy to realize how it is reflective of society (Eppley 2009), O’Malley and Long (2017) conducted a critical content and discourse analysis of 74 print media articles over an eight-year period that captured the development and state-wide struggles associated with the policy. Through their analysis, O’Malley and Long illustrate the

value of a Queer Theory Critical Policy Analysis “as an intellectual tool for problematizing and interrupting normalizing assumptions inscribed in specific educational policies that have the material effect of fostering inequity across multiple manifestations of difference” (p. 78). Paired with discourse analysis, Queer Theory enabled these researchers to do more than read “about” a given policy event; it facilitated the dismantling of the event, the exploration of preconceived notions, the exposure of absences and, among other things, the troubling of loose ends.

If this same policy were explored by another critical policy analyst, using yet a different theoretical perspective, the methodological approach would also look different. For example, if one were using assemblage theory, based on Deleuze and Guattari’s idea of assemblage, one might use an assemblage-informed ethnographic approach to situate the policy process within wider forces of assemblage and disassemblage (Youdell 2015). In contrast, if one were interested in the influence of certain values and beliefs in the process of adopting the partner benefits policy, one might use critical discourse analysis to investigate the inscription of particular policy vocabularies within the policy process.

In addition to critical discourse analysis, CPA scholars draw on the methodological techniques of ethnomethodology, phenomenology, deconstruction, historical and microhistorical designs, cultural studies, hermeneutics, and ethnography. Within these traditions, they utilize a variety of data collection and analytic strategies, including observations, interviews, key informant testimonies, mass media analysis, document analysis, textual and discourse analysis, diffraction, visual methods, and examination of statistical databases and literature reviews (Brewer 2014; Diem 2017; Young and Diem 2017; Ulmer 2016; Young and Reynolds 2017). As Derrida notes, “The idea is not to jettison the classical discipline, but to disturb it by way of exploring what systematically drops through its grid and, by so disturbing it, to open it up” (Caputo 1997, p. 77).

The distinguishing methodological features of CPA include and are linked to the five critical practices delineated earlier in this chapter. For example, in interrogating the roots and development of policy as well as the difference between policy rhetoric and practiced reality, CPA scholars engage in deconstructive processes such as those used by O’Malley and Long (2017) as well as what Morrow and Brown (1994) refer to as “intensive explication” (p. 212). Intensive explication involves questioning the taken-for-granted, interrogating policy constructions, searching for epistemological roots, and identifying and explaining deep patterns by

“empirically lifting into view the underlying semantic, sociocultural, and structural relations that are constitutive of historically unique actors, mediations, and systems” (Morrow and Brown 1994, p. 212).

Similarly, when investigating the distribution of power, resources, and knowledge and the impact of policy on relationships of privilege and inequality, CPA scholars engage in “concentrated looking,” which involves the collection and examination of “contextualizing information, policy texts, observations and interviews” (Diem and Young 2015, p. 845). Through concentrated looking, CPA scholars “pay significant attention to the complex systems and environments in which policy is made and implemented” (Young and Diem 2017, p. 4). They emphasize “the importance of providing a contextualized understanding of their research findings, reflecting the complexity of the policies, people, schools, and communities they impact and tended to take time to provide the historical and or cultural context of the policy issue under examination” (Diem and Young 2015, p. 844). Here, researchers elicit meaning from what is not said or done, what some have referred to as the blank spots, as well as how messages are conveyed, by whom and in what contexts (Young 2003). Concentrated looking is reflected in approaches such as historical reconstruction, deconstruction, ethnographic interpretation, and theory-based analysis.

Finally, CPA scholars engage in self-reflexive practices and discursive reading and rereading of data, the intent of which is to interrogate not only the data but their own sense making of that data (Young and Reynolds 2017). Strongly related to heightened self-reflexivity is the tendency of CPA scholars to be more transparent about their methodological decisions (Young and Reynolds 2017). While many qualitative scholars are less explicit about “how they arrived at the themes they report as their findings,” leaving readers guessing (Brooks and Normore 2015, p. 802), CPA scholars are more likely to discuss how the theory they used to explore a given research problem impacted the way they thought about the problem, the questions they asked, and the methods used to investigate those questions (Diem et al. 2014; Diem and Young 2015; Young 1999). Furthermore, like O’Malley and Long (2017), most CPA scholars are particularly careful to point out the link between theory and method in the analytical phase, that is, how their perspectives influence the patterns, tensions, and inconsistencies they see, the judgments they make about those patterns, and the themes they choose to explicate in their research findings.

APPLICATIONS OF CPA TO THE STUDY OF EDUCATIONAL LEADERSHIP AND POLICY

Critical policy scholars have offered alternative strategies for examining a variety of educational policy issues and, as a result, have offered new perspectives on taken-for-granted policy issues and problems of leadership practice. In this section, we highlight several examples of critical policy studies that make explicit how the “research frame one uses dictates, to a large extent, the way one identifies and describes policy problems [and] the way one researches these problems” (Young 1999, p. 681).

In their article “The influence of values and policy vocabularies on understandings of leadership effectiveness,” Carpenter, Diem, and Young (2014) critically examine the inscription of policies’ vocabularies within four federal and state education reform policies associated with the evaluation of public school leaders: the Obama/Duncan *Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act*, the Obama/Duncan Race to the Top (R2T) policy documents, and the Delaware and Tennessee Race To the Top (R2T) applications. They seek to uncover how the Obama/Duncan Administration, through constructed story lines, developed concrete policy vocabularies that shaped the conceptualization of how educational leaders should be evaluated for effectiveness. Their study is based upon tenets of CPA that characterize policies as “socially constructed products shaped by historically contingent power differentials” (Carpenter et al. 2014, p. 1116; Fischer 2003; Forester 1993). It is organized as a critical discourse analysis and based on Hajer’s (2003) view of discourse, which “suggests policy analysts should examine how actors are able to secure, maintain, and reshape their discursive position in the midst of political negotiations and controversies” (Carpenter et al. 2014, p. 1116). The choice to pair critical discourse analysis with CPA provided Carpenter et al. (2014) a space to pay particular “attention to the value orientation of educational policy solutions by refusing to overlook the social construction of political discourses” (p. 1130), which is not afforded in traditional approaches to educational policy analysis.

In her study of three present-day school desegregation policies, Diem (2017) provides a nuanced historical case study analysis of the complexities behind the development and implementation of policies that use a variety of factors in assigning students to schools with the goal of achieving racial and socioeconomic diversity. She merges a critical policy analysis approach with a policy implementation framework in order to illustrate how and

why decisions were made when designing the policies, the (un)intended consequences of the policy implementation process, and how the politics surrounding student assignment policies (local, state, and federal) has an impact on their design and implementation. “The policy implementation process no longer can be viewed in terms of one group of people working to shape and influence how a policy gets implemented” (Diem 2017, p. 47). Examining the interaction among policy, people, and places is critical to understanding policy design and implementation (Honig 2006). Critical policy scholars are particularly interested in exploring and questioning the roots and development of educational policy as well as “the complex systems and environments in which policy is made and implemented” (Diem et al. 2014, p. 1073). Diem (2012, 2017) illustrates how sociopolitical and geographic contexts matter when shaping, adopting, and garnering support for student assignment policies that seek to racially and economically diversify schools. Her CPA examination of three different policies in distinct contexts sheds light on how policymakers can work to address in the ongoing and increasing levels of racial and economic segregation existent in public schools.

In nearly all of Ontario, Canada’s elementary and secondary schools, school fundraising occurs in order to support school programs and resources. Yet, Milani and Winton (2017) show the contradictory nature of such fundraising and the critical democratic commitment to equality, equity, social justice, and community as fundraising shifts the responsibility of funding education from the public to the private domain. Employing a critical democratic lens in their CPA to illustrate how Ontario’s fundraising policy is undermining the ideals of critical democracy in its public schools, Milani and Winton are particularly interested in the policy’s contexts of influence, text production, and practice. They utilize a CPA approach because “it provides the opportunity to interrogate the policy process, social structures, and power dynamics within the policy field” (Diem et al. 2014; Milani and Winton 2017, p. 194) and assists in determining whether the fundraising policy supports equity, inclusion, participatory decision-making processes, and knowledge inquiry and critical mindedness. The findings of Milani and Winton’s CPA show that if critical democracy is to be achieved in Ontario’s school system, the fundraising policy must be eliminated, and the government must adequately fund its schools.

Finally, Marshall and Young (2013) use a feminist critical policy framework to analyze recent federal and state educational policies and their impact on women. Feminist critical policy analysis, which frames the

gender/power dynamics in issue-framing, coalitions, and funding, enables the search for how thinking is mediated by historically constituted power relations, how constructed “facts” and assumptions became perpetuated as aspects of reality, and how some groups have gained and maintained privilege. Specifically, these authors use feminist critical policy analysis to track the adoption of recent educational policies (e.g., NCLB), to explore how they have been shaped by discourses and relationships of power, and to identify how these policy trends subtly work counter to the interests of women educators and leaders. Through their analysis, Marshall and Young show how certain policies, like accountability and the adoption of national curricula and standards, and policy discourses, such as policy discourses that frames public education, educators, and educational leaders as failures and poor investments, have become avenues for undermining women’s power, pay, status, and chance to speak their voices. Furthermore, Marshall and Young demonstrate how state and national emergencies, like the recession and state budget crises, have been used as reasons and tools for attacking women’s status and power.

As these examples demonstrate, education policy can significantly impact children’s lives, educator’s working conditions, the viability of communities, and the profession in general. Moreover, the methods critical policy scholars employ to explore these issues can help us better understand the structural and individual factors that work to maintain and produce our current inequitable education system. Thus, the implications for the critical analysis of educational policy, policy actors, and policy conditions can be substantial. “Non-traditional framing of policy, policy entities, and policy actors facilitates the development of questions that are rarely asked when traditional perspectives are employed” (Young and Reynolds 2017, p. 40). CPA provides insight into the elements of educational policy that are typically left unquestioned but may contain features that, if left unchecked, can undermine the very outcomes a given policy intended to support (Young and Diem 2017, p. 152).

RECOMMENDATIONS FOR NOVICE AND EMERGENT SCHOLARS

We are often asked by novice scholars where to begin when it comes to utilizing critical approaches to education policy analysis in their own research and scholarship. That question is then usually followed by a second question around how critical policy work is received in a field that operates under a neoliberal, market-driven context, and what we do as

critical policy scholars to convey the importance of critically oriented research to researchers, practitioners, and policymakers alike.

For novice scholars, we offer a number of recommendations that can assist in preparing oneself to begin researching educational policy using critical perspectives. First, it is important to not only read theory, particularly critical theories, and understand them but also to learn how they can be applied across different types of projects. The first step is to become immersed in theory and to develop a deep knowledge of continuities and discontinuities between theories that emerge from different paradigms. We both were fortunate to have faculty mentors (and later colleagues) who introduced us to a variety of theoretical frameworks and engaged us in exploring their epistemological roots and methodological implications. As graduate students, we were assigned readings reflecting the work of important thinkers like Pierre Bourdieu, Nancy Fraser, Jacques Derrida, Patricia Hill Collins, Michel Foucault, Patti Lather, Jacques Lacan, Deborah Stone, Kimberlé Crenshaw, Richard Delgado, Stephen Ball, and Judith Butler, among others, and were required to make meaning of them both on our own as well as within our scholarly classroom communities. We were also fortunate to have worked with talented qualitative and quantitative research faculty, who encouraged us to go beyond introductory methods courses and to struggle with the many ontological issues involved in educational research, including the ethics of speaking for others, epistemological racism, and challenges to validity claims. With each opportunity to engage with theory, our thinking was challenged and our understanding broadened. From our research with other CPA scholars, we have learned that our experiences reflected much of theirs.

We would highly recommend exploring a variety of frameworks from positivism to poststructuralism, spending time thinking through what the frameworks mean for your own research interests. Reading the work of classical theorists is just as important as reading postpositivist and critical theories; in fact, it is crucial as it helps us better understand how theory has evolved over time and why additional theories were needed and eventually created to address concepts that were not previously addressed (e.g., race, gender, sexuality, intersectionality, among others). There are several ways to go about this. If you are still in graduate school, then consider taking courses in theory and human inquiry, often offered outside of schools/colleges of education in departments such as sociology and anthropology. Such courses should facilitate the opportunity to explore ways of theorizing, understand how theory shapes understanding of issues, and practice

using theory to frame research interests. However, if such courses are not available to you, you can move forward on your own or form a reading group with colleagues. Over the course of a semester, you could choose to read a number of standpoint theories, including books like Patricia Hill Collins' *Black Feminist Epistemology*. As you read Collins' book, you might consider what the core themes of a Black Women's Standpoint would reveal if applied to your work, or given Collins' discussion of the development of her perspective, you might think about what the themes of your own or another standpoint or perspective might entail. The point is to think with the theories you are reading, to mull them over, and to apply them to issues of concern to you. Theory "should challenge us to think in new ways and expand our ways of knowing by employing different lenses to examine problems" (Young and Diem 2017, p. 151). As critical policy scholars, we should be interested in understanding the limits and possibilities of theories.

Once one has a strong grasp of theory and how it has been used within a variety of empirical and conceptual projects, we would recommend that novice scholars become similarly familiar with the methodological literature. Taking courses and reading books on a variety of approaches to research design, data collection, description, analysis, and interpretation will not only make you a stronger researcher, but it will aid you in avoiding what Brinkman (2012) calls "methodolatry," which involves the uncritical valorization of a particular methodological technique or research method (p. 49). Moreover, given that critical policy scholars believe that theory "is at the heart of helping us explain social phenomenon" (Young and Diem 2017, p. 151), we suggest that novice scholars become familiar with such texts as the *Handbook of Critical and Indigenous Methodologies* (Denzin et al. 2008), *Disrupting Qualitative Inquiry: Possibilities and Tensions in Educational Research* (Brown et al. 2014), *Critical Race Spatial Analysis: Mapping to Understand and Address Educational Inequity* (Morrison et al. 2017), *Critical Qualitative Inquiry: Foundations and Futures* (Cannella et al. 2015), and *Methods of Critical Discourse Studies* (Wodak and Meyer 2015).

We also believe it is important for novice scholars to be well informed in their subject matter, including how it has been researched in the past. If a firm grasp of the literature has not been established, it is difficult to discover where the holes are and where new critical approaches to inquiry are needed. This knowledge about the subject matter should include the content area, how it has been theorized, and the methods employed in the research.

Finally, we think it is important to provide a key recommendation for expert critical policy researchers that may assist novice scholars in finding a critical voice in their own research: be clear in your application of CPA. If we want to disrupt traditional approaches to education policy analysis and see critical approaches to education policy analysis become more commonplace in our field, we have to do a better job in articulating *how* our research is critical and *why* it is important for it to be critical. We have to make sure our research is accessible to a wide variety of audiences, not just those in academia, by actively participating in diverse communities (e.g., serving on boards of organizations, volunteering in community spaces) and by publishing our research in various outlets in addition to academic journals and books (e.g., policy briefs, policy reports, newspaper articles and op-eds, blogs, etc.). In a current sociopolitical context that is skeptical of higher education in general, we need to do a better job of working alongside those who create and implement policy (i.e., policymakers) and those who are directly impacted by policy (i.e., practitioners, students, families, and communities) to articulate the importance of taking a critical stance on educational policy issues.

CHAPTER SUMMARY

“Understanding policy differently might ultimately lead to better policy” (Ulmer 2016, p. 1392). In this chapter, we discussed the growing subfield of policy analysis within the field of education—a subfield called Critical Policy Analysis (CPA) that is working to understand policy differently. Although critical approaches to educational policy analysis were introduced several decades ago, the foothold of CPA is still gaining traction. To further advance CPA and the opportunity to understand policy differently, we have provided insight into the major theoretical and methodological approaches used by critical policy researchers in the United States today, outlined key methodological and theoretical features, summarized specific examples of critical policy studies, and provided guidance for scholars who are interested in using a CPA approach in their own work. “The profound shifts taking place in contemporary social life require a shift in our research traditions” (Young 1999, p. 705). CPA represents such a shift and offers the promise of broader, deeper, and potentially more complex understandings of educational policy issues.

Recommended Readings

The following texts are offered not as *the* texts to consult, but as providing an entry point for those who are new to CPA and would like to learn more about it.

Carpenter, B., Diem, S., & Young, M. D. (2014). The influence of values and policy vocabularies on understandings of leadership effectiveness. *International Journal of Qualitative Studies in Education*, 27(9), 1110–1113.

Utilizing critical discourse analysis as an analytical framework, Carpenter, Diem, and Young critically examine the inscription of neoliberal policy vocabularies within education reform policies concerning the evaluation of educational leaders. The article explores the ways in which globalized discourses shape federal reform documents regarding the evaluation of educational leaders, and how these federal discourses are woven into state-level policy documents that are relied on by local education agencies to shape the daily practices and evaluation of educational leaders. This article is useful for scholars interested in applying critical policy analysis frameworks in educational leadership studies.

Fischer, F. (2003). *Reframing public policy: Discursive politics and deliberative practices*. New York: Oxford University Press.

Fischer explores the perspectives of public policy discourse, discursive policy analysis, and participatory deliberative policymaking practices to critique the dominant, positivist/empiricist approaches to public policy studies. The book outlines the theoretical, methodological, and political dimensions of these “postempiricist” approaches, and may be useful in helping novice researchers design research that is critically-oriented.

Young, M. D., & Diem, S. (Eds.) (2017). *Critical approaches to education policy analysis: Moving beyond tradition*. Switzerland: Springer International Publishing.

Young and Diem bring together work by emerging and senior scholars in the field of educational policy who apply critical frameworks to their research. The volume offers insight into how theory and method interact and are applied in critical policy analysis. The volume provides concrete examples for engaging in critical policy analysis and is thus useful for novice researchers in conceptualizing educational policy issues from critical perspectives.

REFERENCES

- Anfara, V., & Mertz, N. (2015). *Theoretical frameworks in qualitative research*. Thousand Oaks: Sage.
- Apple, M. (1982). *Education and power*. London: Routledge.
- Bacchi, C. (2012). Why study problematizations? Making politics visible. *Open Journal of Political Science*, 2, 1–8.
- Ball, S. J. (1991). *Politics and policy making in education*. London: Routledge.
- Ball, S. J. (1993). What is policy? Texts, trajectories, and toolboxes. *Discourse*, 13(2), 10–17.
- Ball, S. J. (1994). *Education reform: A critical and post-structural approach*. Buckingham: Open University Press.
- Ball, S. J. (1995). Intellectuals or technicians? The urgent role of theory in educational studies. *British Journal of Educational Studies*, 43(3), 255–271.
- Ball, S. J. (2008). New philanthropy, new networks and new governance in education. *Political Studies*, 56, 747–765.
- Boyd, W. L. (2000). The “R’s of school reform” and the politics of reforming or replacing public schools. *Journal of Educational Change*, 1(3), 225–252.
- Brewer, C. A. (2014). Historicizing in critical policy analysis: The production of cultural histories and microhistories. *International Journal of Qualitative Studies in Education*, 27(3), 273–288.
- Brinkman, S. (2012). *Qualitative inquiry in everyday life: Working with everyday life materials*. London: SAGE.
- Brooks, J. S., & Normore, A. H. (2015). Qualitative research and educational leadership: Essential dynamics to consider when designing and conducting studies. *International Journal of Educational Management*, 29(7), 798–806.
- Brown, R. N., Carducci, R., & Kuby, C. R. (Eds.). (2014). *Disrupting qualitative inquiry: Possibilities and tensions in educational research*. New York: Peter Lang.
- Cannella, G. S., Pérez, M. S., & Pasque, P. A. (Eds.). (2015). *Critical qualitative inquiry: Foundations and futures*. New York: Taylor & Francis.
- Caputo, J. D. (Ed.). (1997). *Deconstruction in a nutshell: A conversation with Jacques Derrida*. New York: Fordham University Press.
- Carpenter, B. W., Diem, S., & Young, M. D. (2014). The influence of values and policy vocabularies on understandings of leadership effectiveness. *International Journal of Qualitative Studies in Education*, 27(9), 1110–1133.
- deLeon, P., & Vogenback, D. M. (2007). The policy sciences at a crossroads. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), *Handbook of public policy analysis: Theory, politics and methods* (pp. 3–14). Boca Raton: Taylor & Francis Group.
- Denzin, N., & Lincoln, Y. (1994). *Handbook of qualitative research*. Thousand Oaks: Sage.

- Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 1–32). Thousand Oaks: Sage.
- Denzin, N. K., Lincoln, Y. S., & Smith, L. T. (Eds.). (2008). *Handbook of critical and indigenous methodologies*. Thousand Oaks: SAGE.
- Diem, S. (2012). The relationship between policy design, context, and implementation in integration plans. *Education Policy Analysis Archives*, 20(23), 1–39.
- Diem, S. (2017). A critical policy analysis of the politics, design, and implementation of student assignment policies. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 43–62). Gewerbestraße, Switzerland: Springer International Publishing.
- Diem, S., & Young, M. D. (2015). Considering critical turns in research on educational policy. *International Journal of Educational Management*, 29(7), 838–850.
- Diem, S., Young, M. D., Welton, A. D., Mansfield, K. C., & Lee, P. (2014). The intellectual landscape of critical policy analysis. *International Journal of Qualitative Studies in Education*, 27(9), 1068–1090.
- Eppley, K. (2009). Rural schools and the highly qualified teacher provision of No Child Left Behind: A critical policy analysis. *Journal of Research in Rural Education*, 24(4), 1–11.
- Fairclough, N. (2001). Critical discourse analysis as a method in social scientific research. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 121–138). London: SAGE.
- Fay, B. (1975). *Social theory and political practice*. London: George Allen & Unwin.
- Fernández, E., & López, G. R. (2017). When parents behave badly: A critical policy analysis of parent involvement in schools. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 111–130). Gewerbestraße, Switzerland: Springer International Publishing.
- Fischer, F. (2003). *Reframing public policy: Discursive politics and deliberative practices*. New York: Oxford University Press.
- Forester, J. (1993). *Critical theory, public policy and planning practice: Toward a critical pragmatism*. Albany: SUNY Press.
- Guba, E. G. (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 17–30). Newbury Park: Sage.
- Gulson, K., Clarke, M., & Petersen, E. B. (2015). Introduction: Theory, policy, methodology. In K. Gulson, M. Clarke, & E. B. Petersen (Eds.), *Education policy and contemporary theory: Implications for research* (pp. 1–12). New York: Routledge.
- Hajer, M. A. (2003). A frame in the fields: Policymaking and the reinvention of politics. In M. A. Hajer & H. Wagenaar (Eds.), *Deliberative policy analysis: Understanding governance in the network society* (pp. 88–110). Cambridge: Cambridge University Press.

- Honan, E. (2015). Thinking rhizomatically: Using Deleuze in education policy contexts. In K. N. Gulson, M. Clarke, & E. B. Peterson (Eds.), *Education policy and contemporary theory* (pp. 208–218). New York: Routledge.
- Honig, M. (Ed.). (2006). *New directions in educational policy implementation: Confronting complexity*. New York: State University of New York Press.
- Lather, P. (2001). Validity as an incitement to discourse: Qualitative research and the crisis of legitimation. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 241–250). Washington, DC: American Educational Research Association.
- Levinson, B. A. U., Sutton, M., & Winstead, T. (2009). Education policy as a practice of power: Theoretical tools, ethnographic methods, democratic options. *Educational Policy*, 23, 767–795.
- Lugg, C. A., & Murphy, J. P. (2014). Thinking whimsically: Queering the study of educational policy-making and politics. *International Journal of Qualitative Studies in Education*, 27(9), 1183–1204.
- Marshall, C. (1997). Dismantling and reconstructing policy analysis. In C. Marshall (Ed.), *Feminist critical policy analysis: A perspective from primary and secondary schooling* (pp. 1–39). London: The Falmer Press.
- Marshall, C., & Young, M. D. (2013). Policy inroads undermining women in education. *International Journal of Leadership in Education*, 16(2), 205–219.
- McDonnell, L. M. (2009). A political science perspective in education policy analysis. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 57–70). New York: Routledge.
- Meyer, M. (2001). Between theory, method, and politics: Positioning of the approaches to CDA. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 14–31). London: SAGE.
- Milani, M., & Winton, S. (2017). Ontario's fourth 'R': A critical democratic analysis of Ontario's fund-'R'aising policy. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 193–214). Gewerbestraße, Switzerland: Springer International Publishing.
- Morrison, D., Annamma, S. A., & Jackson, D. D. (2017). *Critical race spatial analysis: Mapping to understand and address educational inequity*. Sterling: Stylus Publishing, LLC.
- Morrow, R. A., & Brown, D. D. (1994). *Critical theory and methodology*. Thousand Oaks: Sage Publications.
- Nagel, S. S. (1984). *Contemporary public policy analysis*. Birmingham: The University of Alabama Press.
- O'Malley, M. P., & Long, T. A. (2017). Public educational policy as performance: A queer analysis. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 63–82). Gewerbestraße, Switzerland: Springer International Publishing.

- Pillow, W. (1997). Decentering silences/troubling irony: Teen pregnancy's challenge to policy analysis. In C. Marshall (Ed.), *Feminist critical policy analysis: A perspective from primary and secondary schooling* (pp. 134–152). London: Falmer Press.
- Popkewitz, T. S. (1997). A changing terrain of knowledge and power: A social epistemology of educational research. *Educational Researcher*, 26(9), 18–29.
- Popkewitz, T. S. (Ed.). (2000). *Educational knowledge: Changing relationships between the state, civil society, and the educational community*. Albany: SUNY Press.
- Scheurich, J. J. (1994). Policy archaeology: A new policy studies methodology. *Journal of Education Policy*, 9, 297–316.
- Stanfield, J. (1993). Epistemological considerations. In J. Stanfield & R. Dennis (Eds.), *Race and ethnicity in research methods* (pp. 16–36). Newbury Park: Sage.
- Ulmer, J. (2016). Diffraction as a method of critical policy analysis. *Educational Philosophy and Theory*, 48(13), 1381–1394.
- Welton, A. D., Harris, T. O., Altamirano, K., & Williams, T. (2017). The politics of student voice: Conceptualizing a model for critical analysis. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 83–110). Gewerbestraße, Switzerland: Springer International Publishing.
- Wodak, R., & Meyer, M. (Eds.). (2015). *Methods of critical discourse studies* (3rd ed.). London: Sage.
- Youdell, D. (2015). Assemblage theory and education policy sociology. In K. N. Gulson, M. Clarke, & E. B. Peterson (Eds.), *Education policy and contemporary theory* (pp. 110–121). New York: Routledge.
- Young, M. D. (1999). Multifocal educational policy research: Toward a method for enhancing traditional educational policy studies. *American Educational Research Journal*, 36, 677–714.
- Young, M. D. (2003). The leadership crisis: Gender and the shortage of school administrators. In M. D. Young & L. Skrla (Eds.), *Reconsidering feminist research in educational leadership* (pp. 265–298). Albany: SUNY Press.
- Young, M. D., & Diem, S. (2014). Putting critical theoretical perspectives to work in educational policy. *International Journal of Qualitative Studies in Education*, 27(9), 1063–1067.
- Young, M. D., & Diem, S. (2017). Introduction to critical policy analysis. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 1–13). Gewerbestraße, Switzerland: Springer International Publishing.
- Young, M. D., & Reynolds, A. L. (2017). Critically examining policy workers and policy work within state boards of education. In M. D. Young & S. Diem (Eds.), *Critical approaches to education policy analysis: Moving beyond tradition* (pp. 19–42). Gewerbestraße, Switzerland: Springer International Publishing.



Critical Research Perspectives in School Leadership: Putting Dignity and Humanity at the Center

Irene H. Yoon

Critical inquiry is not just a social process; it is identity construction on a shifting and developing foundation. Students and novice researchers with a critical bent to their research interests may experience uncertainty related to their social and political identities and positioning: *Will my committee understand and support what I'm trying to do? Can I do this research if I am not a member of the "community" where I want to conduct my research? How do I take a critical lens on teaching without "teacher bashing"? What does it mean to be an "insider" or "outsider"? Am I unintentionally reproducing oppression or betraying my community by doing research at all?* These questions reflect the complexities of putting dignity, humanity, and activism at the center of all phases of critical research design, process, and writing. I focus this chapter not on "how to" engage in critical research in educational leadership, but rather on what many scholars think about when affirming and centering dignity and humanity in critical research. In particular, I review some trends in how critical research has developed in educational leadership.

I. H. Yoon (✉)

Educational Leadership & Policy, University of Utah, Salt Lake City, UT, USA
e-mail: irene.yoon@utah.edu

© The Author(s) 2018

C. R. Lochmiller (ed.), *Complementary Research Methods for Educational Leadership and Policy Studies*,

https://doi.org/10.1007/978-3-319-93539-3_6

What is clear across fields is that there is no one right way to “do” critical research, whether on educational leadership or elsewhere. Though there are shared principles and stances that define critical inquiry, there are few hard and fast rules. Some research might seem critical in intent (e.g., critiquing systems of oppression in educational institutions), and yet not explicitly identify as such, with no mention of “critical” or “critical theory.” In addition, a range of terms hint at critical perspectives, such as “social justice,” “praxis,” “intersectionality,” “neo-Marxism,” and “neo-liberalism” in educational organizations and leadership. The language of “equity” also comes with political twists and turns.

Despite varied uses of terms, what is common is that critical research takes the stance that all social phenomena are political, meaning that they are inherently infused with power and conflict related to social structure. Thus, critical research is explicitly theory driven (Erickson 2004; Fairclough 2015), intended for social change and advocacy (Kress 2013), and seeks to decenter and dismantle systems of power that result in dehumanization and oppression (Fine et al. 2003; Smith 1999). Critical research does not shy away from this agenda. Because of this systemic view, critical research is situated in historical and political contexts, recognizing that power always comes from somewhere, and power is always at work. Thus, the primary concern of critical research is to ensure that “what is” does not continue to structure realities for marginalized peoples. Critical research emphasizes, at least implicitly, “what should be.” Critical research insists society can “do better” for justice and affirming humanity of all people.

In turn, this perspective suggests a reality that is different from mainstream narratives in U.S. society. Rather than an idyllic past, critical scholars view history in terms of agentic struggles for liberation against intentional, systemic policies and programs designed to strip peoples of their languages, identities, religions, and recognition as human beings. Rather than a postracial or color-blind set of values, critical scholars recognize that schooling and society are inherently biased and normed to exclude and even dehumanize students of color, students with disabilities,¹

¹The politics of language and naming are always evolving. Disability studies, a diverse and interdisciplinary field, deconstructs ableism (the privileging of some abilities over others and erasure of differences in ability) as well as articulating how students are *disabled by* the structures of schooling and society that create barriers for these individuals (Goodley 2017). In this case, calling them “disabled students” reflects this systemic oppression and exclusion. I use different terms depending on the issue and meaning at hand. In this case, I choose the term “students with disabilities” to forefront the human being who is labeled by socially constructed notions of ability and normality.

gender identity, students who live in poverty, and students who are undocumented. At its core, critical perspectives on education in the United States argue that schooling has been a tool for colonization and control since Europeans first arrived on the continent.

Finally, critical inquiry at large recognizes that research has played an important role in colonization, slavery, and legalized oppression (Smith 1999). Research and science have been used to argue for the inferiority of intelligence, ability, and knowledges, and the legacies of these arguments—and the research that spurred them on—continue today. Critical research is interested in challenging hegemonic epistemologies and questioning institutionalized definitions of what constitutes “disciplined inquiry,” or the boundaries of objectivity and rigor in research (Shulman 1997).

In this chapter, I demonstrate how scholars have found various ways to approach critical research in educational leadership. Thus, I discuss what might make research methodologies “critical,” attending to the role of theoretical frameworks in these approaches. Then, I review examples of critical research in educational leadership, with a focus on how critical frameworks have evolved in contesting key assumptions about the realities and structures of schooling, leading, and learning. I conclude with recommendations for conducting critical research as a novice scholar.

REVIEW OF RELEVANT THEORIES

Critical research in education has important historical roots that go back to W.E.B. DuBois (1903) and Carter G. Woodson (1933). The strategic underground fights for literacy among enslaved people is also an ancestor of critical research and ways of knowing (Collins 2009; Walker and Byas 2009; Williams 2005). These roots point to the long history of critical research. Thus, though critical research has entered the scholarly field of educational leadership gradually, it is important to recognize that there have been people arguing for critical orientations to educational leadership for many years. In recent decades, the work of scholars such as William Foster, Khaula Murtadha, Michael Dantley, Linda Tillman, Barbara Jackson, Catherine Lugg, in an incomplete list—and their active efforts to mentor critical scholars of color—have opened up space for the growing diversity and advancements of critical research in educational leadership.

Historically, the term “critical” research had grounding in Marxist views that society is organized around economic capital and that this organization oppresses the majority of society while enriching a select,

small master class of bourgeoisie. This organization around capital disconnects workers from benefiting from their own labor, from recognizing their own oppression. Thus, Marxist critiques connect education to the struggle for liberation and critical consciousness and to identifying the reproduction of oppression (Bowles and Gintis 1976; Freire 2000). The uptake of Marxist ideas began in curriculum theory (Apple 2004; Giroux 1981) but expanded into educational leadership and policy with explorations of institutional structures, school reform, and logics of schooling (Anyon 1981, 1997; Oakes 1985).

After that “critical turn” to Marxist and class-based analyses of oppression, critical race studies entered the landscape (Gottesman 2016). Ladson-Billings and Tate (1995) are credited with bringing Critical Race Theory from U.S. legal studies directly into education research, arguing that racism is the primary organizing factor of U.S. social and economic relations, ideological and cultural values, and therefore, of schooling. CRT’s influence foregrounded, critiqued, and dismantled existing knowledge bases of race and racism in educational research. This iteration of race theory was concurrent with the uptake of intersectional analyses of race and gender, such as Black Feminist epistemologies (Collins 2009; Crenshaw 1991). These intersectional approaches argued that experiences of racism are gender specific, that knowledge bases stemming from these experiences are complicated by multiple oppressive conditions and identities. Critical research, hence, has evolved to include poststructural views, with applications in education most directly drawing from gender studies, feminist theories, and queer theories (Butler 1990, 2004; de Lauretis 1986; Lather 1991; St. Pierre and Pillow 2000). Each of these theoretical families have explored the construction of sexuality and gender in (predominantly Western) societies, arguing for deconstruction of categories that might seem natural, such as male/female. This deconstruction illuminates the socially constructed inadequacy of these categories as normativizing forces that reproduce power and marginalization. Because of the need to understand and transform the social construction of these categories, poststructural scholars have paid close attention to language. In the last decade, perspectives from Foucault (1969, 1975) and Bourdieu (1984) have influenced educational research with discourse-based perspectives on power and normativity. Finally, though most critical scholars of educational leadership have drawn on Western scholars’ work in these areas, increasingly, the work of transnational and decolonization scholars also influences critical research in theory and method (Smith 1999). All these major critical

social theories focus on *power*: examining the reproduction of inequality and oppression and deconstructing the processes and histories that have led to and perpetuate marginalization. This worldview and focus is intended to find leverage points for action and social transformation, from everyday instances, to policy formation, to macro-social ideologies.

Shifting power and sources of knowing suggests that critical research has to do with the power to define, to know, and to be known (or not). This means that critical research has to do with power and conflict. Social identities, social locations, access, and normativity are all at play, in addition to material (economic) privilege; these are studied as units of analysis that connect everyday events and individuals to society because power is institutional, systemic, structural, historical, and social, not only individual.

Rather than decontextualized, “objective” research, critical research views reality and knowledge as deeply *situated in local context and social histories*. Thus, critical research transitions methodologies away from traditional assumptions about the roles and relationships of researcher and research participants and objectivity and validity. It questions the taken-for-granted assumptions in traditional “malestream” paradigms (Code 1991) about who produces and owns knowledge, choosing to center traditionally marginalized perspectives not only in research findings but in the definition of worthy research agendas (Collins 2009; Harding 1993; Smith 1999). Such moves challenge the assumption that objectivity and validity of research have been pursued appropriately (Harding 1993; Lather 2006; Erickson and Gutierrez 2002). It also underscores the importance of questioning who benefits from research and the ethics of research on and with marginalized peoples and groups (Fine et al. 2003). That is, critical research looks inward in response to histories of institutionally sponsored research, which has long been used to oppress and to justify dehumanization of non-White, European peoples (Smith 1999; The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research [The Belmont Report] 1978).

In addition to transforming research traditions, critical research focuses on *social change and activism* because of relationships between individual, community, and society. Critical research centers the agency of individuals and collectives in transforming social structures and policies at local, regional, national, and international levels. Critical research unearths the sources and hidden, invisible structures of power in order to take this action-oriented, explicitly politicized stance. This unearthing process takes place in various ways. Some Critical Race scholars have emphasized

counterstories and critical histories that change taken-for-granted, mainstream narratives (e.g., Delgado Bernal and Villalpando 2002). Others have deconstructed institutional discourse and language use and what they reveal about institutionalized power and exclusion (e.g., Yoon 2012, 2016). Still more highlight ways to do research in partnership with marginalized communities (Green 2017; Ishimaru 2013), in efforts to leave power over knowledge creation in the hands of researcher participants (Cammarota and Fine 2008; Smith 1999).

The centrality of context, the desire for narratives and knowledges from silenced and oppressed people, and the engagement in local projects may be why qualitative methodologies appeal so strongly to many critical scholars. Such advocacy challenges traditional assumptions about objectivity and bias; the potential for the researcher to influence the phenomenon and people in the study is not only taken for granted but also accounted for and desired (e.g., Pillow 2003; Smith 1999). Critical research seeks to upend that relationship for change and empowerment of traditionally excluded and oppressed communities and individuals.

APPLICATIONS TO THE CRITICAL STUDY OF LEADERSHIP

Research on educational leadership has experienced these intellectual tides along with the broader field of educational research. Though perhaps not always the leader in taking up critical theories, much research on educational leadership now adopts these lenses, which are appropriate for studying institutional power and authority—such as the roles of school and system leaders and the organization of schooling. I am not the first to review these trends of critical research in educational leadership (cf., Gooden 2015; Young and López 2008); thus, I begin with a cursory overview of trends in critical research on educational leadership, and discuss some of the shifts over the past 15 years.

Critical research in educational leadership began, by necessity, with development of conceptual frameworks. Critical scholars conceptualize justice and strategic transformation as essential to anything relevant to educational leadership: instruction, curriculum, assessment, community relations, the law, and management. Critical perspectives on justice are not something that need to be included in definitions of leadership; they are the center, the roots, the heart, and the foundation of leadership. For example, Gooden and Dantley (2012) assert that critical race studies of school leadership—in both preparation and practice—argue that scholars

and practitioners must think not only about how race needs to be integrated into components of leadership but also about how race is central to educational leadership itself.

If this is the critical perspective on leadership, then logically, critical research on leadership is interested in understanding not how people “make room for” equity and justice, but rather how they make it the center of their leadership, or otherwise deconstruct what it means to recognize educational leadership as traditionally grounded in racism, heteronormativity, ableism, patriarchy, and classism. For instance, Diem and Carpenter (2013) interrogated silences of race in leadership preparation. From talking to students and professors, they argued that silences around race—even in intentionally structured conversations about race and racism—are strategic. They utilized Mazzei’s (2007) conceptions of silence to identify and characterize when and where conversations about race occur, and when and how they are silenced in different ways, both in classrooms and in program structure.

In addition to being able to identify and deconstruct the hegemony of whiteness and racism, capitalism, heteronormativity, and ability, among other systems, critical scholars also study educational leadership that seeks to transform and disrupt these systems. This approach to research unveils tensions, burdens, and strategic successes of critical leadership that partners with traditionally marginalized communities, directly addresses racist systems, and resists unjust policies. A significant strand of these critical studies examines community-centered paradigms that align with Freire (2000) principles of critical consciousness and praxis (Green 2017). In these ways, critical research on educational leadership recognizes, celebrates, and seeks to understand the depth of leadership that has an explicit focus on empowerment, critical consciousness, and humanization.

Related to this agenda, critical studies of educational leadership reframe paradigms of leadership and of who leaders are (e.g., Dantley and Tillman 2010). The area of work where this has the most depth seeks to learn from historical and contemporary experiences of African American school and system leaders (e.g., Alston 2005; Murtadha and Watts 2005; Walker and Byas 2009; Horsford 2010, 2012; Tillman 2004). Studies now engage Latino/a school leaders (e.g., Mendez-Morse et al. 2015) and LGBTQ school leaders (e.g., Lugg 2006). Thus, critical research on educational leadership is expanding to consider multiple forms, spaces, and types of leadership. For instance, beyond school leadership, community leadership (Green 2017) and youth leadership (Mansfield 2014) are all increasingly areas of interest in reframing and decentering traditional hierarchies and spaces of leadership.

These extensions to conceptualizing and studying leadership, both inside and outside of schools, continues to focus on if and how critical approaches to research and practice can pursue justice through praxis.

Evolving Research Agendas One important evolution is the *centrality* of critical approaches in educational leadership research, which has accelerated in the last 15 years. This is evident in conference programs of Division A of the American Educational Research Association (AERA) and the University Council of Educational Administration (UCEA). Growth is also represented in the issues of *Educational Administration Quarterly* (EAQ) and, to a lesser extent, the *Journal of School Leadership* (JSL), two leading peer-reviewed publications in the field.

While critical research began as limited to special issues of leading journals such as EAQ, it has increasingly become a powerful voice in every issue. For example, the remarkable October 2005 issue (Volume 41, Issue 4) of EAQ explored the stories and philosophies of African American educational leaders, with articles authored entirely by Black scholars. The December 2007 issue (Volume 43, Issue 5) of EAQ was the first one focused on Critical Race Theory applications to educational leadership. And the December 2014 issue of EAQ, on the legacies of *Brown v. Board of Education*, reflected the increasing diversity of critical lenses and approaches to educational leadership, as well as the permanence of racism in educational leadership and schooling (Bell 1992). Beyond special issues, EAQ increasingly publishes critical studies as well as scholars of color who do critical research in educational leadership. This “seat at the table” is to be celebrated, though it is limited. For instance, as of writing this chapter, the last decade of issues in two other leading journals in educational leadership (JSL and *Journal of Research in Leadership Education*) have far less representation of critical research in their pages; there is none in *Journal of Educational Administration*, a non-U.S.-based journal.

It remains to be seen if this critical shift will continue to grow and transform not only scholarly institutions but also the preparation of educational leaders in K-12 schools (McKenzie et al. 2008). Preparing critical educational leaders has increasingly been a focus of research because:

What some preparation programs have found is that too often our students have been ill-prepared to engage the multiple layers of social and cultural realities within which students and school communities live every day. (Gooden and Dantley 2012, p. 238)

Although preparation programs may not yet adequately prepare leaders for realities of school communities, the research base around existing efforts to improve these programs is growing (e.g., the February 2003 issue of EAQ). In addition to the program structures, some scholars also dedicate attention to the experiences of students of color in educational leadership programs (Young and Brooks 2008). Unlike in teacher education, there has been some, but not a proliferation, of critical empirical research on the pedagogical processes of learning and the professional identity development of White, middle-class, straight educational leadership candidates that use critical race, whiteness, and queer lenses of critique (exceptions include Diem and Carpenter 2013; Gooden and Dantley 2012; Marshall and Hernandez 2012). In addition, few studies critically explore leadership preparation program curricula (one exception is O'Malley and Capper 2015). These trends suggest that critical research on leadership preparation is making gains, but still has significant gaps.

Finally, a reflexive study on cultivating critical scholarship in educational leadership may be in order. Gooden (2015) notes the history of UCEA initiatives to redirect focus on equity and, more or less successfully, on race. The preparation of critical scholars in educational leadership (as opposed to practitioners) has been an area of rapid expansion that merits study. This can be credited, I believe, to two major initiatives of the University Council of Educational Administration (UCEA), the national association for educational leadership, preparation, and policy. UCEA has had the Barbara L. Jackson Network for scholars of color who aspire to the professoriate (disclosure: I am a Jackson Scholar alum). The Jackson Scholars Network began in 2003 and has grown to include over 300 scholars and alumni. As noted from conference programs, Jackson Scholars are overwhelmingly interested in critical research and social justice. Thus, critical research in educational leadership has carved out pockets of institutional space vis-à-vis explicit attention to supporting and networking emerging scholars of color, though these exist in tension with institutional histories and norms of erasure (Lugg 2016).

Methodology and Epistemology Moving on from the content and evolution of critical research in educational leadership, I turn now to considering methodologies and epistemologies of this scholarship. Educational leadership has not traditionally been a leader in the broad field of educational research and epistemologies, but two contributions have been particularly notable for their challenges to consider critical frameworks and

self-conscious attention to methodology. As White scholars, Scheurich and Young (1997) addressed an existing debate in *Educational Researcher* on “coloring epistemologies,” arguing that it was time to acknowledge and wrestle with whiteness in research methodologies. Part of the push-back on this article included the recognition that Scheurich and Young were engaging and seen to be initiating critiques that had long been discussed among scholars of color (López and Parker 2003). Two years after Scheurich and Young’s article, Capper (1999) made inroads on “in(queer)y” in education as a matter of not only how research is conducted but also how research problems are conceptualized. Also part of a broader emerging conversation, Capper called for exploration of the ways in which sexuality has been heteronormativized in education research.

As can be noted from the arguments above, scholars of educational leadership draw heavily on critical research methodologies from multiple and often interdisciplinary fields. Some traditions include critical ethnography (Carspecken 1996), youth participatory action research (Cammarota and Fine 2008), and critical discourse analysis (Fairclough 2015). I do not go into these in depth here but suggest that scholars who are new to an area of methodology delve into the relevant texts that problematize and give nuance to what it means to conduct critical research that falls on a spectrum of “participatory” approaches; explore the researcher’s role in designing and conducting research; and offer models of how to shift research participants’ roles in constructing knowledge through research. Attention to trade-offs, who benefits from research (Fine et al. 2003), and researcher’s exploration of discomfort (Pillow 2003) are all important to grappling with issues of validity, humanization, and knowledge construction. I address some of these tangles next.

RECOMMENDATIONS: “IT DEPENDS”

Whenever students ask me about how they should go about a certain decision in designing their critical research, I typically respond, “It depends.” Who will benefit? Who might be put in a compromised position by your research, including yourself as the researcher? Who will have power in your study, and in relation to whom? What will politically position your work to have impact in the communities or school systems—in the ways that you intended (cf., Fine et al. 2003)?

While the preceding work has given an overview of some of the grounding of critical research in educational leadership, I now address the “doing” of such work, organized around several big ideas and burning questions. An underlying thread to these questions relates to critical reflexivity, or recognition, reflection upon, and problematizing of the role of the researcher in knowledge construction as a relational (not power-neutral) task. As such, I shift gears to take a more personal, reflective engagement with questions that have been common among my students, and which I have asked myself regularly. I do not answer the questions in ways that give answers, but rather in ways that point to resources and strategies that can inform your own processes of reflection and decision-making. My thoughts implicitly point to what I think of as essential characteristics of “good” critical research, with the caveat that research methodologies are always evolving.

Who Are You? What Is Your Research Like? I offer my recommendations as a scholar who is learning and growing, and as one who benefited from the support and pathbreaking efforts of those who worked before me. I also make recommendations from the position of someone who has been puzzled by the in-betweens of being a Korean American who identifies as a woman of color, a daughter of immigrants, though one with unearned privilege. I see the world as a critical, progressive scholar whose work is deeply concerned with humanity and justice and the ways in which schools affirm or insult the humanity of marginalized children and their teachers.

I use and integrate a variety of frameworks in my research, depending on the nature of questions and problems in a particular study, as long as they will advance this agenda and concern. For instance, I have overlaid a critical whiteness and intersectionality lens onto sociocultural learning theories to study teacher collaboration using critical discourse analysis (Yoon 2012, 2016). Finding ways to bridge theoretical conversations and fields of research is a challenge that many critical scholars take on in order to fully embrace the complexity of historical, social, political, and cultural contexts. I take the approach of focusing on the everyday and the mundane to see how institutional norms and systems impinge on daily work and learning. Because some of my research critiques these everyday practices of educators, I am committed to grounding analyses in rich contextualization, walking a fine line between critique and appreciating the extraordinarily complex tasks of teaching and leading. For example, I have analyzed

discourse from the perspective of language use as a collective process of racialized, classed, and gendered socialization and learning. However, my methodologies depend on the purpose of the study at hand; I also have utilized grounded theory and critical ethnography.

This push to grapple with the mundane is an approach also argued for by Fine et al. (2003), who outline important ethical questions for researchers who wish to conduct critical research that includes participants who are traditionally marginalized by research and institutions. It is for these reasons, among others, that I utilize research methodologies that require physical and emotional presence, direct engagement, and extended time in school settings. Perhaps because of this approach, I take extensive reflection notes and question myself on how I engaged with people on a given day. I attend to the push and pull of (sometimes artificial-seeming) relationships and to not only the content but also the experiences of data collection and reflection. I was shown how to probe into my research in these ways by mentor scholars, including those who did not conduct critical research. I found my way through a patchwork of resources.

To Whom or Where Do I Go for Advice? Being a novice scholar can be overwhelming. Research requires full engagement of all your senses and cognitive processes. Doing critical research can be even more demanding because of its transformative commitments and politicized nature. Doing critical research means that researchers are directly confronted with painful experiences and injustices, even if they are focusing on stories and processes of resistance and triumph. Often, too, critical research will take a personal toll, particularly as researchers build relationships with participants, or come to see their own stories in those of their participants. However, researchers with supportive mentors on their own campuses have a first line of support. In my own experience, I have found that supportive mentoring and advice does not have to come from scholars who are experts in the methodology or content area that I am immersed in. Instead, mentors have been sounding boards and advocates with no agenda for themselves—experts who can speak to my needs from relevant, though different, experiences of their own.

Similarly, I consider like-minded peers, particularly those who are in similar stages of work, to be a primary source of support. It is incredibly helpful to dissect experiences together and to ask each other probing questions about how things happened, how they make me feel and why, and

how I might respond. Even if my peers are not doing the same kind of work, we can offer each other space for thought-provoking reflection.

In addition to the invaluable support of peers, it is important to build professional networks nationally and locally. In educational leadership, there are multiple structures for networking and mentoring through UCEA and Divisions A and L of AERA, and a good representation of scholars who either do or embrace critical and interdisciplinary research. Though they might begin with awkward conversations at conferences, some of these relationships and networks grow deep over time. I have trusted confidants at other institutions whom I have come to rely on seeing twice a year at conferences, and with whom I can troubleshoot the ins and outs of research partnerships and politics. These networks can be sources of support if novice critical scholars face skepticism from advisors and doctoral committees.

Finally, as always, peer-reviewed, highly respected research can be a source of support for your critical research topics and design. For instance, there is now a wealth of research on a range of critical methodologies and their purposes. There are generally at least several examples of scholarship where researchers have utilized strategies that you might want to use. That is, reading widely is not only fuel for theory and content, but for method. I also have found reading to be necessary to make sense of experiences for which there may not be a robust, coherent research base. As a Korean American woman who does critical research, I am often not fully insider or outsider in schools—and this shifts in relation to school staff, students, or parents. Reading has helped some of the frustrations caused by this sense of belonging-but-not-completely, but most importantly, being forced to read across disparate literatures helps me find rich seams for creativity.

How Should I Do Critical Research If I Am a Member of “the System” or “the Community”? Many educational researchers face a somewhat unique issue of being a member of the organizations and communities where they conduct their research. At the very least, researchers often represent institutions of higher education to participants (for better or worse). For instance, scholar-practitioners may be current or former teachers or leaders in a school or district system where they are partnering with parents on a research project. Scholars may be doing research projects in communities and schools where they reside (Delgado Bernal and Alemán Jr. 2017). These relationships and entanglements point to the ambiguity of critical research, in which scholars may have great privilege and also identify with

marginalized communities. Critical scholars are often both members of and outsiders within institutions of higher education (Collins 2009; Smith 1999). These liminal experiences can be confusing and cause misperceptions from both people in the university and in the field. I have found Tillman's (2006) reflections on positioning and being positioned as an African American researcher with African American research participants to be helpful for understanding my own experiences because no relationships can be predicted or taken for granted simply by race, ethnicity, or other presumptions of shared identity.

Because individuals have multiple facets of identity and community, scholars of color, scholars with disabilities, scholars who are genderqueer or trans*, all likely have skills of navigating educational institutions as both insider and outsider. Yet entering a research setting as researcher is a different position altogether. Within the umbrella of "researcher," critical scholars occupy different positionings (and have different positionings available to them) depending on whether they are a faculty member, a postdoctoral scholar, or a graduate student. These labels are barriers to some people and organizations, while they open doors to others. They also can obfuscate how participants see you and how you want to be perceived. For instance, when I was observing in one racially diverse classroom, most of the male students of color often saw me as an authority figure and were more reserved or even suspicious around me than their White peers and White teachers. On the other hand, several Asian American girls were far more open with and affectionate toward me than any other students. These perceptions of and reactions to me were undoubtedly related to race, gender, age, and were even influenced by my holding a notebook and pen.

Doing critical research in your own educational institution can be especially challenging. Critical research seeks to critique, dismantle, and transform systems of power—yet these institutions could be a researcher's employer and always hold power over permission and access to the research site. In these situations, critical scholars must not be dishonest about their research purposes, but rather find ways to communicate them in terms that make sense in light of the organization's strategic goals, initiatives, and leaders. It also helps to find allies within the school system who can vouch for the value of your work. Finally, it is important to recall that dismantling institutions and critiquing the actions of people in power does not mean treating them as simplistic, monolithic, or uncomplex and evil. I toe this

line when I write about whiteness in teaching and teachers' collaboration. As much as I insist on calling whiteness for what it is—an ideology that allows for dehumanization of particular students and that fuels White supremacy—it is important to recognize and explore how White teachers are not all the same; that many have complex practices that are sometimes equitable or even emancipatory, *and also* sometimes racist; that human beings are inconsistent; and that being a person of minoritized status does not mean I can, in turn, dehumanize the oppressor (Freire 2000).

I'm White, Male, Straight, Middle-Class, Fluent in English, and Typically Able: How Can I Do Critical Research? As I described above, embodiment is an issue for all scholars, in multiple and differing ways. Your self-presentation is an active choice that communicates with your research participants, and yet participants may or may not recognize the nuances of your identity and experience. They may see your physical body and how you present and perform it and form interpretations of you based on the very categories and beliefs that you are trying to dismantle. Your physical body and presence are not only part of collecting data in the field, but part of all phases of the research process. Your social location, too, informs what and how you have known and can possibly know (Collins 2009). White scholars who also occupy other structurally privileged positions have responsibilities on several levels—not to stand in the way of research that can be personally unflattering is one of them. In addition, it is possible for scholars who are not of marginalized identities and backgrounds to do research on issues of social injustice. Critical research is politically and personally tricky for these scholars in different ways than it is for scholars of color and other minoritized groups. I have found Pillow's (2003) advice on reflexivity and discomfort vastly helpful for embracing the complex and probing into discomfort and doubt. Fine et al. (2003) also provide guiding questions for reflection that assume that researchers are not members of, and occupy positions of power in relation to, the groups with which they conduct research.

In my own research, I am often privileged, powerful, and yet the only adult person of color in the room. I find this position deeply unsettling and therefore a productive place to continue double-checking my assumptions, probing for openness, keying in to small moments of reconstructing injustice (e.g., Yoon 2012, 2016). One primary fuel for inspiration and productive self-doubt has been to continue reading widely and ravenously, even texts that I disagree with.

How Do I Know If I Am Finding the Right Theories to Inform My Thinking? Reading widely and across diverse perspectives is a key component of rigorous and interesting scholarship. Creativity and insight are frequently generated by others' work. This generativity may point to relevant frameworks and lenses for a study. In fact, reading before, during, and after data collection and formal phases of analysis can lead to inductively finding frameworks that fit the data, and not only the other way around.

There are a few hints that can be helpful for the role of critical theories in critical research. In educational leadership, particularly, many existing theories of leadership were developed by White men for a predominantly White male field. As Young and López (2008) described, educational leadership can and has benefited from critical race studies, poststructuralist feminisms and gender theories, queer theories, and critical neo-Marxist theories. In addition to applying theories from other disciplines, several critical scholars have developed conceptual frameworks specifically for educational leadership (e.g., Alston 2005; Foster 1989; Horsford 2012).

In all research, theoretical frameworks ideally inform and are epistemologically aligned with methodological decisions, even implicitly. In critical theory, these ties are generally made explicit. This is because critical methodologies are designed for epistemological and ethical transformation; critical scholars shift even the collection of data away from epistemic roots in dehumanizing histories of objectifying research toward more participatory and participant-controlled research (Smith 1999). In general, critical methodologies are explicitly tied to theories about power and, as such, connect theories about social structure to the interpretation of everyday occurrences and practices. These theories also decry essentializing individuals while building collectivities and solidarity in diversity. These are not considered weaknesses of critical theory, but rather strengths (Capper 1999; Collins 2009).

Many of these critical methodological approaches will require questioning and politicization of researchers' positionalities and identities. It will also mean keeping a sharp eye out for the ethics of research—who “owns” data or “creates” knowledge? Who benefits from the research findings? Who presents the research, and where? These questions also hold for critical readings of critical research. Just like in “mainstream” research, not all critical research is argued well, done ethically, or offers important contributions to scholarship, policy, or practice. Scholars also argue that, despite

intentions, critical research methodologies do not necessarily resolve the problems of the traditions whence they came (Noblit et al. 2004). It is important to think critically about critical research, and to recognize the contributions and risks people took so that we could do so.

CHAPTER SUMMARY

Educational leadership is a field in motion, and critical scholars in this area fluidly move between understanding leadership and systems as they currently are, and imagining them as they need to be. Taken together, critical research of educational leadership relates to transforming the focuses, epistemological frameworks, and faces of inquiry. This means that critical research on educational leadership breaks past normative conceptions of educational leaders as White, straight, cis-male, typically abled, English-speaking, school principals. Further, if you think of educational leadership as co-constructed and contested practice, as critical scholars do, then it logically follows that critical research in educational leadership has blasted past traditional schema of leadership practices and settings. Though there is no right way to “do” critical research in educational leadership, there are specific histories and tensions that must be recognized in order to engage in research that is transformative intellectually, conceptually, spiritually, and pragmatically. This chapter has explored these histories and tensions to consider some of the ways that critical research can affirm dignity and humanity while confronting injustice. It is offered in the spirit of moving forward against oppression while recognizing the wind at our backs.

Recommended Readings

Foster, W. (1989). Toward a critical practice of leadership. In J. Smyth (Ed.), *Critical perspectives on educational leadership* (pp. 39–92). London: Falmer Press.

William Foster’s work was foundational for the recent decades of scholarship around social justice in educational leadership. In this chapter, Foster connected moral, communitarian, transformative, and critical perspectives to reframe educational leadership.

Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). New York: Continuum.

Pedagogy of the oppressed is an essential text for anyone interested in interrogating power and connecting theory to practice. Freire reviews dialectical relationships between oppressor vs. oppressed, dehumanization vs. critical

consciousness, and banking model of education vs. the pedagogy of the oppressed (now considered a precursor to critical pedagogy)

Smith, L.T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. New York: St. Martin's Press.

Beginning with a review of traditional research methodologies, particularly related to anthropology, Smith reviews the historically oppressive and inhumane roots of Western scholarship. The second half of the book imagines an indigenous research agenda and how it should be pursued with indigenous peoples and knowledges at the center.

REFERENCES

- Alston, J. A. (2005). Tempered radicals and servant leaders: Black females persevering in the superintendency. *Educational Administration Quarterly*, 41(4), 675–688.
- Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry*, 11(1), 3–42.
- Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform*. New York: Teachers College Press.
- Apple, M. W. (2004). *Ideology and curriculum* (3rd ed.). New York: Routledge.
- Bell, D. (1992). *Faces at the bottom of the well: The permanence of racism*. New York: Basic Books.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Cambridge, MA: Harvard University Press.
- Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. New York: Basic Books.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.
- Butler, J. (2004). *Undoing gender*. New York: Routledge.
- Cammarota, J., & Fine, M. (2008). *Revolutionizing education: Youth participatory action research in motion*. New York: Routledge.
- Capper, C. (1999). (Homo)sexualities, organizations and administration: Possibilities for in(queer)y. *Educational Researcher*, 28(5), 4–11.
- Carspecken, F. P. (1996). *Critical ethnography in educational research: A theoretical and practical guide*. New York: Routledge.
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43, 1241–1299.
- Code, L. (1991). *What can she know? Feminist theory and the construction of knowledge*. Ithaca: Cornell University Press.

- Collins, P. H. (2009). *Black feminist thought: Knowledge, consciousness, and the politics of empowerment* (2nd ed.). New York: Routledge.
- Dantley, M. E., & Tillman, L. C. (2010). Social justice and moral transformative leadership. In C. Marshall & M. Oliva (Eds.), *Leadership for social justice: Making revolutions in education* (pp. 19–34). Boston: Allyn & Bacon Publishers.
- De Lauretis, T. (1986). *Feminist studies, critical studies*. Bloomington: Indiana University Press.
- Delgado Bernal, D., & Alemán, E., Jr. (2017). *Transforming educational pathways for Chicana/o students: A critical race feminista praxis*. New York: Teachers College Press.
- Delgado Bernal, D., & Villalpando, O. (2002). An apartheid of knowledge in academia: The struggle over the “legitimate” knowledge of faculty of color. *Equity & Excellence in Education*, 35(2), 169–180.
- Diem, S., & Carpenter, B. W. (2013). Examining race-related silences: Interrogating the education of tomorrow’s educational leaders. *Journal of Research on Leadership Education*, 8(1), 56–76.
- Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life*. Cambridge, UK: Polity Press.
- Erickson, F., & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher*, 31(8), 21–24.
- Fairclough, N. (2015). *Language and power* (3rd ed.). Harlow: Longman.
- Fine, M., Weis, L., Weseen, S., & Wong, L. (2003). For whom? Qualitative research, representations, and social responsibilities. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research: Theories and issues* (2nd ed., pp. 167–207). Thousand Oaks: Sage.
- Foster, W. (1989). The administrator as a transformative intellectual. *Peabody Journal of Education*, 66(3), 5–18.
- Foucault, M. (1969). *The archaeology of knowledge*. Paris: Editions Gallimard.
- Foucault, M. (1975). *Discipline and punish: The birth of the prison* (Trans.: Alan Sheridan, 2nd Vintage Books). New York: Vintage Books.
- Freire, P. (2000). *Pedagogy of the oppressed* (30th anniversary ed.). New York: Continuum.
- Giroux, H. A. (1981). *Ideology, culture, and the process of schooling*. Philadelphia: Temple University Press.
- Gooden, M. A. (2015). Evidencing the effort: (Re)defining UCEA’s role in using leadership to center and advance equity in schools. *UCEA Review*, 56(1), 1–12.
- Gooden, M. A., & Dantley, M. (2012). Centering race in a framework for leadership preparation. *Journal of Research on Leadership Education*, 7(2), 237–253.
- Goodley, D. (2017). *Disability studies: An interdisciplinary introduction* (2nd ed.). London: Sage.
- Gottesman, I. (2016). *The critical turn in education: From Marxist critique to poststructuralist feminism to critical theories of race*. New York: Routledge.

- Green, T. L. (2017). Community-based equity audits: A practical approach for educational leaders to support equitable community-school improvements. *Educational Administration Quarterly*, 53(1), 3–39.
- Harding, S. (1993). Rethinking standpoint epistemology: What is “strong objectivity”? In L. Alcott & R. E. Potter (Eds.), *Feminist epistemologies* (pp. 49–82). New York: Routledge.
- Horsford, S. D. (2010). Mixed feelings about mixed schools: Superintendents on the complex legacy of school desegregation. *Educational Administration Quarterly*, 46(3), 287–321.
- Horsford, S. D. (2012). This bridge called my leadership: An essay on Black women as bridge leaders in education. *International Journal of Qualitative Studies in Education*, 25(1), 11–22.
- Ishimaru, A. (2013). From heroes to organizers: Principals and education organizing in urban school reform. *Educational Administration Quarterly*, 49(1), 3–51.
- Kress, T. M. (Ed.). (2013). *Using critical research for educational and social change*. New York: Routledge.
- Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record*, 97(1), 47–68.
- Lather, P. (1991). *Getting smart: Feminist research and pedagogy with/in the post-modern*. New York: Routledge.
- Lather, P. (2006). Paradigm proliferation as a good thing to think with: Teaching research in education as a wild profusion. *International Journal of Qualitative Studies in Education*, 19(1), 35–57.
- López, G. R., & Parker, L. (Eds.). (2003). *Interrogating racism in qualitative research methodology*. New York: Peter Lang.
- Lugg, C. A. (2006). Thinking about sodomy: Public schools, legal panopticons, and queers. *Educational Policy*, 20(1), 35–58.
- Lugg, C. A. (2016). *U.S. public schools and the politics of queer erasure*. New York: Palgrave Macmillan.
- Mansfield, K. C. (2014). How listening to student voices informs and strengthens social justice research and practice. *Educational Administration Quarterly*, 50(3), 392–430.
- Marshall, J. M., & Hernandez, F. (2012). “I would not consider myself a homophobe”: Learning and teaching about sexual orientation in a principal preparation program. *Educational Administration Quarterly*, 49(3), 451–488.
- Mazzei, L. A. (2007). *Inhabited silence in qualitative research: Putting poststructural theory to work*. New York: Peter Lang.
- McKenzie, K. B., Christman, D. E., Hernandez, F., Fierro, E., Capper, C. A., Dantley, M., ... Scheurich, J. J. (2008). From the field: A proposal for educating leaders for social justice. *Educational Administration Quarterly*, 44(1), 111–138.

- Méndez-Morse, S., Murakami, E. T., Byrne-Jiménez, M., & Hernandez, F. (2015). Mujeres in the principal's office: Latina school leaders. *Journal of Latinos and Education, 14*, 171–187.
- Murtadha, K., & Watts, D. M. (2005). Linking the struggle for education and social justice: Historical perspectives of African American leadership in schools. *Educational Administration Quarterly, 41*(4), 591–608.
- Noblit, G. W., Flores, S. Y., & Murillo, E. (2004). Postcritical ethnography: An introduction. In G. W. Noblit, S. Y. Flores, & E. Murillo (Eds.), *Postcritical ethnography: Reinscribing critique* (pp. 1–45). Cresskill: Hampton Press.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.
- O'Malley, M. P., & Capper, C. A. (2015). A measure of the quality of educational leadership programs for social justice: Integrating LGBTIQ identities into principal preparation. *Educational Administration Quarterly, 51*(2), 290–330.
- Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *Qualitative Studies in Education, 16*(2), 175–196.
- Scheurich, J. J., & Young, M. D. (1997). Coloring epistemologies: Are our research epistemologies racially biased? *Educational Researcher, 26*(4), 4–16.
- Shulman, L. S. (1997). Disciplines of inquiry in education: A new overview. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 3–31). Washington, DC: American Educational Research Association.
- Smith, L. T. (1999). *Decolonizing methodologies: Research and indigenous peoples*. New York: St. Martin's Press.
- St. Pierre, E. A., & Pillow, W. S. (Eds.). (2000). *Working the ruins: Feminist post-structural theory and methods in education*. New York: Routledge.
- The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Bethesda: Author.
- Tillman, L. C. (2004). African American principals and the legacy of *Brown*. *Review of Research in Education, 29*, 101–146.
- Tillman, L. C. (2006). Researching and writing from an African-American perspective: Reflective notes on three research studies. *International Journal of Qualitative Studies in Education, 19*(3), 265–287.
- Walker, V. S., & Byas, U. (2009). *Hello professor: A Black principal and professional leadership in the segregated south*. Chapel Hill: University of North Carolina Press.
- Williams, H. A. (2005). *Self-taught: African American education in slavery and freedom*. Chapel Hill: The University of North Carolina Press.
- Yoon, I. H. (2012). The paradoxical nature of whiteness-at-work in the daily life of schools and teacher communities. *Race Ethnicity and Education, 15*(5), 587–613.

- Yoon, I. H. (2016). Trading stories: Middle-class White women teachers and the creation of collective narratives about students and families in a diverse elementary school. *Teachers College Record*, *118*(2), 54.
- Young, M. D., & Brooks, J. S. (2008). Supporting graduate students of color in educational administration preparation programs: Faculty perspectives on best practices, possibilities, and problems. *Educational Administration Quarterly*, *44*(3), 391–423.
- Young, M. D., & López, G. R. (2008). Putting alternative perspectives to work in the politics of education. In E. A. Samier (Ed.), *Political approaches to educational administration and leadership* (pp. 155–172). New York: Routledge.



The Potential of (Participatory) Action Research for School Leaders, Local Policy Makers, and University-Based Researchers

Meagan Call-Cummings and Melissa Hauber-Özer

I walked tentatively into the modular classroom in rural Idaho, USA that October morning. Opening the door, the din of 52 high school students chatting and laughing and complaining all at once, in at least two languages, flooded me. I instinctively closed my eyes to adjust to the lack of natural light. When I opened them again, I saw only a few students sitting down. The others were standing, leaning into groups with friends and smiling.

After the bell rang, the teacher, Mrs. James (pseudonyms have been used throughout the chapter), cleared her throat to get their attention. They quieted. She introduced me as her niece, a doctoral student, who focused on participatory action research. The students seemed to look at me all at once, intrigued, but at the same time, bored. Nervous, I introduced myself.

“My name is Meagan. My aunt, your teacher, Mrs. James, invited me to come meet you today and work with you this week – if you want. Like she said, I do what’s called Participatory Action Research, or PAR. In PAR,

M. Call-Cummings (✉) • M. Hauber-Özer
Graduate School of Education, George Mason University, Fairfax, VA, USA
e-mail: mcallcum@gmu.edu; mhauberr@masonlive.gmu.edu

researchers like me work with everyday people like you, to solve problems or make changes. You become researchers, just like me.” I babbled for a bit, I think, unsure when to stop. Looking around at the students and at my aunt for some indication that I should let the students speak. After probably too long, I paused.

Jaime, who, I found out later, was the class president, spoke up: “So, Miss Meagan, what do you want us to *do*?”

I hesitated, trying to choose my words carefully. I was fully aware of my privilege—a young, naïve, White doctoral student in the middle of a classroom of 52 Latino/a students, many of whom are considered “undocumented” by the United States immigration system because their parents brought them to the United States when they were young from countries like El Salvador, Honduras, and Mexico without necessary papers.

“Well,” I started, “I guess it’s not what *I* want *you* to do. It’s more what *you* want to do. Maybe we could start by doing a group discussion or brainstorming session.” They seemed to be paying attention a bit more closely now. “What if I asked you what makes you mad? What do you not like or think is unfair about your lives, your community, this world? What pisses you off?” They seemed to perk up a bit when I used language they do not hear teachers use. “What if you could ask any question and get an answer? What if you could talk to anybody—even Obama—and they would listen? What would you say?”

As we started discussions that day, the students became lively. They wanted to tell me and Mrs. James what was unfair, what was wrong, and what questions they had. And they had ideas about how to fix things. Ideas ranged from how to recycle more at school, to how to improve the immigration system, to how to provide more equitable access to resources and opportunities in the United States. After two days of brainstorming and discussions, the students settled on a question they all wanted to work on together: *Why are our teachers racist?*¹

AN INTRODUCTION TO (PARTICIPATORY) ACTION RESEARCH

I entered the classroom that autumn day sure only of one thing: that I was interested in how participatory action research (PAR) could work in a school or classroom. I had taken a class on PAR and had strong ethical leanings toward a methodological approach to research that would include and prioritize voices of students historically disenfranchised by the research

¹All details of this project can be found at www.researchforempowerment.com/burned/

process itself, but other than that I was not sure of much. I did not know what the students would want to research, or if they would be interested at all. I did not know how they would react to me entering their classroom, their space, and their group.

I learned a lot that first day and then during the 18 months I worked with Mrs. James and her students. While the number of students fluctuated with class enrollment, a core group of about 30 students worked together with Mrs. James and me to start a critical conversation about race and racism in the school and community with those in positions of power. We used PAR to guide this conversation.

The purpose of this chapter is to provide school leaders, local policy-makers (e.g. school board members, superintendents, and elected leaders), and novice researchers with a broad introduction to Action Research (AR) and PAR. This chapter positions AR/PAR as one methodological approach that school leaders and local policy actors can use to understand local challenges and induce changes in response to these challenges. Throughout the chapter, we discuss the origins and theoretical underpinnings of AR and PAR, present one commonly used AR/PAR process, and provide recommendations about how novice scholars can engage in AR/PAR. Throughout the chapter, we draw on examples from scholarly literature and the first author's research in rural Idaho, which took place as part of a research project entitled *Why are our teachers racist?* Throughout the chapter, references to I or we are made in reference to the first author.

ORIGINS AND THEORETICAL UNDERPINNINGS OF AR AND PAR

Scholars have offered multiple accounts of the historical origin of AR and PAR (Santos 2015). Many scholars credit Kurt Lewin with the development of AR through his attempts to instigate change in social behaviors in the 1940s (Kemmis and McTaggart 2005; Levin 1999; Lind 2008; Winter 1998). Lewin is widely viewed as the founder of AR because he challenged the dominant positivist paradigm tied to the scientific method (Levin 1999) by adopting an inquiry process involving reconnaissance, planning, and action (Winter 1998). His experiments and those of other early action researchers, such as Ronald Lippitt and Marian Radke, sought to produce practical solutions to real-world issues in the United States (Lind 2008), including racism, oppression, and intergroup conflict (Glassman et al. 2013).

Several strands developed out of Lewin's efforts as well as early work in Britain (Elliott and Adelman 1973) and Australia (Carr and Kemmis 1986) that sought after more "practical," "critical," and explicitly "emancipatory" research (Kemmis and McTaggart 2005). All of these efforts resulted in a diverse and overlapping web of frameworks and methodologies applied in fields across the social sciences, education, and health sciences (Stark 2014; Santos 2015).

In the 1970s, the participatory strand grew out of a commitment to holistic epistemology, knitting together reason and personal ethics, or "head and heart" (Fals-Borda 2001, p. 29). Early PAR adopters were in part inspired by Paulo Freire's (1970) *Pedagogy of the Oppressed*, which exhorted readers to listen to the voices of the oppressed and seek their liberation from unjust, systemic marginalization by enabling them to take control of their circumstances (Levin 1999; Lind 2008). Freire's critical theory privileged "ordinary people's knowledge" (Lind 2008, p. 223), melding naturally with the participatory, democratic approach of AR. This resulted in a type of AR focused on the empowerment of marginalized individuals to inform and lead change-focused inquiry, thereby dismantling the traditional research hierarchy (Fals-Borda and Rahman 1991; Santos 2015).

From a brief glance at the literature, it may seem that rather than a cohesive methodology, AR/PAR became a broad, collaborative movement of "social activists, organizational/community leaders, and scholars" (Glassman et al. 2013, p. 274). However, Kemmis and McTaggart (2005), Santos (2015), and others have made clear that AR and PAR are distinguished from other approaches to research in unique and important ways. While Santos (2015) focuses on PAR's strong connection to the political realm as its defining and distinguishing feature, Kemmis and McTaggart (2005) offer three distinct features differentiating AR and PAR from more conventional research:

- Shared ownership of research projects,
- Community-based analysis of social problems, and
- An orientation toward community action. (p. 273)

Understanding each of these unique attributes and how they can be authentically achieved is important for those interested in pursuing AR/PAR because otherwise one runs the risk of what Fals-Borda and Rahman (1991) call inauthentic engagement or "cooptation" of AR/PAR

(pp. 28–30). Rather than treating AR/PAR as an “easy blueprint” or “panacea” approach for school leaders, local policy makers, or others, the methodology should be taken up along with its epistemological roots in authentic participation, inclusive knowledge production, and local-level action. These three critical aspects frame our discussion of AR/PAR in the remainder of the chapter, including the further description of my research in rural Idaho to more clearly illustrate how these features can look in a real-world AR/PAR project.

SHARED OWNERSHIP

One of the central tenets of AR/PAR epistemology is the idea that expertise lies in everyday lived experience and should not be conceptualized as outside of the reach of “ordinary” people (Fals-Borda and Rahman 1991; Freire 1970, 1974; Kemmis and McTaggart 2005; Santos 2015; Sohng 1996). Going further, AR/PAR is rooted firmly in an epistemological stance that not only includes but also prioritizes the experiential knowledge(s) of those often seen as non-experts. Sohng (1996) explains why flipping the knowledge hierarchy on its head is so crucial for AR/PAR:

The production of knowledge has become a specialized profession and only those trained in that profession can legitimately produce it...In modern society, knowledge has been increasingly concentrated in the hands of “experts” and the elite class they represent. The ideology of the knowledge society has...historically privileged the pronouncements of trained experts over the discourses of “ordinary” people (Foucault 1980). Today this ideology manifests itself in deference to experts, and ultimately the subordination of people’s own experiences and personal meanings to expertise. As a result, decisions affecting ordinary people are based on “expert” knowledge, denying the rationality of individual citizens and their life experiences. Understanding human nature and the problems of living becomes the purview of scientists, rendering people dependent on experts to explain and oversee their life experiences (Berman 1981). Hence, the specialists dominate any debate concerning issues of public interest because ordinary people are unable to enter the scientized debate, as they lack the technical terminology and specialized language of argumentation. (Habermas 1979, pp. 79–80)

Because it has become expected or unquestioned that “specialists” and “experts” maintain firm control over public debate around issues like education, health care, or criminal justice, “ordinary” people are stuck in a

dependent position, rather than in one that puts them in control of their own lives and decisions that will affect them.

Both AR and PAR seek to address problems of power, hierarchy, and dependence in the production of knowledge by engaging members of local communities as “the primary agents of change” (Glassman et al. 2013), whatever that change may be. The relationship between the researcher and the researched, as well as the situation being investigated, becomes *interdependent* (Levin 1999; Sohng 1996). AR/PAR redistributes power through equal participation in the investigation, and repositions a university-based researcher—sometimes referred to as “outside” or “external”—as a facilitator (Lind 2008) or co-researcher, equal in power with all other co-researchers—sometimes referred to as “inside” or “internal”—in a research collective (Call-Cummings 2017; Fine et al. 2004; Freire 1982; Gaventa 1991).

This shared ownership ideally extends to all parts of the AR/PAR research process. Bourke (2009) says participatory research is “a research process which involved those being researched in the decision-making and conduct of the research, including project planning, research design, data collection and analysis, and/or the distribution and application of research findings” (p. 458). Fine and Torre (2006) aver that no matter how we define participatory research, it must have an “understanding that people—especially those who have experienced historic oppression—hold deep knowledge about their lives and experiences, and should help shape the questions, [and] frame the interpretations’ of research” (p. 458).

One of the many ways this shared ownership can be achieved is through co-authorship of scholarly publications or other co-dissemination of process and findings. An example of co-authorship can be seen in Fine et al. (2004), cited above, where Michelle Fine published a scholarly book chapter with several of her co-researchers from a project conducted around the impact of college education on inmates at a women’s prison in New York. These co-researchers were, in fact, inmates at the prison and conducted the investigation alongside Dr. Fine. The decision to include them as co-authors of the publication meant that they literally co-owned those findings.

Another approach to co-dissemination of findings involves collectively deciding what findings are most important, who needs to hear or see or experience the findings, and how that will happen. In my study in rural Idaho, I asked students who participated in the research project to think how and with whom would they share our findings if money, time, and

energy were not an issue. Some students suggested creating a documentary film and sending it to President Obama. Other students suggested writing a book describing their experiences with racism in schools. Finally, other students recommended holding a mandatory school assembly at which the research team could present the most important aspects of our work together. Through a process of group brainstorming, we decided to start a conversation at the school and with the local community about how racism affected the everyday lives of these students.

A few of the students found out when the local school board would next meet. The group selected four students to act as representatives at the meeting, and Mrs. James got us on the agenda. We attended the meeting, and there the four selected students shared their stories of how they experienced racism at their school and in the community. There was no “ask.” They simply wanted to be heard and have their experiences acknowledged.

COMMUNITY-BASED ANALYSIS

Analysis of data conjures up images of a lone researcher in a white lab coat, surrounded by beeping computers, smoking beakers, and chalkboards full of confusing algorithms. Like research design and data collection, this stage is often seen as the purview of the “experts.” Dodson and Schmalzbauer (2005) remark that “the interpretive stage of research is conventionally one in which the data become the sole province of the researcher community for analysis and construction of meaning” (p. 954). Curry (2012) referred to this practice of extracting data from participants and then hoarding and interpreting it as “hit and run” research and cites Reinharz (1979) calling it “the rape model of research” (p. 95). Whatever metaphor is used, this practice reproduces, extends, and further entrenches unequal power relationships between the researcher and the researched. Meaningful participation of community members in the analysis process, on the other hand, can enhance insight into participants’ ways of knowing, which gives AR/PAR the potential to produce more “authentic” knowledge.

Community-based analysis is an essential yet often overlooked or at best thinly discussed aspect of AR/PAR. While participatory analysis is not overly emphasized in methodological literature, the few who do discuss it provide helpful examples of how it can be carried out effectively. Curry (2012) provides a useful example of collaborative analysis with teachers and school reformers, describing “an onion-skin order whereby researchers

shared their emerging analytic claims first and then study participants shared their interpretations last so that they could contest or affirm the research team's understandings" (p. 94). This process allowed all involved in the analysis to articulate their own views while simultaneously remaining open to learning through dialogue. Curry explains a second approach as a pulling away of a curtain or a fishbowl approach, through which participants watched researchers conducting analysis. University-based researchers strategically distributed opportunities for community-based researchers to speak back, critique, and question that analysis. Both approaches highlighted AR/PAR's commitment to level power dynamics between "outside" and "inside" researchers.

Dodson and Schmalzbauer (2005) refer to the process of participatory analysis as collaborative interpretation and co-production of knowledge. From their work with poor mothers, they provide a useful example of how participants can be meaningfully included throughout data analysis. In this study, participants were invited to be co-analysts of the data, working together with the "outside" researchers to see inside the data, to make explicit what was implicit. They did this by asking a few simple questions:

1. Are we hearing what is really going on in these people's lives? Does this sound like what is going on in the lives of people you know?
2. What else do you think is going on that is not represented here?
3. We do not understand the meaning of what is being said here. Can you explain to us what this means? (List points where clarity is missing.)
4. If you were trying to find out what we are trying to find out about... what would *you* expect people not to share or talk about? Why would they be hesitant to talk about this? What's at stake?
5. After going over all of the information that we have gathered from our interviews, we have concluded the following... Do you think we have this right? Are we getting it? (pp. 955–956)

Dodson and Schmalzbauer note that they always improvised this process and asked questions that seemed appropriate given the particular context. They find that "when participants believed the interaction was open to their knowledge and their critique of our limitations, the most unexpected and revealing commentaries emerged" (p. 956). Grounded in the epistemological commitment that ordinary people are the experts of their own lives, this type of collaborative data analysis can uncover participants' deep knowledge and insight.

In my research in rural Idaho, I worked to achieve equal power relationships among members of the collective. But no matter how hard I tried, I remained the outsider. At one point I decided to capitalize on that position and, with the permission of the students with whom I worked, conducted one-on-one interviews with a few of their White teachers. The students felt uncomfortable having conversations with some of their teachers about race and racism, so, in my position as outsider, I was able to do that with low personal risk. After I conducted the interviews, I anonymized the data and brought it back to the students for interpretation. Similar to the Dodson and Schmalzbauer (2005) approach, our collective read through the data together and I asked them questions like, “What do you think she meant by that?” “I don’t understand what he’s saying there. Can you help me understand?” “This is what I thought about that. Is that right or am I missing another possible interpretation?” By positioning myself as learner rather than expert, I placed value on the students’ interpretations based in their lived experiences. They became experts and their analysis became a useful tool. In this way, we redistributed power over and within the research process.

ORIENTATION TOWARD ACTION

AR/PAR engages groups of researchers without deference to title or position with the aim of some practical action or change, such as developing a social justice-oriented high school curriculum (Cammarota and Romero 2011); understanding barriers to housing for women fleeing domestic violence (Ponic et al. 2010); challenging bullying in K-12 schools (Stoudt 2007); or influencing women’s health policy (Wang 1999). In fact, Kemmis and McTaggart (2005) argue that the research-action dualism typically presented in other forms of more traditional research, where research is conducted according to some predetermined process by those in authority to do so and then is followed by information-giving and action-taking based on that information, does not exist in AR/PAR. Instead, they suggest that “research and action converge *in* communicative action aimed at practical and critical decisions about what to do in the extended form of exploratory action” (p. 319). Kemmis and McTaggart, referencing Habermas’s theory of the system and lifeworld, echo the sentiment that “research and action are to be understood not as separate functions but rather as different moments in a unified process of struggle” (p. 320).

This breaking down of the research-action dualism was clear in my work in rural Idaho. From the moment I stepped in to the students' classroom, our focus was on what we needed to *do*. The process of gathering data and analyzing it became steps toward taking action. Indeed, it centered on taking action and calling for change. For example, soon after we decided to try and better understand the students' relationships with their White teachers, the students expressed their interest in interviewing their teachers themselves about how they see their Latino/a students. This, we envisioned, would be a type of data collection and would put our collective in a position to conduct analysis that would form the basis for a call to action of some sort. Individual students volunteered to invite teachers to be interviewed. Eight teachers were invited and four agreed to participate.

I worked with the student co-researchers to develop a flexible interview protocol. We brainstormed about what types of questions they should ask the teachers. Initially, students wanted to ask questions like, "Why are you racist?" or "Why do teachers at this school favor White kids?" or "Why is it always the brown kids who get in trouble?" We had conversations about how questions like that might make defense mechanisms flare up in teachers so that students really would not get useful or "authentic" data. We role-played with students becoming teachers that would be interviewed and thought about how various questions would garner different types of responses. We settled on questions that seemed to get at racism in more circuitous ways:

1. Why did you start teaching?
2. Without giving names, can you talk about some students you are concerned about? Why are you concerned about them?
3. Let's say a student comes to class and tells you he hasn't done his homework that day. What would a good teacher do? What would a weak teacher do?
4. Let's say a student falls asleep in class almost every day. What is the right thing to do? (After the teacher has answered, ask the follow up question: How have you come to know that this is the right thing to do?)
5. What does it take to be a really good teacher of Latino students? (After the teacher has answered, ask the follow up question: How have you come to know these things?)
6. What does it take to be a really good teacher of at-risk students? (After the teacher has answered, ask the follow up question: How have you come to know these things?)

7. Can you tell us about a situation you have had with a student you were concerned about? How did you deal with that student or that situation?
8. Can you tell us about the best student you've ever had? Tell us all about this student.
9. What teaching methods do you use for those that have a hard time understanding the subject you teach?

Looking back, I see that in addition to asking questions to gather data, we asked questions that facilitated the teachers' authentic reflection on their own experiences. Giving White teachers the opportunity to reflect on and then discuss their experiences with Latino/a students, make explicit their own biases, and no longer take for granted preconceived ideas about teaching and learning relationships was action in the form of information gathering.

THE AR/PAR PROCESS

Traditionally, research is conceptualized as "detached discovery and empirical verification of generalizable patterns" (Sohng 1996, p. 78) that is based on a systematic, linear process of well-defined steps, which include: (a) identifying a research problem and questions, (b) designing the study, and (c) collecting and analyzing data. However, like much inquiry with roots in critical theory, AR/PAR demands a more flexible and responsive procedure. Kemmis and McTaggart (2005) suggest that instead of some linear, mechanistic procedure to which a researcher is bound in order to claim validity or rigor, AR/PAR is much more like a "spiral of self-reflective cycles" (p. 276) that includes aspects of planning, acting/observing, and reflecting over and over again. Knapp (2016) dubs the process "design-in-practice" (p. 31), which evolves as the researcher, or, in the case of AR/PAR, a research collective made up of multiple stakeholders, adjusts and refines due to unforeseen events, revealed complexities, or unexplored opportunities. It is important to emphasize that AR/PAR must be an iterative process and that with each iteration new understandings are generated that further inform the work as it moves forward. Indeed, this flexibility and iterative responsiveness are at the very heart of the methodological approach and its validity.

It is crucial that practitioners, such as school leaders, local policymakers, and novice researchers, understand that AR/PAR is not intended to be a simple or straightforward process to follow and will not look the same

every time one engages in it. Rather, a keen focus on the epistemological underpinnings of the process at each step of the action research process is critical for authentic engagement. Figuring out creative ways to bring together a collective that is both inclusive and representative; making learning explicit and folding new understandings back into the process; and taking action that is rooted in and speaks to the needs and expressed desires of a community should be paramount.

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY

One of the most foreboding questions is: why should school leaders or local policy makers engage in AR/PAR? What kinds of unique knowledge and understandings can school leaders and local policy makers gain from using AR/PAR in their schools and communities? And how can they work effectively with university-based researchers throughout the process? The illustrative study we have described throughout this chapter provides a useful example of how AR/PAR can be enacted to help school leaders and local policymakers access authentic knowledge of the individuals and groups they serve. The case also helps novice researchers understand the flexible, iterative process of AR/PAR as well as the potential it holds for conducting community-based research.

School leaders are bombarded with many different types of data, including enrollment information, achievement data, graduation rates, student demographics, poverty measures, and much more. This data can be very helpful in answering many questions, particularly those related to improving student achievement outcomes. But what if a school leader wants to improve the community in her school? What if the students want to eliminate large amounts of food waste? What if classroom teachers want to adopt grading practices that reflect the true learning of the child? What if students say racism, bullying, or safety is an issue and strive to develop a response? AR/PAR holds promise in these and many similar situations.

ESTABLISHING SHARED OWNERSHIP IN LEADERSHIP AND POLICY RESEARCH

When confronted with an issue like one of those mentioned above, school leaders and policymakers should first think about who should be at the proverbial table. Who has a stake in the problem and potential solution? To take the example of school cafeteria food waste, students should of

course be at the table, but who else? A school leader might partner with a representative from the local school board in charge of making decisions about nutrition. They could invite a representative of the Parent Teacher Association (PTA) to attend an initial meeting at which those stakeholders implicated in such an issue could start a process of investigation. They should probably also ask some of the lunch staff (school-based and central staff) to attend such a meeting. Teachers may want to play a role as well. They may also choose to invite a university-based researcher who focuses on nutrition and/or education to join the research collective as a methodological facilitator. The leader and policymaker should make these invitations to participate in an AR/PAR project purposefully to assure all who have a stake in the outcome are represented.

Once everyone implicated in the issue is part of a research collective, they can make collaborative decisions about what they want to achieve and how they want to achieve it. Perhaps they decide they need to collect data about what students prefer to eat. They may decide it is important to examine other school district nutrition models. Or they may hold a town hall meeting to garner a lot of feedback in a short amount of time. Whatever they decide, the participatory nature of the process is most important and is what will lead to change that speaks to the needs of those implicated.

ENGAGING IN COMMUNITY-BASED ANALYSIS IN LEADERSHIP AND POLICY RESEARCH

While the entire AR/PAR process should be based on stakeholders' experiences and needs, analysis can be a time in which communities feel a sense of buy-in because they are involved in a part of the research process typically reserved for "experts". Using a consecutive, "onion skin" analysis procedure (Curry 2012), school officials could examine the data, draw out initial themes and findings, and then share them with the collective's student, parent, teacher, lunchroom staff, and other members for confirmation or correction. Alternatively, the collective could use the "curtain" or "fishbowl" approach described by Curry, inserting their insights as school leaders analyze the data, perhaps by responding to guiding questions. To apply Dodson and Schmalzbauer's (2005) approach, the entire collective could engage in collaborative analysis, working through raw data to identify themes and apply interpretations. In any of these procedures, all participants have the opportunity to weigh in on the data collected, share their own insights, and influence the next steps.

ADOPTING AN ORIENTATION TOWARD ACTION IN LEADERSHIP AND POLICY RESEARCH

Once data is collected and analyzed, all stakeholders should be meaningfully involved in decisions about how to disseminate findings and what action(s) to take in relation to those findings. In this hypothetical project about school lunches, different parts of the collective may choose to take various forms of action. Perhaps a group of students decides to take action by advertising new school lunch options in exciting ways. A group of parents may disseminate key information to the PTA. Lunch staff may hold professional training meetings or create regular opportunities for students and parents to provide constructive feedback or suggestions. The school leader and school board member may work to make policy changes, depending on findings. Everyone can play a meaningful role and take action in important ways.

RECOMMENDATIONS FOR NOVICE RESEARCHERS: A REORIENTATION OF IDENTITY

In providing recommendations for novice researchers, we offer suggestions regarding the three points discussed throughout the chapter: shared ownership, community-based analysis, and orientation toward action.

Challenge 1: Negotiating New Roles in AR/PAR In our experience, one of the trickiest aspects of engaging in the shared ownership of AR/PAR is in the negotiation of roles (Call-Cummings 2017). As university-based researchers, we often take for granted ownership of the research process and products. We fill out Institutional Review Board (IRB) forms. We are listed as Principal Investigators (PIs). We author scholarly manuscripts alone or with other “scholarly” colleagues. We collect data and make decisions about how the data are interpreted and shared. In other words, we own research. Yet AR/PAR upends this seemingly rigid and supposedly necessary structure. As university-based researchers become part of a research collective, they/we must reorient their/our identities and relationships to research participants. For example, as a university-based researcher engaging with a particular community around a specific issue, I am most likely, by definition, an outsider because of my status as a university-based researcher. But AR/PAR requires that we work together with communities in such a way as to toggle between being an outsider

and an insider. This comes with ethical considerations as well as methodological quandaries that should be explicitly and collaboratively considered within the research collective.

This reorientation of our identity as researchers also relates to expertise. As university-based researchers, we are trained as experts and think of ourselves as experts. We are required to publish in scholarly journals that treat us and refer to us as experts. Students speak to us with deference. Media representatives ask us for our opinions as if the public should pay attention to us. AR/PAR breaks that hierarchy. It can be difficult for a novice—and a more senior—researcher to feel comfortable in the role of learner or co-producer of knowledge. It may feel to the researcher as if she is losing her authority.

Our first recommendation is for researchers who are interested in engaging in an AR/PAR approach to embrace the reorientation of identity and the fluidity of roles within the collective. Transparently negotiate roles with members of the research collective. Make this an explicit part of the process from beginning to end. Ask questions you may normally only journal about or discuss with a colleague, mentor, or advisor. By dealing with these issues up front you will be more likely to authentically share ownership of the process and products. Your work will gain validity because you will wrestle with power relations openly rather than allowing them to remain unquestioned and unchallenged, potentially undermining your findings (Call-Cummings 2017).

Challenge 2: Approaching Participants as Co-analysts Challenges also arise as we seek to engage in community-based analysis. This analytic approach requires a further reorientation of our identity. As university-based researchers, we are most likely experts in a particular approach to data analysis. As we engage in community-based analysis, however, our role becomes that of equal participant and, if necessary, facilitator of conversations and discussions that make meaning explicit. The protocol provided by Dodson and Schmalzbauer (2005) provides a helpful guide to begin this process.

The keys for successful community-based analysis are sincere humility and authentic engagement in learning about the lived experiences of those community researchers, as well as about the meanings they place on those experiences. As we reorient our identity to engage as humble learners and

facilitators, analysis will become a unique opportunity for consciousness raising for all those involved in the knowledge production process—including university-based researchers. Analysis carries with it the potential for access to understandings and knowledges otherwise inaccessible to “outsiders.”

Challenge 3: Recognizing Action The final challenge involves shifting our orientation as researchers toward action and recognizing what actions are appropriate given the capacity and context where the AR/PAR research occurs. This is central to AR/PAR and yet, in my own research in Idaho (Call-Cummings 2015), I have found it to be a stumbling block. I assumed that the “action” entailed in AR/PAR was supposed to be a particular type of action—a visible or tangible change; an inequitable policy struck down or an unfair practice stopped. At the conclusion of my 18-month study in Idaho, I was disappointed that no one appeared to be clamoring for change, the school had not established a formal process to address racism among teachers, and no school-wide anti-racism campaign had been initiated. I felt like the project had failed, or I had done something completely wrong during the process. However, upon reflection, I learned that in many instances it is the *process* of AR/PAR that is important and that can lead to a critical empowering of those involved. Rather than some monumental change in other students, practices, or policies, the change came to those who participated in the project—those members of the collective as well as those teachers who were interviewed and those who engaged with the student co-researchers in the dissemination of findings (school board members and other community leaders). Several of the students who had collaborated with Mrs. James and I reported after their project concluded that they felt inspired and were in a better position to stand up for themselves and their peers when they were in situations where they felt discriminated against. They felt more confident speaking with their White teachers and administrators and believed they had developed the skills to effectively articulate their feelings about racism at their school and in their community. Meanwhile, teachers became more aware of their own biases and presuppositions about Latino/a/x students and families. School board members may have felt freer to discuss racism as an important issue at this school and in this community rather than ignoring it or pretending it was not an issue. So, even though racism had not disappeared, the change came through and in the research process itself.

The recommendation for novice researchers is thus to think about action, change, and transformation more broadly than one might normally, and to take cues from the other members of the collective in terms of the actions they want and need to take and the change(s) they are seeking. Being transparent and explicit about one's own research agenda at the beginning of the process as one builds relationships with members of the collective will help in doing this.

CHAPTER SUMMARY

This chapter has presented an introduction to three critical aspects of action research and participatory action research for school leaders, local policy makers, and novice researchers: shared ownership, community-based analysis, and an orientation toward action. While AR/PAR includes much more than these three ingredients, and is more flexible and iterative than simply adding these three together and expecting a clean process to emerge, we believe that a strong and reflective emphasis on all three can help school leaders, local policymakers, and novice researchers as they engage in AR/PAR projects. Drawing on my research experience and on scholarly literature, we have included several examples of ways in which various stakeholders can reach toward these three aspects of AR/PAR, especially, but not exclusively, in relation to projects that focus on issues of social justice. Overall, this chapter illustrates that AR/PAR is a unique and accessible methodological approach and that school leaders, local policymakers, and novice researchers should feel confident in engaging in it with communities that seek change.

Recommended Readings

Fals Borda, O. & Rahman, M. A. (1991). *Action and knowledge: breaking the monopoly with participatory action-research*. New York: Apex.

Fals Borda and Rahman highlight applications of PAR to social change projects in the developing world, in which participants gain power over their circumstances through generating knowledge and devising local solutions to local problems.

Reason, P., & Bradbury, Hilary. (2006). *Handbook of action research: The concise paperback edition*. London; Thousand Oaks, Calif.: SAGE.

Reason and Bradbury provide an excellent introduction to the philosophical and theoretical groundings and development of AR/PAR methodology and the potential of collaborative knowledge production. Contributors, who

include Fals Borda, describe AR/PAR projects carried out in numerous national and professional contexts, all sharing a commitment to democratic, reflective research.

REFERENCES

- Berman, M. (1981). *The reenchantment of the world*. Ithaca: Cornell University Press.
- Bourke, L. (2009). Reflections on doing participatory research in health: Participation, method and power. *International Journal of Social Research Methodology*, 12(5), 457–474.
- Call-Cummings, M. (2015). Exploring the intersections and implications of pedagogies of empowerment and critical PAR through youth research on racism in schools (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses Global (3737223).
- Call-Cummings, M. (2017). Establishing communicative validity: Discovering theory through practice. *Qualitative Inquiry*, 23(3), 192–200.
- Cammarota, J., & Romero, A. (2011). Participatory action research for high school students: Transforming policy, practice, and the personal with social justice education. *Educational Policy*, 25(3), 488–506.
- Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge and action research*. London: Falmer.
- Curry, M. W. (2012). In pursuit of reciprocity: Researchers, teachers, and school reformers engaged in collaborative analysis of video records. *Theory Into Practice*, 51(2), 91–98.
- Dodson, L., & Schmalzbauer, L. (2005). Poor mothers and habits of hiding: Participatory methods in poverty research. *Journal of Marriage and Family*, 67(4), 949–959. <https://doi.org/10.2307/3600249>.
- Elliott, J., & Adelman, C. (1973). Supporting teachers' research in the classroom. *New Era*, 54(9), 210–215.
- Fals-Borda, O. (2001). Participatory (action) research in social theory: Origins and challenges. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: Participative inquiry and practice* (pp. 27–37). London: Sage.
- Fals Borda, O., & Rahman, M. A. (1991). *Action and knowledge: Breaking the monopoly with participatory action-research*. New York: Apex Press.
- Fine, M., & Torre, M. E. (2006). Intimate details: Participatory action research in prison. *Action Research*, 4(3), 253–269. <https://doi.org/10.1177/1476750306066801>.
- Fine, M., Torre, M. E., Boudin, K., Bowen, I., Clark, J., Hylton, D., & Upegui, D. (2004). Participatory action research: From within and beyond prison bars. In L. Weis & M. Fine (Eds.), *Working method: Research and social justice* (pp. 95–120). New York: Routledge.

- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings* (pp. 1972–1977). New York: Pantheon.
- Freire, P. (1970). *Pedagogy of the oppressed*. New York: Herder and Herder.
- Freire, P. (1974). *Education for critical consciousness*. New York: Seabury.
- Freire, P. (1982). Creating alternative research methods: Learning to do it by doing it. In B. Hall, A. Gillette, & R. Tandon (Eds.), *Creating knowledge: A monopoly? Participatory research in development* (pp. 29–37). New Delhi: Participatory Research Network Series 1.
- Gaventa, J. (1991). Toward a knowledge democracy: Viewpoints on participatory research in North America. In O. Fals-Borda & M. A. Rahman (Eds.), *Action and knowledge: Breaking the monopoly with participatory action-research* (pp. 121–131). New York: Apex Press.
- Glassman, M., Erdem, G., & Bartholomew, M. (2013). Action research and its history as an adult education movement for social change. *Adult Education Quarterly*, 63(3), 272–288. <https://doi.org/10.1177/0741713612471418>.
- Habermas, J. (1979). *Legitimation crisis in the modern society. Communication and the evolution of society*. Boston: Beacon Press.
- Kemmis, S., & McTaggart, R. (2005). Communicative action and the public sphere. *The Sage Handbook of Qualitative Research*, 3, 559–603.
- Knapp, M. S. (2016). The practice of designing qualitative research on educational leadership: Notes for emerging scholars and practitioner-scholars. *Journal of Research on Leadership Education*, 12(1), 26–50.
- Levin, M. (1999). Action research paradigms. In D. J. Greenwood (Ed.), *Dialogues on work and innovation: Action research: From practice to writing in an international action research development program* (pp. 25–38). Philadelphia: John Benjamins Publishing Company.
- Lind, C. (2008). Knowledge development with adolescents in a PAR process. *Educational Action Research*, 16(2), 221–233.
- Ponic, P., Reid, C., & Frisby, W. (2010). Cultivating the power of partnerships in feminist participatory action research in women's health. *Nursing Inquiry*, 17(4), 324–335. <https://doi.org/10.1111/j.1440-1800.2010.00506.x>.
- Reinharz, S. (1979/1984). *On becoming a social scientist: From survey research and participant observation to experiential analysis*. New Brunswick: Transaction Books.
- Santos, D. (2015). (Participatory) action research and the political realm. In B. Dennis, L. Carspecken, & P. F. Carspecken (Eds.), *Qualitative research. A reader in philosophy, core concepts, and practice* (pp. 492–514). New York: Peter Lang.
- Sohng, S. (1996). Participatory research and community organizing. *The Journal of Sociology & Social Welfare*, 23(4), 77–98.
- Stark, J. L. (2014). The potential of Deweyan-inspired action research. *Education and Culture*, 30(2), 87–101. <https://doi.org/10.1353/eac.2014.0013>.

- Stoudt, B. G. (2007). The co-construction of knowledge in “safe spaces”: Reflecting on politics and power in participatory action research. *Children Youth and Environments*, 17(2), 280–297.
- Wang, C. (1999). Photovoice: A participatory action research strategy applied to women’s health. *Journal of Women’s Health*, 8(2), 185–192. <https://doi.org/10.1089/jwh.1999.8.185>.
- Winter, R. (1998). Finding a voice – Thinking with others: A conception of action research. *Educational Action Research*, 6(1), 53–68. <https://doi.org/10.1080/09650799800200052>.

PART II

Quantitative Approaches to the Study of Educational Leadership and Policy

Part II of this volume presents quantitative approaches to the study of educational leadership and policy. Specifically, this part of the volume addresses secondary data analysis, matching strategies for causal inference, quantitative methods suitable for studying curriculum, value-added and growth models, Social Network Analysis, fiscal research strategies, and cost estimation.



Secondary Data Analysis in the Field of Educational Leadership and Policy Studies

Angela Urick

The use of large-scale, secondary data with representative samples allows researchers to draw generalizable conclusions for research questions related to educational leadership and policy issues. Most common to this field, administrative school or district data address local issues or give examples of practice. Further, qualitative data answer more in-depth questions that address the what, how, and why. In contrast, established national and international data sets can be used to analyze trends across the country, or countries. This type of analysis can better inform state and federal policy, extend research on commonalities or differences across systems of education, and provide a means for advancing quantitative analysis techniques.

There are several benefits to the secondary data analysis approach. First, large organizations, agencies and/or centers collect data and, for the most part, have it available in a format that is ready to use. With many sources and data sets available, a researcher will likely find an existing study and survey designed to assess one or more target research interests. A scholar

A. Urick (✉)
Educational Leadership & Policy Studies, University of Oklahoma,
Norman, OK, USA
e-mail: urick@ou.edu

can then spend more time on study conceptualization and data analysis to rigorously address state, national, or international issues with generalizability. Second, established national and international data sets have large sample sizes and several administrations over time. With sufficient power and multiple time points across samples, there are many options for statistical analysis. Third, common data sets provide an opportunity for scholars to replicate research. Scholars can refer to past studies which analyze their data set of interest to understand how variables have been used, and their results, which guides future study. This potential for replication supports emerging researchers with a clearer direction of how to use the data, as well as with a clearer comparison of what has and has not been studied.

This chapter focuses on six main data sets: three national and three international (refer to Table 8.1). These surveys serve different purposes as evidenced by the research framework themes, which detail the intended design of the survey by sponsoring organizations (see right column, Table 8.1). National Assessment of Educational Progress (NAEP), Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) are each designed with a different intent, but all include student achievement scores. NAEP seeks to measure the progress of academic literacy in grades 4, 8, and 12 with the nation, states, and districts compared across time. With similar goals, TIMSS measures formal learning in grade 4 and 8 mathematics and science with varying degrees of cognitive skill (i.e. knowledge, application, reasoning), but TIMSS also allows for cross-national comparison. Whereas PISA is noted for the assessment of knowledge applied to everyday life for a sample based on age 15, rather than grade. Each of these data sets has background and perception questionnaires which connect home and school environments to these unique academic assessments. High School Longitudinal Study of 2009 (HSLs) also includes baseline and follow-up measures of academic achievement. However, the purpose of HSLs is to follow a cohort of ninth graders in 2009 through high school and beyond. This data set is ideal for the study of postsecondary transitions and success as well as to connect high school to career aspirations and training. This is the newest cohort study. There are previous waves of cohort studies available through National Center for Education Statistics, or NCES, (i.e. Educational Longitudinal Study of 2002). Finally, Schools and Staffing Survey (SASS) (U.S. data) and Teaching and Learning International Survey (TALIS) (international data) both focus on the work conditions of teachers and principals in schools. SASS is best known as a source to track

Table 8.1 Main surveys for educational leadership and policy secondary data analysis

<i>Survey</i>	<i>Years</i>	<i>Organization</i>	<i>Population</i>	<i>Data</i>	<i>Research framework themes</i>
Schools and Staffing Survey (SASS) https://nces.ed.gov/surveys/sass/	88, 91, 94, 00, 04, 08, 12	National Center for Education Statistics (NCES)	U.S. public schools (traditional, charter, military and special purpose); private schools; teachers provided by sampled schools	District, school, principal, teacher and library media center questionnaires	School quality, school management, school choice, teacher quality (see Cox et al. 2017)
National Assessment of Educational Progress (NAEP) https://nces.ed.gov/nationsreportcard/	69–17	National Center for Education Statistics (NCES)	U.S.: public and private schools selected, then students	Student, school, students with disabilities/English language learners, teacher, National Indian Education Study, long-term trend student questionnaires; Grade level subject assessments;	Opportunities to learn in and out of classroom, educational experiences, teacher training and instructional practices, school policies and characteristics, assessment of academic literacy (see NCES 2009)
High School Longitudinal Study of 2009 (HSLs) https://nces.ed.gov/surveys/hsls09/index.asp	09, 12, 13, 14, 17	National Center for Education Statistics (NCES)	U.S. public and private high schools, then grade 9 students in base year of study	High school transcript study Student, parent, teacher, school counselor and school administrator questionnaires; Only relevant questionnaires collected for subsequent follow ups	Math and science education, the changing environment of high school, postsecondary education (see Ingels et al. 2011)

(continued)

Table 8.1 (continued)

<i>Survey</i>	<i>Years</i>	<i>Organization</i>	<i>Population</i>	<i>Data</i>	<i>Research framework themes</i>
Teaching and Learning International Survey (TALIS) http://www.oecd.org/edu/school/talis.htm	08, 13, 18	Organisation for Economic Co-operation and Development (OECD)	Around 24 to 45 countries which includes United States in 2013 and 2018	Principal and teacher questionnaires	Teacher education, school leadership, teacher appraisal and feedback, school climate and ethos, teachers' pedagogical beliefs and practices (see Rutkowski et al. 2013)
Programme for International Student Assessment (PISA) http://www.oecd.org/pisa/ ; https://nces.ed.gov/surveys/pisa/	00, 03, 06, 09, 12, 15, 18	Organisation for Economic Co-operation and Development (OECD)	Around 88 countries or regions across 2000–2018; schools selected then 15-year-old students	Student, school, parent and teacher questionnaires; Education career and information and communication technology familiarity questionnaires for students; Science, reading, math and financial literacy assessments	Non-cognitive outcomes, teaching and learning processes, school policies and governance, student background, application of academic knowledge (see Vayssettes et al. 2016)
Trends in International Mathematics and Science Study (TIMSS) https://timssandpirls.bc.edu/ ; https://nces.ed.gov/timss/	95, 99, 03, 07, 11, 15	International Association for the Evaluation of Educational Achievement (IEA)	Around 49 countries; schools selected then 4 and 8 grade students	Student, home, teacher, school and country curriculum questionnaires; Math and science assessments	Home contexts, student characteristics and attitudes toward learning, principal leadership, resources, instruction, teacher preparation and experience; assessment of formal learning with degrees of cognitive skill (see Mullis and Martin 2013)

teacher turnover trends, and includes school leadership measures. TALIS includes school leadership measures, as well as measures around instruction. Further, TALIS can be linked to the student data available in PISA. Particularly, these data sets, and this chapter primarily focus on K-12 education rather than higher education.

Instead of a “how to” guide, the purpose of this chapter is to provide novice researchers with enough information about secondary data analysis to decide whether it is a fit for their research interests. Likely, the most valuable information in this chapter is the collection of references provided in an organized way. Within the narrative and tables, a reader can identify additional articles and texts that will support the development knowledge and skills needed to apply this method beyond this chapter. Due to the vastness of available data sets and possible statistical analyses to apply to them, it would be difficult to supply the reader with a singular, introductory and relevant example. Rather, the literature review identifies research topics which can be studied with these six data sets. An application section describes which statistical procedures a researcher would want to learn to use with this data, and finally, the recommendations detail the nature of future work for this approach. By the end of this chapter, a novice researcher should be able to decide if secondary data analysis aligns with their research agenda as well as their desired methodological skill set.

REVIEW OF RELEVANT LITERATURE

This review of literature broadly demonstrates what has been studied using these six national and international surveys in the field of educational leadership and policy studies. Organizations sponsoring these data sets create research frameworks to describe the extent and nature of variables included, which are often found in user guides. However, researchers can apply new perspectives, theories, and concepts to existing items. In addition, intended research themes can be left unexplored, which gives direction for new contributions. However, rather than compare new perspectives to these intended research frameworks and identify unexplored areas—which would be difficult to do in one review—this review is a synthesis of what has been studied and how content in these data sets might overlap. This synthesis supplies the reader with a list of potential research topics suitable for secondary data analysis across these main surveys in order to stimulate ideas about how to link or compare surveys in future use. This review includes broad topics of study connected to these six data

sets since 2010. While it is comprehensive, it is not exhaustive. Across these six main data sets, scholars have investigated teachers, principals, and students as individuals, the contexts of classrooms and schools, postsecondary transitions, and national policy issues.

Teachers, principals, and students have different backgrounds, roles, beliefs, and attitudes, which make them individual actors within the organization and/or policy. For these reasons, many questionnaires are designed to measure their individual contributions or enactments within the education system. For example, teachers are hired into positions with varying qualifications (see Lee 2012a; Meroni et al. 2015) and receive different pay (Woessmann 2011). Within the position, teachers have unique attitudes toward their commitment and satisfaction (Price 2012; Reback et al. 2014), unique perceptions about their efficacy (Conley and You 2017; Ware and Kitsantas 2011), and distinct needs for professional development (Phillips et al. 2011). Further, teachers experience their position based on race and gender, among other characteristics (Farinde-Wu and Fitchett 2018), and the conditions in their schools (Grissom et al. 2014; Ni 2012). Altogether, background, beliefs, and attitudes have predicted their decisions to stay in a school or the profession (Grissom and Keiser 2011; Urick 2016b).

Similarly, principals experience their role based on their background (Shen et al. 2012) and perceived work conditions (Sun and Ni 2016). These characteristics help to explain or moderate their attitudes toward commitment and satisfaction (Boyce and Bowers 2016; Price 2012) as well as their own efficacy (Ware and Kitsantas 2011), and how teachers may view them as effective (Grissom and Harrington 2010). Principal background, beliefs, and attitudes also contribute to their decisions about whether to remain as a principal in a school or in the profession (Boyce and Bowers 2016; Sun and Ni 2016; see Table 8.2). Likewise, students experience schools and learning based on their background: race (Anderson 2016), gender (Pope and Sydnor 2010), and their parents and home environment (Noble and Morton 2013; Froiland and Davison 2016; Tsai et al. 2017). Students have individual views of their own academic self-concept (Marsh 2016), motivation, and interest (Yu and Singh 2018; Ainley and Ainley 2011), which relates to their sense of belonging (Akgul et al. 2016), persistence (Ashford et al. 2016), and resilience (Sandoval-Hernández and Bialowolski 2016; see Table 8.3).

These individual backgrounds, attitudes and beliefs of teachers, principals, and students—studied separately or as interactions—help to explain

Table 8.2 Teachers and principals studied across six main surveys

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
Teachers	Teacher satisfaction	SASS, TFS	Price (2012), Tickle et al. (2011)
	Intent to stay/ attrition risk	SASS	Grissom et al. (2014)
	Teacher commitment	SASS, ECLS	Reback et al. (2014)
	Teacher efficacy, collective efficacy	SASS, TALIS	Ware and Kitsantas (2011), Conley and You (2017) and Gálvez et al. (2017)
	Teacher cognitive self-regulation	PISA	Mattern and Bauer (2014)
	Teacher turnover	SASS, TFS	Grissom and Keiser (2011). Urick (2016b)
	Teacher compensation	SASS, census, PISA	Goldhaber et al. (2010), Woessmann (2011)
	Teacher hiring	SASS	Engel (2012)
	Teacher qualifications	SASS, TFS, NAEP, F-33, TMSS, PISA, PIAAC	Lee (2012a), Ronfeldt et al. (2014), Woo and Henfield (2016) and Meroni et al. (2015)
	Teacher professional development	SASS, Education Week's Quality Counts, American Federation of Teachers' "Making Standards Matter," Fordham Foundation's "State of State Standards" reports	Phillips et al. (2011), Wieczorek (2017)
	Teacher work conditions/ attitudes	SASS, TALIS	Ni (2012), Grissom et al. (2014) and Duyar et al. (2013)
	Black female teachers	SASS	Farinde-Wu and Fitchett (2018)
Principals	Principal turnover	SASS, PFS	Boyce and Bowers (2016), Sun and Ni (2016)
	Principal background	SASS	Shen et al. (2012)
	Principal satisfaction	SASS, PFS	Price (2012), Boyce and Bowers (2016)

(continued)

Table 8.2 (continued)

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
	Principal professional development	SASS, TALIS	Grissom and Harrington (2010), Gumus and Bellibas (2016)
	Principal efficacy	SASS	Ware and Kitsantas (2011)
	Principal commitment	SASS	Price (2012)
	Principal attitude	SASS, PFS	Boyce and Bowers (2016)
	Principal effectiveness	SASS	Grissom and Harrington (2010), Sun and Ni (2016)
	Principal work conditions	SASS	Sun and Ni (2016)

Notes: *TFS* Teacher Follow-up Survey, *ECLS* Early Childhood Longitudinal Survey, *F-33* School District Finance Survey, *PIAAC* Programme for the International Assessment of Adult Competencies, *PFS* Principal Follow up Survey

the variance within organizations, and how learning is supported or experienced by each actor. While these roles can be studied separately, when studied together, they bring a more complete and accurate measure of perceptions in a school. For instance, using SASS, Urick (2016b) investigated the influence of both teacher and principal perceptions of leadership on teacher retention, and found different types of perceptions of leadership for both, which predicted whether a teacher stayed in their current school. Since schools are social organizations where different perceptions, experiences and beliefs merge, an understanding of the convergence or divergence of these interests is important (further, see Goldring et al. 2015).

While individuals have their own experiences and beliefs, they overlap within common school and classroom contexts within the school system. Questionnaires often ask respondents not only about themselves but also about the community, practices, or structures that are shared with others. Teachers and principals respond to the degree of teacher classroom autonomy, practices of school leadership, teacher collaboration, and teacher-principal relationships within the school (Duyar et al. 2013; Liu et al.

Table 8.3 Students studied across six main surveys

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
Students	Gender differences in math	NAEP	Pope and Sydnor (2010)
	Academic achievements of Black students	HSLs	Anderson (2016), Young (2017)
	Home influence	NAEP, American community survey	Dunst and Hamby (2016)
	Home educational resources	TIMSS; PISA	Azina and Halimah (2012), Tsai et al. (2017)
	Parental involvement	HSLs; PISA	Froiland and Davison (2016), Tramonte and Willms (2010)
	African American parental influence	NAEP	Noble and Morton (2013)
	Student motivation	HSLs, TIMSS, PISA	Yu and Singh (2018), Liou (2017), Marsh et al. (2013) and Guzel and Berberoglu (2010)
	Interest development	PISA	Ainley and Ainley (2011)
	Academic self-concept	TIMSS, PISA	Min et al. (2016), Marsh (2016)
	Self-efficacy	PISA	Wu (2016), Akgul et al. (2016)
	Sense of belonging	PISA	Akgul et al. (2016)
	Student perception of learning participation	TIMSS	Chen et al. (2012)
	Big-fish-little-pond effect	TIMSS, PISA	Marsh et al. (2014), Nagengast and Marsh (2012)
	STEM persistence	HSLs	Andersen and Ward (2014), Ashford et al. (2016)
	Academic resilience	TIMSS	Sandoval-Hernández and Bialowolski (2016)
	Math course enrollment	HSLs	Froiland and Davison (2016)

2016; Oberfield 2016; Price 2012; Urick 2016a, b). They also perceive the disciplinary climate or student behavior within the classroom and across the school (Jenkins and Ueno 2017; Curran 2016). Teachers provide information about their classroom resources, instructional practices, and other teaching-learning processes with students (Akyuz and Berberoglu 2010; Dee et al. 2013; Lambert et al. 2015). These school and classroom contexts include the relationships within them, and also the characteristics of groups or structures like class size (Li and Konstantopoulos 2017), socioeconomic composition of the classroom (Chudgar et al. 2012),

or race of those served in the school community (Kelly 2010). These school and classroom context items, or constructs provide measures of the common practices and environment shared by the individuals within it (refer to Table 8.4). Many of these processes and characteristics of the school and classroom can be tested together to build frameworks of the systems through which schools operate and improve (see Urick et al. 2018, as example).

Table 8.4 Classrooms and schools studied across six main surveys

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
Classroom	Teacher perception of classroom resources and demands	SASS	Lambert et al. (2015)
	Teacher classroom autonomy	SASS	Oberfield (2016)
	Classroom resources	SASS, F-33	Dee et al. (2013)
	Instructional time	SASS, F-33, NAEP, ECLS	Fitchett et al. (2014), Dee et al. (2013), Blank (2013) and Reback et al. (2014)
	Classroom disciplinary climate	TALIS	Jenkins and Ueno (2017)
	Instructional practices	TALIS, TIMSS, PISA, NAEP	Shi et al. (2014), Charalambous and Kyriakides (2017), Rivkin and Schiman (2015) Jiang and McComas (2015) and Bittman and Russell (2016)
	Teaching-learning process	TIMSS	Akyuz and Berberoglu (2010)
	Class size	TIMSS	Li and Konstantopoulos (2017)
	Classroom composition of socioeconomic status	TIMSS	Chudgar et al. (2012)
School	School processes	SASS	Shen et al. (2012), Farinde-Wu and Fitchett (2018)
	School resources and facilities	TIMSS	Afana et al. (2013), Hopland (2013)
	Work environment/ conditions	SASS, TFS, TIMSS	Grissom (2011), Ingersoll and May (2012) and Winnaar et al. (2015)

(continued)

Table 8.4 (continued)

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
	Teacher Cohesion and Interaction	SASS	Price and Collett (2012)
	Teacher sense of community	SASS	Weathers (2011)
	Teacher collaboration	TALIS	Duyar et al. (2013)
	School climate	SASS, PFS, F-33, HSLS	Dee et al. (2013), Boyce and Bowers (2016) and Froiland et al. (2016)
	Teacher and principal perceptions of student behaviors	SASS, U.S. Dept. of Ed. OCR data, Chicago School Study and Partnership for Literacy Study	Curran (2016), Kelly (2010)
	Administrator support	SASS	Tickle et al. (2011), Conley and You (2017)
	Principal influence	SASS	Boyce and Bowers (2016)
	Leadership types, styles, and practices	SASS, TALIS	Weathers (2011), Urick (2016a, b), Liu et al. (2016) and Bellibas and Liu (2018)
	Principal-teacher relationship	SASS	Price (2012)
	Teacher leadership/autonomy	SASS	Fitchett et al. (2014), Xie and Shen (2013)
	School innovation	SASS	Preston et al. (2012)
	Organizational quality	SASS	Jackson and Marriott (2012)
	School quality	TIMSS, World Bank, International Monetary Fund	Bouhlila (2017)
	School effectiveness	TIMSS	Chen (2014)
	Production function, school inputs, outputs	PISA	Woessmann (2016), Giménez et al. (2017)
	Predominately black schools	SASS, Chicago School Study and Partnership for Literacy Study	Kelly (2010)

Notes: *F-33* School District Finance Survey, *ECLS* Early Childhood Longitudinal Survey, *TFS* Teacher Follow-up Survey, *PFS* Principal Follow up Survey

These measures of individuals and their school and classroom contexts extend beyond research topics immediately relevant to practice or leadership in K-12 schools. Because of the generalizability of samples, and how K-12 schools are situated within a larger system of national education, these measures also address research questions related to postsecondary transitions, as well as pressing policy issues. Scholars have studied college supports for K-12 students' academic readiness and planning for college financing (Hillman et al. 2015; Hurwitz and Howell 2014; Schneider and Saw 2016; refer to Table 8.5). While the main data sets in this chapter are situated within K-12 schools, scholars have linked these to others focused on higher education or adult learning, which are noted in Tables 8.2 and 8.6 (see Integrated Postsecondary Educational Data System, IPEDS, Programme for the International Assessment of Adult Competencies, PIAAC, and Baccalaureate and Beyond Survey, B&B). However, an overarching purpose of national and international data is to assess country trends and policies. For example, scholars have studied how students are sorted across the education system based on race, socioeconomic status, language, and/or immigrant status (Fairchild et al. 2012; Spees et al. 2016), access to education resources (Lee 2012a), and opportunities to learn (Schmidt et al. 2015), as well as achievement trajectories (Lee 2010). Policies specific to the United States have been assessed such as No Child Left Behind (Dee et al. 2013) and Common Core State Standards (Porter et al. 2011), as well as those that compare similar issues or reforms across countries, such as grade retention (Goos et al. 2012), pre-school education (Pholphirul 2016), or the underrepresentation of women (Stoet and Geary 2015). Ties between the education system and the economy have also been studied (Condron 2011; see Table 8.6).

Table 8.5 Post-secondary transitions studied across six main surveys

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
Postsecondary transitions	College support for k-12 students	SASS	Hurwitz and Howell (2014)
	Gaps in math for college readiness	NAEP, TIMSS, ECLS	Lee (2012b)
	Race and college readiness/enrollment	HSLs	Alvarado and An (2015), Schneider and Saw (2016)
	College financing	HSLs	Hillman et al. (2015)

Notes: ECLS Early Childhood Longitudinal Survey

Table 8.6 National policy issues studied across six main surveys

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
National policy issues	Accountability / testing / no child left behind	SASS, F-33, ECLS, NAEP, PISA, World Bank	Dee et al. (2013), Reback et al. (2014), Fitchett et al. (2014), Wieczorek (2017), Lee and Reeves (2012) and Smith (2017)
	Charter schools	SASS	Preston et al. (2012), Sun and Ni (2016)
	Teacher supply or shortage	SASS, TFS, IPEDS, B&B	Ingersoll and Perda (2010)
	Alternative schools and special education teacher qualifications	SASS	Mason-Williams and Gagnon (2017)
	Private schools and teacher compensation	SASS	Goldhaber et al. (2010)
	Racial composition: Students, teachers, principals	SASS	Grissom and Keiser (2011), Renzulli et al. (2011) and Fairchild et al. (2012)
	Common Core state standards	NAEP, State Standards and Assessment Database, TIMSS	Porter et al. (2011), Schmidt and Houang (2012)
	Zero tolerance policies	SASS, U.S. Dept. of Ed. OCR data	Curran (2016)
	Age and grade	TIMSS	Cliffordson (2010)
	Grade retention	PISA, TIMSS	Goos et al. (2012)
	Pre-school education effects	PISA	Pholphirul (2016)
	Gender equity and sex differences in performance	PISA	Stoet and Geary (2015)
	Religiosity in nation, sex differences, performance	PISA, World Values Survey, TIMSS, World Factbook, Education at a Glance Reports	Stoet and Geary (2017)
	Language policies	NAEP, PISA, PIRLS	McEneaney et al. (2014), Arya et al. (2016)
Immigrant status and achievement	NAEP; HSLs, PISA, TIMSS	Spees et al. (2016), Bozick et al. (2016)	

(continued)

Table 8.6 (continued)

<i>Category</i>	<i>Research concepts</i>	<i>Surveys</i>	<i>Example of authors</i>
	Long-term academic literacy	PISA, PIAAC	Gustafsson (2016)
	Achievement trajectories	NAEP, ECLS, NELS, HS&B, ELS, NLS	Lee (2010)
	Equity and adequacy resource gaps	NAEP, F-33, SASS	Lee (2012a)
	National Education and the economy	PISA, TIMSS, Fraser Institute, Heritage Foundation	Condron (2011), Rindermann and Thompson (2011)
	Environment, wealth and achievement	PISA, bio-temperature calculations, World Bank	He et al. (2017)
	Socioeconomic status composition and mediators	TIMSS, PISA	Chudgar et al. (2012), Schmidt et al. (2015)
	Opportunity to learn	TIMSS, PISA, World Bank	Carnoy et al. (2016), Schmidt et al. (2015)

Notes: *F-33* School District Finance Survey, *ECLS* Early Childhood Longitudinal Survey, *PIRLS* Progress in International Reading Literacy Study, *PIAAC* Programme for the International Assessment of Adult Competencies, *NELS* National Education Longitudinal Study, *HS&B* High School and Beyond, *ELS* Educational Longitudinal Study of 2002, *NLS* National Longitudinal Surveys, *TFS* Teacher Follow-up Survey, *IPEDS* Integrated Postsecondary Educational Data System, *B&B* Baccalaureate and Beyond Survey

The study of these postsecondary transitions and national policies demonstrate how the context of K-12 schools fit within a larger societal system and purpose. For instance, Grissom et al. (2014) tested how teachers' perceptions of their work environment changed during a period of No Child Left Behind legislation. Further, Condron (2011) explains, with a cross-national comparison, that the association between economic inequality and low academic achievement might be due to the extent of egalitarianism and of economic inequality across racial/ethnic groups. Through comparisons to other policy contexts across countries, data collection during policy implementation, or connections to long-term workforce or adult outcomes, large-scale data sets allow researchers to situate findings within politics/policy, economics, and an understanding of the overall purpose of education.

For this review, a broad search was conducted for studies that applied secondary data analysis using these six main surveys since 2010. While it is not exhaustive due to the volume of publications, most research topics relevant to educational leadership and policy are represented in Tables 8.2, 8.3, 8.4, 8.5 and 8.6. From the literature searched, a few suggestions can be made for future research. First, the largest number of articles on topics of principals and school leadership used SASS data. Similar principal and leadership measures are available in TALIS, TIMSS, and PISA data sets but are not used as often. In fact, comparatively, there were few articles that used TALIS, which is most like SASS but with an international sample. Further, scholars who publish with TIMSS and PISA are more interested in issues of instruction, curriculum, and achievement, and often do not connect leadership or organizational concepts. Second, NAEP is underutilized. The sampling structure for NAEP allows for an analysis of states and districts. There are longitudinal data as well as questionnaires that have not been used to their capacity. Most studies that used NAEP found in this search were descriptive or only used test scores. Finally, scholars across different fields—educational leadership, education policy, public policy, educational psychology, teacher education, curriculum studies, math/science education, and economics—were using similar and/or related items, but discussions were disconnected across this literature. Scholars would benefit from reviews of interdisciplinary research on a select data set.

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY

Like most quantitative research methods, the application of secondary data analysis is dependent on how the research question aligns with available data. Scholars who use large-scale data rely on the existing body of research to understand which items to include as variables and how these variables fit together. From the nature of the research question, either correlational or causal, and the known relationship among the variables in the literature, a theory is formulated and tested with an appropriate statistical analysis.

While some research studies test hypotheses without an overarching framework, theoretical frameworks add direction and understanding to the use of large-scale data. Sponsoring organizations of data sets provide research frameworks which explain what measures were collected and suggest how they can be used, but a theoretical framework, which guides a research study of secondary data use, explains why and how select variables

answer a research question and extend knowledge. For example, scholars have applied expectancy value theory to formulate testable models to show how student motivation influences science learning (Liou 2017), or theories of organizational quality to demonstrate the importance of teacher-principal interactions (Jackson and Marriott 2012; Price 2012). Scholars have used secondary data to test and extend theory about the overlap and relationships among leadership styles and practices (e.g. Bellibas and Liu 2018; Liu et al. 2016; Urick 2016a, b). These applicable theoretical frameworks are interdisciplinary and stem from psychology, leadership and organizational theory, and sociology, as well as economics with frameworks of school quality, efficiency, equity, and production (e.g. Giménez et al. 2017; Lee 2012a; Woessmann 2016). Past literature, available measures, and a theoretical framework provide a map for which statistical analysis will answer a research question.

There are two main quantitative design approaches for secondary data analysis: (1) correlational, which tests the relationships among variables, or (2) causal, which tests the efficacy or impact of a “treatment.” Since the data are already collected, and there is no control over the assignment of participants to treatment or control groups, quasi-experimental designs are followed to estimate causal inference. Both correlational and causal designs have basic and advanced statistical methods which can be applied to secondary data.

A vast number of articles that have applied secondary data analysis in educational leadership and policy use some form of regression. For a basic approach, scholars apply regression from bivariate to multivariate approaches. Descriptive studies have used bivariate regressions or correlations to demonstrate simple relationships. Inferential studies extend these descriptive models with multivariate regression, nonparametric or parametric, and sometimes with a stepwise test of competing models (for statistical procedures, see Howell 2017; Mertler and Vannatta Reinhart 2017; e.g. Bellibas and Liu 2018; Ingersoll and May 2012). If a scholar is trained in multiple regression, this can be expanded to more advanced methods of theory testing with path analysis and structural equation models (for statistical procedures, see Bowen and Guo 2012; Kaplan 2009; e.g. Chen et al. 2012). Often, scholars apply exploratory or confirmatory factor analysis as a separate study or in preparation of some type of regression analysis (e.g. Marsh et al. 2013).

While sponsoring organizations of data sets publish reports and manuals with the psychometric properties of items and composites, scholars

may want to test the way in which these existing items measure new theories or constructs, which can be explored with factor analysis, and then used to build structural equation models. However, the sampling frames of most data sets have a nested structure, for example, teachers in schools, or students in classrooms. Because of this nested structure, or levels in unit of analysis, if a scholar wants to simultaneously test the effects of multiple, nested units (e.g. teachers nested in schools, or students nested in classrooms) on an outcome, a multilevel analysis is applied. In these cases, researchers use multilevel regression, also called hierarchical linear models (see Raudenbush and Bryk 2002), which can be extended to multilevel path analysis, multilevel factor analysis, and multilevel structural equation models (see Harring et al. 2015; Heck and Thomas 2015; e.g. Tsai et al. 2017).

If the purpose of a study is to estimate causal inference, a researcher can apply a quasi-experimental design which uses measures to account for non-random selection or assignment to treatment versus control, as well as possible covariates in the relationship of the treatment to the outcome. Rather than test how variables are related as in correlational design, quasi-experimental researchers present a clear logic for what variables matter and why, and try to present evidence which negates possible selection or assignment bias and missing covariates (for procedures, see Shadish et al. 2002). Some statistical techniques include propensity score matching (e.g. Ni 2012) to remove selection or assignment bias and regression discontinuity (e.g. Hurwitz and Howell 2014) or comparative interrupted time series (e.g. Dee et al. 2013) to test the change in outcome before and after treatment using multiple time points. Quasi-experimental design prompts researchers to think carefully about reasons for the inclusion of variables, as well as extent of bias and error in models. Because of the large-scale, representative samples, secondary data analysis offers a unique opportunity to test the impact of policy, reforms, and implementation of instructional and leadership practices.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

Scholars who are new to large-scale secondary data have three important areas to learn. First, sampling procedures often incorporate multiple characteristics as strata, then calculate proportions of strata for the sample to represent population, and nest stages of selection, such as school then participant. All components of these sampling frames must remain intact to generalize to the

population. Thus, researchers need to understand the application of sampling weights and missing data analyses which adjust collected data back to the intended sampling frame. Second, organizations that administer these assessments forgo a design process in which frameworks are created to guide the kinds of research topics or content that the items measure. They provide technical reports on the frameworks used to guide the intended content in surveys and assessments, and on the statistical results of the extent that items measure a concept or skill. Finally, a new secondary data analysis researcher must learn how to access each data set, and each year of each data set. Some data are publicly available through online downloads, and other data are restricted, requiring a formal request to use. There are different ways in which data are downloaded. For example, PISA offers a “manual” download in which a researcher saves a zip folder of multiple raw data files on their computer, and downloads and edits syntax for SPSS software, or other package, to extract data. To avoid edits of syntax and data merges which might compromise sampling frame structure, some organizations—like IEA for TIMSS—offer a free software download which helps a researcher construct data sets in their chosen statistical software package, and even provides access to run basic statistics to explore available data before creating formatted files.

Data download is the first step, next a researcher must understand data structure, data labels, composite variables, and how to prepare for an analysis. This is specific to each data set and learned with practice. For these reasons, learn one data set, then expand to other years of the same data set, and next other data sets. Once the structure and procedures of most national and international data sets are understood, then a scholar can think about how to link multiple data sets across years to answer substantial, nuanced, national leadership and policy questions.

Scholars who are advanced users of large-scale data and statistical techniques can work to extend the application of secondary data analysis. Scholars have begun to create their own data sets from funded research projects to share widely with other researchers, for example, the National Center for Teacher Effectiveness at Harvard University. Further, with technology and large information systems, data mining from websites is another means to create a secondary data set. The availability and collection of “big data” widens the scope of data analysis. Scholars have started to include intensive longitudinal data with numerous time points for each participant in statistical models (see Asparouhov et al. 2017). Because of the size and structure of this longitudinal data, the quantitative methods mentioned above—correlational, like structural equation models, and

causal, like time series analysis—have been extended to these “dynamic” versions. New dynamic statistical analyses can account for a larger number of time points, and potentially smaller samples of participants. Secondary data analysis makes it possible to apply new statistical theory and models.

CHAPTER SUMMARY

The National Center for Education Statistics (NCES), as well as international organizations, administer large-scale surveys and assessments which can be used for secondary data analysis. Existing, generalizable data allow researchers to address pressing national and systemic issues in educational leadership and policy. These common data sets provide an opportunity for an easier comparison and replication of findings in the literature across fields. Due to the large scale of the data sets, there are more options for advanced statistical analysis. Once researchers learn how to analyze one of these main data sets, this knowledge can be applied across additional years of the same data set, extended to other available national/international surveys, and can provide an example of how future original data collections may be shared with other researchers.

Recommended Readings

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issue in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151.

These authors explain common issues in the analysis and reporting of results of international data sets including TIMSS and PISA. They provide further discussion and resources on sampling, weights and estimation of test scores. This article presents key analysis issues which are important to understand in order to use these data appropriately.

Schneider, B., Saw, G., & Broda, M. (2016). A future for the National Education Longitudinal Program. *AERA Open*, 2(2), 1–13.

This article reviews the history of surveys administered by the U.S. National Center for Education Statistics. Schneider, Saw, and Broda provide readers with an understanding of how research using these federal data sets has informed policy, as well as suggested areas of growth for NCES.

Strayhorn, T. (2009). Accessing and analyzing national databases. In T. Kowalski and T. Lasley (Eds.), *Handbook of Data-Based Decision Making in Education* (pp. 105–122). New York: Routledge.

Strayhorn's (2009) chapter is an introduction to the use of U.S. secondary data in research. Data sets included in this chapter are relevant to early childhood, elementary, secondary and, specifically postsecondary interests. This chapter guides researchers' application of data with discussions of generalizability, combining data, common analysis challenges and supports, as well as an illustrative example.

REFERENCES

- Afana, Y., Lietz, P., & Tobin, M. (2013). The relationship between school resources and grade 8 mathematics achievement: A comparison of Palestinian Authority, Israeli Hebrew and Israeli Arab schools in TIMSS 2007. *Journal for Educational Research Online*, 5(1), 59.
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12.
- Akgul, G., Cokamay, G., & Demir, E. (2016). Predictors of teacher support: Turkey and Shanghai in the programme for international student assessment, 2012. *Eurasian Journal of Educational Research*, 16(63), 115–132.
- Akyuz, G., & Berberoglu, G. (2010). Teacher and classroom characteristics and their relations to mathematics achievement of the students in the TIMSS. *New Horizons in Education*, 58(1), 77–95.
- Alvarado, S. E., & An, B. P. (2015). Race, friends, and college readiness: Evidence from the High School Longitudinal Study. *Race and Social Problems*, 7(2), 150–167.
- Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability students: A comparison between black, hispanic, and white students. *Science Education*, 98(2), 216–242.
- Anderson, K. A. (2016). Examining organizational practices that predict persistence among high-achieving black males in high school. *Teachers College Record*, 118(6), 1–26.
- Arya, D. J., McClung, N. A., Katznelson, N., & Scott, L. (2016). Language ideologies and literacy achievement: Six multilingual countries and two international assessments. *International Journal of Multilingualism*, 13(1), 40–60.
- Ashford, S. N., Lanehart, R. E., Kersaint, G. K., Lee, R. S., & Kromrey, J. D. (2016). STEM pathways: Examining persistence in rigorous math and science course taking. *Journal of Science Education and Technology*, 25(6), 961–975.
- Asparouhov, T., Hamaker, E., & Muthen, B. (2017). Dynamic structural equation models. Technical Report. Version 2. Retrieved on May 18, 2017, from <https://www.statmodel.com/TimeSeries.shtml>

- Azina, I. N., & Halimah, A. (2012). Student factors and mathematics achievement: Evidence from TIMSS 2007. *Eurasia Journal of Mathematics, Science and Technology Education*, 8(3), 249–255.
- Bellibas, M. S., & Liu, Y. (2018). The effects of principals' perceived instructional and distributed leadership practices on their perceptions of school climate. *International Journal of Leadership in Education*, 21(2), 226–244.
- Bittman, B., & Russell, W. (2016). Civic education in the United States: A multiple regression of civic education scores from the National Assessment of Educational Progress. *Research in Social Sciences and Technology*, 1(2), 1–16.
- Blank, R. (2013). Science instructional time is declining in elementary schools: What are the implications for student achievement and closing the gap? *Science Education*, 97(6), 830–847.
- Bouhlila, D. S. (2017). Parents' education and literacy skills: Evidence on inequality of socioeconomic status in Arab countries. *World Development Perspectives*, 5, 34–43.
- Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. Oxford/New York: Oxford University Press.
- Boyce, J., & Bowers, A. (2016). Principal turnover: Are there different types of principals who move from or leave their schools? A latent class analysis of the 2007–8 Schools and Staffing Survey and the 2008–2009 Principal Follow-up Survey. *Leadership and Policy in Schools*, 15(3), 237–272.
- Bozick, R., Malchiodi, A., & Miller, T. (2016). Premigration school quality, time spent in the United States, and the math achievement of immigrant high school students. *Demography*, 53(5), 1477–1498.
- Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W. H., & Zakharov, A. (2016). Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal*, 53(4), 1054–1085.
- Charalambous, C. Y., & Kyriakides, E. (2017). Working at the nexus of generic and content-specific teaching practices: An exploratory study based on TIMSS secondary analyses. *The Elementary School Journal*, 117(3), 423–454.
- Chen, Q. (2014). Using TIMSS 2007 data to build mathematics achievement model of fourth graders in Hong Kong and Singapore. *International Journal of Science and Mathematics Education*, 12(6), 1519–1545.
- Chen, S. F., Lin, C. Y., Wang, J. R., Lin, S. W., & Kao, H. L. (2012). A cross-grade comparison to examine the context effect on the relationships among family resources, school climate, learning participation, science attitude, and science achievement based on TIMSS 2003 in Taiwan. *International Journal of Science Education*, 34(14), 2089–2106.
- Chudgar, A., Luschei, T. F., & Zhou, Y. (2012). Science and mathematics achievement and the importance of classroom composition: Multicountry analysis using TIMSS 2007. *American Journal of Education*, 119(2), 295–316.

- Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: The between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation, 16*(1), 39–52.
- Condron, D. J. (2011). Egalitarianism and educational excellence: Compatible goals for affluent societies? *Educational Researcher, 40*(2), 47–55.
- Conley, S., & You, S. (2017). Key influences on special education teachers' intentions to leave: The effects of administrative support and teacher team efficacy in a mediation model. *Educational Management Administration and Leadership, 45*(3), 521–540.
- Cox, S., Parmer, R., Stizek, G., Thomas, T., & Spiegelman, M. (2017). Documentation for the 2011–12 schools and staffing survey. National Center for Education Statistics. Retrieved on May 18, 2017, from <https://nces.ed.gov/pubs2016/2016817.pdf>
- Curran, F. C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis, 38*(4), 647–668.
- Dee, T., Jacob, B., & Schwartz, N. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35*(2), 252–279.
- Dunst, C., & Hamby, D. (2016). Effects of cumulative family risk factors on American students' academic performance. *American Journal of Educational Research, 4*(2), 150–154.
- Duyar, I., Gumus, S., & Sukru Bellibas, M. (2013). Multilevel analysis of teacher work attitudes: The influence of principal leadership and teacher collaboration. *International Journal of Educational Management, 27*(7), 700–719.
- Engel, M. (2012). The timing of teacher hires and teacher qualifications: Is there an association? *Teachers College Record, 114*, 1–29.
- Fairchild, S., Tobias, R., Corcoran, S., Djukic, M., Kovner, C., & Noguera, P. (2012). White and black teachers' job satisfaction: Does relational demography matter? *Urban Education, 47*(1), 170–197.
- Farinde-Wu, A., & Fitchett, P. G. (2018). Searching for satisfaction: Black female teachers' workplace climate and job satisfaction. *Urban Education, 53*(1), 86–112.
- Fitchett, P., Heafner, T., & Lambert, R. (2014). Assessment, autonomy, and elementary social studies time. *Teachers College Record, 116*, 1–34.
- Froiland, J. M., & Davison, M. L. (2016). The longitudinal influences of peers, parents, motivation, and mathematics course-taking on high school math achievement. *Learning and Individual Differences, 50*, 252–259.
- Froiland, J. M., Davison, M. L., & Worrell, F. C. (2016). Aloha teachers: Teacher autonomy support promotes Native Hawaiian and Pacific Islander students' motivation, school belonging, course-taking and math achievement. *Social Psychology of Education, 19*(4), 879–894.

- Gálvez, I. E., López-Martín, E., Ayuso, J. M., & López, J. M. V. (2017). Determining factors of teachers' self-efficacy in countries of the European Union. Results from TALIS 2013 (pp. 225–248). *Educación, XXI*.
- Giménez, V., Thieme, C., Prior, D., & Tortosa-Ausina, E. (2017). An international comparison of educational systems: A temporal analysis in presence of bad outputs. *Journal of Productivity Analysis, 47*(1), 83–101.
- Goldhaber, D., Destler, K., & Player, D. (2010). Teacher labor markets and the perils of using hedonics to estimate compensating differentials in the public sector. *Economics of Education Review, 29*, 1–17.
- Goldring, E., Cravens, X., Porter, A., Murphy, J., & Elliott, S. (2015). The convergent and divergent validity of the Vanderbilt Assessment of Leadership in Education (VAL-ED) Instructional leadership and emotional intelligence. *Journal of Educational Administration, 53*(2), 177–196.
- Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2012). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review, 57*(1), 54–84.
- Grissom, J. (2011). Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record, 113*(11), 2552–2585.
- Grissom, J., & Harrington, J. (2010). Investing in administrator efficacy: An examination of professional development as a tool for enhancing principal effectiveness. *American Journal of Education., 116*(4), 583–612.
- Grissom, J., & Keiser, L. (2011). A supervisor like me: Race, representation, and the satisfaction and turnover decisions of public sector employees. *Journal of Policy Analysis and Management, 30*(3), 557–580.
- Grissom, J., Nicholson-Cotty, S., & Harrington, J. (2014). Estimating the effects of No Child Left Behind on teachers' work environments and job attitudes. *Educational Evaluation and Policy Analysis, 36*(4), 417–436.
- Gumus, E., & Bellibas, M. S. (2016). The effects of professional development activities on principals' perceived instructional leadership practices: Multi-country data analysis using TALIS 2013. *Educational Studies, 42*(3), 287–301.
- Gustafsson, J. E. (2016). Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence, 57*, 66–72.
- Guzel, C. I., & Berberoglu, G. (2010). Students' affective characteristics and their relation to mathematical literacy measures in the programme for international student assessment (PISA) 2003. *Eurasian Journal of Educational Research, 40*, 93–112.
- Harring, J., Stapleton, L., & Beretvas, S. (2015). *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications*. Charlotte: Information Age Publishing.

- He, J., Van de Vliert, E., & Van de Vijver, F. J. (2017). Extreme response style as a cultural response to climato-economic deprivation. *International Journal of Psychology, 52*(S1), 67–71.
- Heck, R., & Thomas, S. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York: Routledge.
- Hillman, N. W., Gast, M. J., & George-Jackson, C. (2015). When to begin? Socioeconomic and racial/ethnic differences in financial planning, preparing, and saving for college. *Teachers College Record, 117*(8), 1–16.
- Hopland, A. O. (2013). School facilities and student achievement in industrial countries: Evidence from the TIMSS. *International Education Studies, 6*(3), 162.
- Howell, D. (2017). *Fundamental statistics for the behavioral sciences* (9th ed.). Boston: Cengage Learning.
- Hurwitz, M., & Howell, J. (2014). Estimating causal impacts of school counselors with regression discontinuity designs. *Journal of Counseling and Development, 92*, 316–327.
- Ingels, S. J., Pratt, D. J., Herget, D. R., Burns, L. J., Dever, J. A., Ottem, R., Rogers, J. E., Jin, Y., & Leinwand, S. (2011). *High School Longitudinal Study of 2009 (HSLS:09): Base-year data file documentation*. NCES. Retrieved on May 18, 2017, from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011328>
- Ingersoll, R., & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis, 34*(4), 435–464.
- Ingersoll, R., & Perda, D. (2010). Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal, 47*(3), 563–594.
- Jackson, K., & Marriott, C. (2012). The interaction of principal and teacher instructional influence as a measure of leadership as an organizational quality. *Educational Administration Quarterly, 48*(2), 230–258.
- Jenkins, A., & Ueno, A. (2017). Classroom disciplinary climate in secondary schools in England: What is the real picture? *British Educational Research Journal, 43*(1), 124–150.
- Jiang, F., & McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education, 37*(3), 554–576.
- Kaplan, D. (2009). *Structural equation modeling: foundations and extensions* (Vol. 10). Los Angeles: SAGE.
- Kelly, S. (2010). A crisis of authority in predominately black schools? *Teachers College Record, 112*(5), 1247–1274.
- Lambert, R., McCarthy, C., Fitchett, P., Lineback, S., & Reiser, J. (2015). Identification of elementary teachers' risk for stress and vocational concerns using the national schools and staffing survey. *Education Policy Analysis Archives, 23*(43), 1–37.

- Lee, J. (2010). Tripartite growth trajectories of reading and math achievement: Tracking of national academic progress at primary, middle, and high school levels. *American Educational Research Journal*, 47(4), 800–832.
- Lee, J. (2012a). Educational equity and adequacy for disadvantaged minority students: School and teacher resource gaps toward national mathematics proficiency standard. *Journal of Educational Research*, 105(1), 64–75.
- Lee, J. (2012b). College for all: Gaps between desirable and actual P-12 math achievement trajectories for college readiness. *Educational Researcher*, 41(2), 43–55.
- Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, 34(2), 209–231.
- Li, W., & Konstantopoulos, S. (2017). Does class-size reduction close the achievement gap? Evidence from TIMSS 2011. *School Effectiveness and School Improvement*, 28(2), 292–313.
- Liou, P. Y. (2017). Profiles of adolescents' motivational beliefs in science learning and science achievement in 26 countries: Results from TIMSS 2011 data. *International Journal of Educational Research*, 81, 83–96.
- Liu, Y., Bellibas, M. S., & Printy, S. (2016). How school context and educator characteristics predict distributed leadership: A hierarchical structural equation model with 2013 TALIS data. *Educational Management Administration & Leadership*. <https://doi.org/10.1177/1741143216665839>.
- Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology*, 108(2), 256.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J., Abdelfattah, F., Leung, K. C., Xu, M., Nagengast, B., & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, 105(1), 108.
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of US and Saudi Arabian TIMSS responses. *Journal of Cross-Cultural Psychology*, 45(5), 777–804.
- Mason-Williams, L., & Gagnon, C. (2017). An analysis of teacher sorting in secondary special education and alternative schools. *The Journal of Special Education*, 50(4), 239–250.
- Mattern, J., & Bauer, J. (2014). Does teachers' cognitive self-regulation increase their occupational well-being? The structure and role of self-regulation in the teaching context. *Teaching and Teacher Education*, 43, 58–68.

- McEneaney, E., Lopez, F., & Nieswandt, M. (2014). Instructional models for the acquisition of English as bridges into school science: Effects on the science achievement of U.S. Hispanic English language learners. *Learning Environments Research, 17*, 305–318.
- Meroni, E. C., Vera-Toscano, E., & Costa, P. (2015). Can low skill teachers make good students? Empirical evidence from PIAAC and PISA. *Journal of Policy Modeling, 37*(2), 308–323.
- Mertler, C., & Vannatta Reinhart, R. (2017). *Advanced and multivariate statistical methods: Practical application and interpretation* (6th ed.). New York: Routledge.
- Min, I., Cortina, K. S., & Miller, K. F. (2016). Modesty bias and the attitude-achievement paradox across nations: A reanalysis of TIMSS. *Learning and Individual Differences, 51*, 359–366.
- Mullis, I., & Martin, M. (2013). *TIMSS 2015 assessment framework*. IEA. Retrieved on May 18, 2017, from <https://timssandpirls.bc.edu/timss2015/frameworks.html>
- Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033.
- National Center for Education Statistics (NCES). (2009). *NAEP technical documentation*. Retrieved on May 18, 2017, from <https://nces.ed.gov/nationsreportcard/tdw/>
- Ni, Y. (2012). Teacher working conditions in charter schools and traditional public schools: A comparative study. *Teachers College Record, 114*(3), 1–26.
- Noble, R., & Morton, C. (2013). African Americans and mathematics outcomes on National Assessment of Educational Progress: Parental and individual influences. *Journal of Child and Family Studies, 22*(1), 30–37.
- Oberfield, Z. (2016). A bargain half fulfilled: Teacher autonomy and accountability in traditional and public schools and public charter schools. *American Educational Research Journal, 53*(2), 296–323.
- Phillips, K., Desimone, L., & Smith, T. (2011). Teacher participation in content-focused professional development and the role of state policy. *Teachers College Record, 113*(11), 2586–2630.
- Pholpirul, P. (2016). Pre-primary education and long-term education performance evidence from Programme for International Student Assessment (PISA) Thailand. *Journal of Early Childhood Research*. <https://doi.org/10.1177/1476718X15616834>.
- Pope, D., & Sydnor, J. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives, 24*(2), 95–108.
- Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Researcher, 40*(3), 103–116.

- Preston, C., Goldring, E., Berends, M., & Cannata, M. (2012). School innovation in district context: Comparing traditional public schools and charter schools. *Economics of Education Review*, *31*, 318–330.
- Price, H. (2012). Principal-teacher interactions: How affective relationships shape principal and teacher attitudes. *Educational Administration Quarterly*, *48*(1), 39–85.
- Price, H., & Collett, J. (2012). The role of exchange and emotion on commitment: A study of teachers. *Social Science Research*, *41*, 1469–1479.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage.
- Reback, R., Rockoff, J., & Schwartz, H. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, *6*(3), 207–241.
- Renzulli, L., Parrott, H., & Beattie, I. (2011). Racial mismatch and school type: Teacher satisfaction and retention in charter and traditional public schools. *Sociology of Education*, *84*(1), 23–48.
- Rindermann, H., & Thompson, J. (2011). Cognitive capitalism the effect of cognitive ability on wealth, as mediated through scientific achievement and economic freedom. *Psychological Science*, *22*, 754–763.
- Rivkin, S. G., & Schiman, J. C. (2015). Instruction time, classroom quality, and academic achievement. *The Economic Journal*, *125*(588), F425–F448.
- Ronfeldt, M., Schwartz, N., & Jacob, B. (2014). Does preservice preparation matter? Examining an old question in new ways. *Teachers College Record*, *116*(10), 1–46.
- Rutkowski, D., Rutkowski, L., Belanger, J., Knoll, S., Weatherby, & Prusinski, E. (2013). *Teaching and Learning International Survey TALIS 2013 conceptual framework*. OECD. Retrieved on May 18, 2017, from http://www.oecd.org/edu/school/TALIS%20Conceptual%20Framework_FINAL.pdf
- Sandoval-Hernández, A., & Białowolski, P. (2016). Factors and conditions promoting academic resilience: A TIMSS-based analysis of five Asian education systems. *Asia Pacific Education Review*, *17*(3), 511–520.
- Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher*, *41*(8), 294–308.
- 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.
- Schneider, B., & Saw, G. (2016). Racial and ethnic gaps in postsecondary aspirations and enrollment. *Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 58–82.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

- Shen, J., Leslie, J., Spybrook, J., & Ma, X. (2012). Are principal background and school processes related to teacher job satisfaction? A multilevel student using Schools And Staffing Survey 2003–04. *American Educational Research Journal*, *49*(2), 200–230.
- Shi, Q., Zhang, S., & Lin, E. (2014). Relationships of new teachers' beliefs and instructional practices: Comparisons across four countries. *Action in Teacher Education*, *36*(4), 322–341.
- Smith, W. C. (2017). National testing policies and educator based testing for accountability. *OECD Journal: Economic Studies*, *2016*(1), 131–149.
- Spees, L., Potochnick, S., & Perreira, K. (2016). The academic achievement of limited English proficient (LEP) youth in new and established immigrant states: Lessons from the National Assessment of Educational Progress (NAEP). *Education Policy Analysis Archives*, *24*(99), 1–27.
- Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, *48*, 137–151.
- Stoet, G., & Geary, D. C. (2017). Students in countries with higher levels of religiosity perform lower in science and mathematics. *Intelligence*, *62*, 71–78.
- Sun, M., & Ni, Y. (2016). Work environments and labor markets: Explaining principal turnover gap between charter and traditional public schools. *Educational Administration Quarterly*, *52*(1), 144–183.
- Tickle, B., Chang, M., & Kim, S. (2011). Administrative support and its mediating effect on US public school teachers. *Teaching and Teacher Education*, *27*, 342–349.
- Tramonte, L., & Willms, J. D. (2010). Cultural capital and its effects on education outcomes. *Economics of Education Review*, *29*(2), 200–213.
- Tsai, S. L., Smith, M. L., & Hauser, R. M. (2017). Families, schools, and student achievement inequality: A multilevel MIMIC model approach. *Sociology of Education*, *90*(1), 64–88.
- Urlick, A. (2016a). Examining US principal perception of multiple leadership styles used to practice shared instructional leadership. *Journal of Educational Administration*, *54*(2), 152–172.
- Urlick, A. (2016b). The influence of typologies of school leaders on teacher retention: A multilevel latent class analysis. *Journal of Educational Administration*, *54*(4), 434–468.
- Urlick, A., Wilson, A. S., Ford, T. G., Frick, W. C., & Wronowski, M. L. (2018). Testing a framework of math progress indicators for ESSA: How opportunity to learn and instructional leadership matter. *Educational Administration Quarterly*, *54*(3), 396–438.
- Vaysettes, S., Achiron, M., & Limoges, S. (2016). *PISA 2015 assessment and analytical framework*. OECD. Retrieved on May 18, 2017, from <http://www.oecd.org/publications/pisa-2015-assessment-and-analytical-framework-9789264255425-cn.htm>

- Ware, H., & Kitsantas, A. (2011). Predicting teacher commitment using principal and teacher efficacy variables: An HLM approach. *Journal of Educational Research, 104*(3), 183–193.
- Weathers, J. (2011). Teacher community in urban elementary schools: The role of leadership and bureaucratic accountability. *Education Policy Analysis Archives, 19*(3), 1–39.
- Wieczorek, D. (2017). Principals' perceptions of public schools' professional development changes during NCLB. *Education Policy Analysis Archives, 25*(8), 1–45.
- Winnar, L. D., Frempong, G., & Blignaut, R. (2015). Understanding school effects in South Africa using multilevel analysis: Findings from TIMSS 2011. *Electronic Journal of Research in Educational Psychology, 13*(1), 151–170.
- Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review, 30*(3), 404–418.
- Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *The Journal of Economic Perspectives, 30*(3), 3–31.
- Woo, H., & Henfield, M. S. (2016). Student and teacher factors' impact on fourth grade students' mathematics achievement: An HLM analysis of TIMSS 2007. *Journal of Mathematics Education, 9*(1), 69–87.
- Wu, Y. (2016). Universal beliefs and specific practices: Students' math self-efficacy and related factors in the United States and China. *International Education Studies, 9*(12), 61.
- Xie, D., & Shen, J. (2013). Teacher leadership at different school levels: Findings and implications from the 2003–04 Schools And Staffing Survey in US public schools. *International Journal of Leadership in Education, 16*(3), 327–348.
- Young, J. L. (2017). Access, achievement, and academic resilience: The relationship between AVID and Black student participation in advanced placement courses. *Journal of Multicultural Affairs, 1*(2), 4.
- Yu, R., & Singh, K. (2018). Teacher support, instructional practices, student motivation, and mathematics achievement in high school. *The Journal of Educational Research, 111*(1), 81–94.



Matching Strategies for Causal Inference with Observational Data in Education

Yongnam Kim, Stanley A. Lubanski, and Peter M. Steiner

In educational research, causal questions like “how does retention affect children’s learning development?”, “do Catholic schools produce higher student achievement?”, or “did the No Child Left Behind program improve graduation rates?” are frequently posed. Well-implemented randomized experiments answer such questions because random assignment of study units (e.g., students, schools, districts) creates treatment and control groups that are statistically equivalent, on average, before treatment implementation. However, due to ethical or financial concerns, randomized experiments are often infeasible in practice. Instead, matching non-equivalent treatment and control units on the basis of observational data can be a viable alternative.

In observational studies, matching is applied to treatment and control groups that were formed according to *non-randomized* selection processes (e.g., self-selection or third-person selection via program administrators). Thus, the treatment and control groups almost surely differ not only in

Y. Kim (✉) • S. A. Lubanski • P. M. Steiner
Department of Educational Psychology, University of Wisconsin-Madison,
Madison, WI, USA
e-mail: ykim379@wisc.edu; lubanski@wisc.edu; psteiner@wisc.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_9

their observed but also unobserved characteristics. Consider the effect of student retention on achievement scores: if student ability determines both retention and achievement, then any achievement difference between retained and promoted students might just be due to differential selection rather than retention. If so, the estimated retention effect is *confounded* with differences in the retained and promoted students' ability. Since many characteristics may simultaneously confound the treatment effect, strong subject-matter theory about treatment selection and the outcome-determining factors is required to identify and reliably measure the confounding factors. In the absence of other complications like missing data or attrition, matching treatment and control units on observed covariates allow us to estimate the causal effect if the set of covariates can successfully capture all the confounding. Otherwise, causal claims are not warranted.

In this chapter, we give a non-technical overview of matching strategies, emphasizing propensity score (PS) techniques. We first highlight the relevant matching literature, introduce the causal quantities of interest, and discuss the assumptions required for identifying causal effects. Then, we present the most important matching strategies for single and multi-level data and discuss their implementation using an applied example. We conclude with practical recommendations about variable selection and the choice of a specific matching technique. Throughout the chapter, we discuss ideas and concepts with respect to the effect of retaining students based on student achievement.

REVIEW OF RELEVANT LITERATURE

Methodological Literature Historically, matching methods to equate non-equivalent treatment and control groups have played a central role in causal inference. The idea of matching can be traced back to J. S. Mill (1864) who emphasized the importance of “treated and control units that are identical but for the treatment” (Rosenbaum 2005, p. 147). This philosophical justification was accepted in statistics by Cochran and Rubin (1973), and Rubin (1973), and with the development of the potential outcomes framework for causal inference (Rubin 1974), matching methods were rigorously formalized.

Building on this framework, Rosenbaum and Rubin (1983) developed the PS, a univariate score computed from the observed baseline characteristics, to simplify the matching process. Since then, a large class of matching

techniques has been developed (for a summary, see Schafer and Kang 2008; also Austin 2011). More recently, researchers have investigated matching strategies for hierarchically structured data (e.g., when students are nested within classrooms and schools) and how to deal with peer effects among study participants (Hong and Raudenbush 2006; Steiner et al. 2012).

Applied Literature In educational research, PS-matching has been used to estimate the effects of retention policies on math and reading achievement growth (West and Hughes 2008) and the effects of school size on math achievement (Wyse et al. 2008). Kelcey (2011) used PS-stratification to study the effects of teacher reading knowledge on student achievement, while Hong and Raudenbush (2006) used PS-stratification to assess the effects of kindergarten retention on student achievement. PS-weighting has been used to assess the effects of private school voucher plans (Lara et al. 2011) and NCLB (No Child Left Behind)'s test-driven accountability system on academic performance (Lee and Reeves 2012).

ESTIMANDS AND ASSUMPTIONS

Causal Estimands The Rubin Causal Model (Holland 1986) provides a convenient and rigorous framework for defining causal quantities of interest. The key concepts are the potential treatment and control outcomes which we would observe if a unit experiences the treatment and control condition, respectively. In our example, the *potential treatment outcome*, $\mathcal{Y}_i(1)$; refers to the achievement score we would observe if student i were to be retained ($Z_i = 1$, where Z_i is the retention indicator). If student i were to be promoted ($Z_i = 0$), we would observe their *potential control outcome*, $\mathcal{Y}_i(0)$. Given the pair of potential outcomes, the individual causal effect at the unit level is defined as the difference in the potential treatment and control outcome: $\tau_i = \mathcal{Y}_i(1) - \mathcal{Y}_i(0)$. However, since a student cannot be retained and promoted at the same time, only one of the two potential outcomes is realized. Thus, we generally cannot identify the unit-level causal effect.

Since individual causal effects are hard to identify and estimate, researchers frequently focus on the *average treatment effect* (ATE), which is the average of the individual causal effects across all units in the population, $E[\tau_i] = E[\mathcal{Y}_i(1) - \mathcal{Y}_i(0)]$. Sometimes, they might not be directly interested in the treatment's effect on the entire population, but only in the *average*

treatment effect for the treated (ATT). The ATT is the average effect for those who actually received the treatment: $E[\tau_i | Z_i = 1] = E[\Upsilon_i(1) - \Upsilon_i(0) | Z_i = 1]$. For instance, the ATT might be more useful for assessing retention effects because retention policies only target potentially disadvantaged students who are at risk to be retained.

Identification Assumptions The causal estimands defined above cannot directly be estimated, because they involve *unobserved* potential outcomes. However, if two assumptions are met, we can express the ATE and ATT in terms of *observable* quantities and we say that the causal effects are *identified*.

The first assumption is the *strong ignorability* assumption which ensures that all the confounding bias can be removed (Rosenbaum and Rubin 1983). This assumption addresses the major challenge in causal inference with observational studies: the non-equivalence of the treatment and control group at baseline. Assume that retained students tend to be less able and from households with lower incomes than promoted students. Then, retained students will be lower achieving at the posttest even if retention has no effect. However, if a retained student i and a promoted student j have the same levels of ability and household income, then the difference between the two students' posttest scores may be safely attributed to the retention policy. This is possible if (a) ability and income are the only confounding characteristics and they are reliably measured, and if (b) the retention decision is probabilistic (not deterministic). The two conditions are formally expressed as *unconfoundedness*, $\{\Upsilon(1), \Upsilon(0)\} \perp Z | \mathbf{X}$ (i.e., the potential outcomes are independent of the treatment Z , given the covariates \mathbf{X}), and *positivity*, $0 < \Pr(Z = 1 | \mathbf{X}) < 1$, where \mathbf{X} is a set of carefully chosen baseline covariates that removes all the confounding bias.

The second assumption is the *stable unit treatment value assumption* (SUTVA) which requires that (a) a unit's potential outcomes do not vary with other unit's treatment assignment (i.e., no peer effects), and (b) that there is only one well-defined version of the treatment (Rubin 1980). When defining the causal estimands above, we assumed that each student i has only two potential outcomes, $\Upsilon_i(1)$ and $\Upsilon_i(0)$, depending only on student i 's treatment status, $Z_i = 1$ or $Z_i = 0$, respectively. However, as with all assumptions, SUTVA might be frequently violated because humans' decisions and behaviors are regularly affected by their friends or neighbors. Suppose that student i 's potential achievement depends not only on

her retention status but also whether her peer j is retained or promoted. Then, this student i has not only two but *four* potential outcomes: $\Upsilon_i(1, Z_j = 0)$, $\Upsilon_i(1, Z_j = 1)$, $\Upsilon_i(0, Z_j = 0)$, $\Upsilon_i(0, Z_j = 1)$. In this case, it is unclear how to define unit-level and average causal effects because we have two potential treatment outcomes and two potential control outcomes for each student. If student i 's potential outcomes depend on the retention status of multiple peers, such as students j, k, l, m , then the number of possible potential outcomes increases exponentially and the comparison is nearly intractable. Thus, SUTVA is essential for making causal inference feasible. Nonetheless, since peer effects are often a reality in educational contexts, researchers have tried to relax SUTVA and we will discuss one such example later.

MATCHING STRATEGIES

If the identifying assumptions hold, the causal effect can be estimated by different strategies. The choice of a specific matching strategy depends on the level of treatment implementation and selection. We first discuss the most popular matching techniques for single-level data and then techniques for multilevel data.

Matching with Single-Level Data There are many ways to match non-equivalent treatment and control groups with single-level data. Individual-case matching is one possibility that can be implemented as multivariate matching or PS-matching. Other strategies use the PS for stratifying or weighting the data.

Multivariate and PS-Matching The basic idea of individual-case matching is to find a control unit for each treated unit that is as similar as possible (ideally identical) with respect to the baseline covariates \mathbf{X} or the corresponding PS, and to discard the unmatched units. The resulting subsample of treatment and control units is then comparable in baseline covariates and the ATT can be directly estimated without bias (if the identifying assumptions hold).

The similarity of two units can be *quantified* with a distance metric. Researchers frequently use the Mahalanobis distance, a standardized version of the Euclidian distance which also accounts for the correlation among covariates. Using pairwise differences, treatment and control units

with the smallest difference are matched. Since the units are matched with respect to the difference on the multivariate covariates, Mahalanobis distance matching is an example of a *multivariate matching* technique.

A popular alternative is matching on the PS, which quantifies each unit's probability of receiving the treatment given baseline covariates \mathbf{X} , $PS = \Pr(Z = 1 | \mathbf{X})$ (Rosenbaum and Rubin 1983). Then, treatment and control units with the same or at least a very similar PS are matched. Since students with *different* values of ability and income can have the *same* PS, finding close matches on the PS is easier than with the multivariate covariates \mathbf{X} . Since the true PSs are unknown, they are typically estimated via logistic regression where the binary treatment variable is regressed on covariates \mathbf{X} . Alternatively, ensemble methods like classification trees, random forests, or neural networks that do not rely on the linearity assumption can be used to estimate the PS (Keller et al. 2015; McCaffrey et al. 2004). Irrespective of the chosen method for estimating the PS, the predicted treatment probabilities are used as the PS. The similarity between a treated and control unit is then measured by the PS distance, the absolute difference between the units' raw PSs or the corresponding logits (logit = $\log \{PS/(1 - PS)\}$).

Figure 9.1a shows an example plot where the outcome, \mathcal{Y} , of the treated (“T”) and control units (“C”) is plotted against their PS-logit. The plot indicates that the two groups are not comparable at baseline because the treated group has on average higher logits (i.e., a greater probability of being retained) than the control group. Thus, the unadjusted group mean difference—the difference between two horizontal dashed lines which is almost zero—cannot be interpreted as a valid causal effect because it is likely contaminated with confounding bias. However, if the PS has been estimated from a set of baseline covariates that establishes a strongly ignorable selection process, the causal effect is identified and estimable via PS-matching as illustrated in Fig. 9.1b. Based on the PS-logit, each treated unit is matched to its closest control unit, as indicated by the vertical segments connecting the matched pairs. Treated units without a close match remain unmatched and, together with the unmatched control units, get discarded from the data set before estimating the ATT. The ATT is then the average of the matched pairs' outcome differences (i.e., the average height of the vertical lines in Fig. 9.1b). Since each treated case is matched to a single control case, this type of matching is called one-to-one matching. Alternative matching techniques allow for multiple control units (one-to-many matching), or even multiple treatment and control units

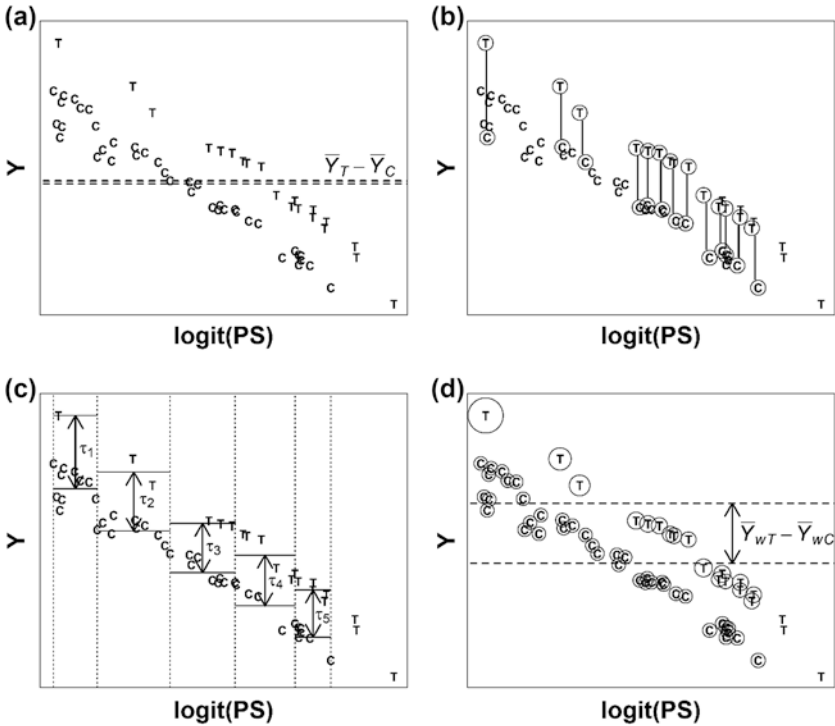


Fig. 9.1 Illustration of PS techniques (a) Initial status. (b) PS-matching (c) PS-stratification. (d) PS-weighting. (Note. “T” denotes treated units, while “C” denotes control units)

(optimal full matching). Matching techniques may also involve the replacement of matched control units (allowing a control unit to be matched to two or more treated units), calipers (treatment units with no close match inside the caliper remain unmatched), or matching algorithms (greedy, genetic, or optimal). For more details, see Austin (2011), Steiner and Cook (2013), or Stuart (2010).

A variety of software can implement such algorithms and matching options. The R packages *optmatch* (Hansen and Klopfer 2006) and *MatchIt* (Ho et al. 2011) are useful, and STATA offers the commands *psmatch2* (Leuven and Sianesi 2003) and *teffects psmatch* (StataCorp 2015). For SAS, *gmatch* and *vmatch* macros (Kosanke and Bergstralh

2004) are available. For SPSS users, Thoemmes (2012) provides an SPSS custom dialog for PS-matching. Stuart (2010) provides an extensive list of software for matching methods.

PS-Stratification PS-stratification creates subclasses of individuals based on the quantiles of the PS distribution (or, equivalently, the PS-logit). Figure 9.1c illustrates the stratification approach with five strata, each containing 20% of the units. The vertical dashed lines represent the boundaries of the strata. The rationale of PS-stratification is that units within each stratum are homogeneous (comparable) with respect to the covariate distribution and thus allow for an unbiased estimation of stratum-specific treatment effects (i.e., the differences in stratum-specific group means, bi-headed arrows, as shown in Fig. 9.1c). Average causal effects are then obtained as a weighted average of the stratum-specific effects. For ATE, the stratum weights reflect the number of units in each stratum; for ATT, the weights reflect the distribution of treated units across strata.

In contrast to PS-matching, PS-stratification uses all units in the sample (maybe with the exception of non-overlapping units at the tail of the PS distribution). Thus, PS-stratification is usually more efficient (i.e., results in a smaller standard error). However, with only five strata, the assumption of within-stratum homogeneity may not be perfectly met such that about 10% of the initial bias could remain (Cochran 1968). Increasing the number of strata can eliminate more bias, but small stratum sizes may result in unstable stratum-specific effects. Typically, five to ten strata have been used in practice (Stuart 2010).

Inverse-Propensity Weighting Following survey sampling designs, the PS can be viewed as the probability that a unit of the population gets sampled into the treatment group. The basic idea of PS-weighting is to create *comparable* pseudo-populations by weighting the observed treatment and control units. For instance, a retained (i.e., treatment) student with a PS of 0.2 receives a weight of $1/0.2 = 5$, whereas a retained student with a PS of 0.8 receives a weight of $1/0.8 = 1.25$. The first student has a larger weight because he is underrepresented with respect to the overall population due to the lower propensity of being retained, while the student with a high propensity is overrepresented. Analogously, we compute the weights for the promoted (i.e., control) students, but instead of $1/PS$, we use $1/(1-PS)$ because $(1-PS)$ is the propensity that a student gets promoted.

PS-weighting is illustrated in Fig. 9.1d, where the area of the circles reflects the size of the weights (non-overlapping units receive a weight of zero, thus they are not circled). Treated units with a lower PS (or PS-logit) receive relatively larger weights than corresponding control units, suggesting that the underrepresented treatment units are up-weighted, while the overrepresented control units are down-weighted (and vice versa for units with high PSs). The ATE is then computed as the difference between weighted treatment and control means—the difference between two dashed lines in Fig. 9.1d. The ATE can also be conveniently estimated with a *weighted least squares* regression of the outcome on the treatment variable with the corresponding inverse-propensity weights. If the ATT is of interest, a weight of 1 is assigned to the treated units and a weight of $PS/(1-PS)$ to the control units. Compared to PS-matching or PS-stratification, PS-weighting is rather sensitive to outliers. To deal with these units, researchers can trim large weights or drop them altogether (but then the PSs must be re-estimated with the remaining units). For more details, see Schafer and Kang (2008), who also provide standard error formulas for PS-weighting estimators. Griffin et al. (2014) developed *twang* (toolkit for weighting and analysis of non-equivalent groups) which exists as an R package, STATA command, and SAS macro, and provide useful tutorials.

Doubly Robust Strategies All the PS techniques can be combined with an additional covariance adjustment in the outcome analyses. With matched, stratified, or weighted data, the treatment effect can be estimated by regressing the outcome on the treatment plus any additional baseline covariates. Such doubly robust strategies have two advantages. First, researchers have two chances to remove confounding bias, either via a correctly estimated PS (i.e., treatment selection model) or a correctly specified outcome model. Second, controlling for additional covariates yields more efficient estimators—that is, standard errors are smaller. A detailed discussion of the doubly robust methods can be found in Kang and Schafer (2007).

Matching with Multilevel Data All the matching strategies discussed so far assumed a single-level data structure. However, in education, we typically have hierarchically structured data, where students are nested within classes, classes within schools, and schools within districts and states. For simplicity, assume a two-level structure where students are nested within

schools. If entire schools choose to participate in the treatment or control condition, then the same matching strategies as already discussed apply, but schools are matched on school-level covariates, including aggregated student-level covariates. Once the schools are matched, stratified, or weighted, a multilevel model is used to estimate the causal effect and its standard error. The R package *matchMulti* (Keele and Pimentel 2016) can be useful for this type of matching where the treatment selection occurs at the cluster-level.

Matching strategies may differ if students select into treatment and control conditions within a school (e.g., students get retained within schools). In this case, we can pursue one of three major matching strategies: treatment and control students can be matched within each school separately, students can be matched across all available schools, or students can be matched across homogeneous schools only. See J.S. Kim and Steiner (2015) or Thoemmes and West (2011) for the details.

Within-Cluster Matching Within-cluster matching first matches treatment and control units within each cluster and then estimates the ATE or ATT for each cluster separately. The weighted average of cluster-specific effects can be used to estimate the causal effect for the entire population of units or clusters. Within-cluster matching requires that strong ignorability holds within each cluster. Thus, only level-one covariates are needed to estimate cluster-specific PSs—no cluster-level covariates are required. However, the drawback of within-cluster matching is that we might lack sufficient overlap between treatment and control units, particularly so when cluster sample sizes are small and selection is strong—as it may be when retaining or promoting students.

Across-Cluster Matching In order to increase the overlap among treated and control units, we may allow for matching across clusters. For instance, if we cannot find a comparable promoted student for a retained student in the same school, we might borrow a more similar control student from another school. To make across-cluster matching work, we need both level-one and level-two (cluster-level) covariates because the selection process might differ across clusters. The ATE or ATT is identified only if level-one and cluster-level covariates together establish strong ignorability with respect to all level-one units. Thus, the estimation of the PS requires a joint PS model that includes both levels of covariates. The major disadvantage

of this strategy is that the PS model needs to correctly reflect selection differences across clusters. Hence, all relevant cluster-level covariates must be reliably measured and correctly modeled. For this reason, across-cluster matching relies on stronger assumptions than within-cluster matching.

Across-Cluster Matching Within Homogenous Groups of Clusters In order to increase the chances of meeting the strong ignorability assumptions across clusters, researchers might first try to identify homogeneous groups of clusters (based on observed or latent covariates) and then allow for across-cluster matches but only within the homogeneous groups. Homogeneity can be defined with respect to the selection process, outcome generating model, or both. Other advantages of this matching strategy are the larger sample sizes of the homogeneous groups and the better overlap between the treated and control units as compared to within-cluster matching.

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY

In order to illustrate the application of a PS-matching strategy, we describe the kindergarten retention study of Hong and Raudenbush (2006, henceforth HR) in more detail. HR used the Early Childhood Longitudinal Study Kindergarten cohort (ECLS-K) data of the National Center for Education Statistics to investigate the effect of kindergarten retention on children's reading and math achievement. This study has made two major methodological contributions. First, HR tried to relax SUTVA, which is often questionable in educational settings because of peer effects. Second, HR took the nesting of students within classrooms/schools into account. HR implemented across-cluster matching via PS-stratification, with the PS estimated from child- and school-level covariates. Using the kindergarten retention study, we illustrate the four major steps of a PS analysis: (1) Defining causal estimands and assessing assumptions, (2) PS estimation and balance/overlap check, (3) causal effect estimation, and (4) sensitivity analysis.

Step 1: Causal Estimands and Assumptions At the beginning of the study, it is important to clarify the causal estimands of interest and to assess the strong ignorability assumption. HR defined several causal effects of interest, but here we only focus on one: the average effect of retention for at-risk students in low-retention schools. They estimated the ATE for the

students at risk to be retained because evaluating the retention effect for the entire population of at-risk students is of high relevance for policy decisions. To identify the causal effect, they assumed that the large sets of more than 200 child-level (\mathbf{X}) and school-level covariates (\mathbf{W}) capture all the confounding, that is, the covariates establish unconfoundedness. Implicitly, HR also assumed positivity (each at-risk child has a positive probability of being retained).

Step 2: PS Estimation and Balance/Overlap Check To estimate the PS of being retained in low-retention schools, HR ran a two-level logistic regression model, with retention status as the dependent variable and child- and school-level covariates as the explanatory variables. They then used the estimated retention probabilities to stratify all students into ten strata. However, for successful bias removal, the estimated PS has to guarantee satisfactory balance and overlap. First, in order to check overlap of the retained and promoted students, HR investigated the number of retained and promoted students in each stratum. One stratum did not contain any retained students and thus was discarded (note that discarding strata restricts the generalizability of results). Then, within each remaining stratum, they tested the balance of the 213 child- and school-level covariates. Since almost no stratum-specific group mean differences in covariates were statistically significant (at the 5% level), HR concluded that the PS-strata sufficiently balanced the retained and promoted students' covariate distributions. In addition to testing for significant differences, which is sensitive to the within-stratum sample sizes, HR could have also reported *standardized mean differences* in covariates. Standardized mean differences, and also variance ratios before and after the PS adjustment, are frequently used to demonstrate the improvement in balance. For more details, see Imbens and Rubin (2015). If the balance checks would indicate any remaining imbalances in covariates, one would need to re-specify the PS model with different covariates, interaction, or higher-order terms, or increase the number of strata. Once balance on the overlapping treatment and control cases is satisfactory, the causal effect can be estimated.

Step 3: Causal Effect Estimation With well-balanced PS-strata, estimating the causal effect is straightforward. HR implemented PS-stratification in a two-level hierarchical model by regressing the reading achievement scores on the retention indicator and a set of dummy variables representing the nine strata with overlap. In order to deal with residual bias left within each

stratum, HR also included the PS-logit as an additional predictor in their model. For the reading outcome, HR found a retention effect of -8.18 with a standard error of 0.94 , which is statistically significant at the 5% level.

Step 4: Sensitivity Analysis Although HR assumed that their baseline covariates \mathbf{X} and \mathbf{W} established unconfoundedness (step 1), the negative retention effect HR found might still be due to unobserved confounding. To consider a reasonable worst-case confounding scenario, HR modeled a hypothetical unobserved confounders U_X (child-level) and U_W (school-level) according to the second strongest measured confounders (i.e., they used the observed confounders' association with the treatment and outcome to create the corresponding associations for the hypothetical confounders). Under this scenario, the 95% confidence interval for the new effect estimate was $(-2.44, 1.24)$. Thus, if such unmeasured confounding were actually present, then there would be no evidence for a significant (negative) retention effect. This reveals that the estimated retention effect is rather sensitive to violations of the strong ignorability assumption. Though HR's analysis relied on the linearity assumption, non-parametric sensitivity analyses are also possible (see, for instance, Rosenbaum 2002).

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

Compared to conventional regression methods, matching methods may result in smaller sample sizes and less efficient estimators due to discarding unmatched or non-overlapping units, or extreme weights when using PS-weighting. The deletion of non-overlapping units particularly restricts the generalizability of causal effect estimates. But the matching methods' drawbacks might be outweighed by their major advantages: (a) Treatment and control groups can be made comparable without looking at the outcome, which helps in avoiding fishing for significant estimates (Rubin 2007); (b) Matching methods do not rely on linear modeling assumptions—whether balance in observed covariates is achieved via parametrical or non-parametrical methods does not matter; And (c) matching methods do not rely on extrapolation into areas of non-overlap because the non-overlapping units are discarded before estimating the effects. In the following, we discuss three issues that are important for implementing matching methods in practice.

Covariate Selection In matching studies, selecting the “right” covariates for matching treatment and control units is the main practical challenge. Theoretically, matching estimators succeed in removing all the confounding bias only when the measured covariates establish strong ignorability. In practice, researchers often think that the availability of a large set of baseline covariates guarantees the removal of at least a major portion of the confounding bias, if not all the bias. However, recent studies have shown that thoughtless selection of covariates does not necessarily decrease bias and may even increase it (Elwert and Winship 2014; Pearl 2010; Steiner and Kim 2016). Thus, matching treatment and control units on more covariates does not imply that more bias is reduced!

Subject-matter theory helps in selecting covariates. One should avoid matching covariates that only affect the treatment but are otherwise (nearly) unrelated to the outcome (instrumental variables), or covariates (colliders) that are not confounders but are affected by two unobserved variables—one affecting the treatment, the other the outcome. Including these covariates in a PS model (or directly matching on them) either amplifies any remaining bias or induces collider bias. Thus, researchers need to carefully think about construct domains and single constructs that might have confounded the causal effect.

Reliable Covariate Measurement Once the constructs for establishing strong ignorability have been determined, they need to be reliably measured with regard to the actual data-generating process. Their unreliable measurement would result in a violation of the unconfoundedness assumption. For example, students’ ability is very likely a confounder because ability affects both the retention decision and reading scores. Thus, if ability measures like the pretest of the reading outcome are unreliable measures of the latent ability construct, they cannot remove all the bias. For a single confounder, each 0.1 decrease in the confounder’s reliability (e.g., from 0.9 to 0.8), results in 10% less bias removal (Steiner et al. 2011). Though analytic techniques to correct for measurement error exist, they generally rely on strong modeling assumptions or detailed knowledge about the measurement error. In the case of multiple confounders, the relation between covariate reliability and bias reduction is more complex. A decrease in the reliability of one or multiple covariates no longer implies that less bias is removed (Aiken and West 1991; Kim and Steiner 2017).

Generally, since measurement error is inevitable in practice, important single constructs should always be measured with multiple items. The measurement of strongly related constructs (e.g., teachers' assessment of student ability in addition to the pretest of the reading score) may at least partially compensate for unreliability and help to further decrease the confounding bias.

Choice of Matching Technique After the covariates have been carefully selected and measured, a specific matching technique needs to be chosen. Multivariate matching based on the Mahalanobis distance is suitable when the number of covariates is rather small. With a large number of covariates and observations, PS techniques—matching, stratification, or weighting—are preferable. Standard one-to-one matching estimates the ATT. In contrast, PS-stratification and PS-weighting can directly estimate both the ATT and ATE, and typically produce more efficient estimates because all (overlapping) units are used. But, as discussed for Hong and Raudenbush's (2006) study, PS-stratification may leave some residual bias due to the heterogeneity of units within strata. PS-weighting may result in unstable estimates when outliers with extreme PSs are present.

What Matters in Practice? The factors considered thus far—covariate selection, reliable measurement, choice of matching technique—are not equally important for bias reduction in practice. Most important is covariate selection, which directly relates to the strong ignorability assumption. If one fails to identify a set of covariates that removes all the confounding, valid causal inference is almost impossible even if all covariates are reliably measured and an optimal matching strategy is chosen. Demographical covariates like sex, age, or race rarely remove a major part of the bias (Steiner et al. 2010). If an appropriate set of bias-removing covariates has been identified, their reliable measurement is the next important factor. Steiner et al. (2011) demonstrate how unreliability in covariates creates bias in effects estimates, irrespective of the choice of PS technique. The choice of a specific matching technique is, in general, the least important factor, provided the observed covariates have been well-balanced. Especially when combined with additional covariance adjustments (doubly robust methods), matching, stratification, and weighting produce very similar results (Schafer and Kang 2008; Shadish et al. 2008).

CHAPTER SUMMARY

The challenge of making causal inferences from observational data consists not only in mastering the statistical matching techniques but also in meeting the causal assumptions. In this chapter, we explained what the causal estimands are and how they can be identified and then estimated via PS-matching, PS-stratification, or PS-weighting in single-level or multi-level settings. We also highlighted that the selection of baseline covariates and their reliable measurement is much more important than the choice of a specific matching technique. Sensitivity analyses can be very helpful in assessing the effect of unobserved confounding.

Recommended Readings

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.

Austin's paper provides a focused introduction to propensity score methods (not the general matching methods). Both conceptual and practical aspects regarding the propensity scores are well discussed in plain language.

Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 1, pp. 237–259). New York: Oxford University Press.

Steiner and Cook provide a comprehensive overview of matching strategies. Many technical details not thoroughly discussed in this chapter are well introduced in the paper.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.

Stuart provides a systematical and accessible review for applied researchers. Besides the theoretical aspect, the paper also emphasizes the practical aspect of implementing matching methods. The way to diagnose the matching quality and the extensive list of various software for matching methods may be useful for applied researchers.

REFERENCES

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424.

- Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, *24*, 295–313.
- Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya: The Indian Journal of Statistics*, *35*(4), 417–446.
- Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology*, *40*, 31–53.
- Griffin, B. A., Ridgeway, G., Morral, A. R., Burgette, L. F., Martin, C., Almirall, D., Ramchand, R., Jaycox, L. H., & McCaffrey, D. F. (2014). *Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) Website*. Santa Monica: RAND Corporation. Retrieved from <http://www.rand.org/statistics/twang>. Accessed 28 July 2017.
- Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*, 609–627.
- Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric pre-processing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960.
- Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*(475), 901–910.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York: Cambridge University Press.
- Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*(4), 523–539.
- Keele, L., & Pimentel, S. (2016). *matchMulti: Optimal multilevel matching using a network algorithm*. R package version 1.1.5. <https://CRAN.R-project.org/package=matchMulti>
- Kelcey, B. (2011). Assessing the effects of teachers' reading knowledge on students' achievement using multilevel propensity score stratification. *Educational Evaluation and Policy Analysis*, *33*(4), 458–482.
- Keller, B., Kim, J.-S., & Steiner, P. M. (2015). Neural networks for propensity score estimation: Simulation results and recommendations. In L. A. van der Ark, D. M. Bolt, S.-M. Chow, J. A. Douglas, & W.-C. Wang (Eds.), *Quantitative psychology research* (pp. 279–291). New York: Springer.
- Kim, J.-S., & Steiner, P. M. (2015). Multilevel propensity score methods for estimating causal effects: A latent class modeling strategy. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & S.-M. Chow (Eds.), *Quantitative psychology research: Proceedings of the 79th annual meeting of the psychometric society* (pp. 293–306). New York: Springer.

- Kim, Y., & Steiner, P. M. (2017, March). *The mechanics of omitted variable bias and the effect of measurement error*. Invited talk at the 2017 ENAR Spring Meeting of the Eastern North American Region International Biometric Society. Washington, DC.
- Kosanke, J., & Bergstralh, B. (2004). *gmatch*: Match 1 or more controls to cases using the greedy algorithm. Retrieved from <http://www.mayo.edu/research/documents/gmatchsas/doc-10027248>; *vmatch*: Match cases to controls using variable optimal matching. Retrieved from <http://www.mayo.edu/research/documents/vmatchsas-05-14-14/doc-20094471>. Accessed 26 July 2017.
- Lara, B., Mizala, A., & Repetto, A. (2011). The effectiveness of private voucher education: Evidence from structural school switches. *Educational Evaluation and Policy Analysis*, 33(2), 119–137.
- Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources state NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, 34(2), 209–231.
- Leuven, E., & Sianesi, B. (2003). *Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Statistical Software Components S432001, Boston College, Department of Economics.
- McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9, 403–425.
- Pearl, J. (2010). On a class of bias-amplifying variables that endanger effect estimates. In P. Grunwald & P. Spirtes (Eds.), *Proceedings of the 26th conference on uncertainty in artificial intelligence* (pp. 425–432). Corvallis: AUAI Press.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer-Verlag.
- Rosenbaum, P. R. (2005). Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *The American Statistician*, 59(2), 147–152.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1), 185–203.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.
- Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371), 591–593.
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, 26, 20–36.

- Schafer, J. L., & Kang, J. D. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279–313.
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*(484), 1334–1344.
- StataCorp. (2015). *Stata treatment-effects reference manual: Potential outcomes/counterfactual outcomes*. College Station: Stata Press. Retrieved from <http://www.stata.com/manuals14/te.pdf>. Accessed 26 July 2017.
- Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 1, pp. 237–259). New York: Oxford University Press.
- Steiner, P. M., & Kim, Y. (2016). The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases. *Journal of Causal Inference, 4*(2). DOI: <https://doi.org/10.1515/jci-2016-0009>. Advance online publication.
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods, 15*(3), 250.
- Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. *Journal of Educational and Behavioral Statistics, 36*(2), 213–236.
- Steiner, P. M., Kim, J.-S., & Thoemmes, F. J. (2012). Matching strategies for observational multilevel data. In *Joint statistical meetings proceedings* (pp. 5020–5032). Alexandria: American Statistical Association.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*(1), 1–21.
- Thoemmes, F. (2012). *Propensity score matching in SPSS*. Unpublished manuscript. Retrieved from <https://sourceforge.net/projects/psmspss/>. Accessed 26 July 2017.
- Thoemmes, F., & West, S. G. (2011). The use of propensity scores for nonrandomized designs with clustered data. *Multivariate Behavioral Research, 46*(3), 514–543.
- West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years. *Journal of Educational and Psychological Measurement, 100*(4), 727–740.
- Wyse, A. E., Keesler, V., & Schneider, B. (2008). Assessing the effects of small school size on mathematics achievement: A propensity score-matching approach. *Teachers College Record, 110*(9), 1879–1900.



Using Quantitative and Qualitative Methods to Study the Content and Effects of Curriculum Materials

Morgan Polikoff, Shauna E. Campbell, and Shira A. Korn

Curriculum materials are an important educational input that has received limited attention in education policy research. High-quality quantitative research (e.g., Agodini et al. 2010; Bhatt and Koedel 2012; Koedel et al. 2017) suggests that choice of curriculum materials may have substantial impacts on student achievement (a tenth to a fifth of a standard deviation in most studies). However, it is difficult to identify the true effects of curriculum materials on student outcomes because of limited data available to conduct high-quality quantitative studies.

This material is based upon work supported by the National Science Foundation under Grant No. 1445654, the William T. Grant Foundation, the Bill & Melinda Gates Foundation, and an anonymous foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

M. Polikoff (✉) • S. E. Campbell • S. A. Korn
Rossier School of Education, University of Southern California,
Los Angeles, CA, USA
e-mail: polikoff@usc.edu; shaunaec@usc.edu; shira.korn@usc.edu

The selection, implementation, and effects of curriculum materials are important issues for scholarly research because high-quality curriculum materials can potentially produce measurable differences in student achievement. For school and district leaders, curriculum materials are likely less controversial and lower in cost than human resource interventions—such as staffing changes or performance evaluation. Curriculum interventions also have the potential for very high benefit/cost ratios, given the low marginal cost between curriculum options (Chingos and Whitehurst 2012).

We begin with a review of relevant literature on the adoption, use, and effects of curriculum materials. Throughout this chapter, we use the term “textbook” to refer to the traditional paper volume that contains a comprehensive subject-specific curriculum—textbooks have been the focus of virtually all of the research we discuss below. “Curriculum materials” includes textbooks, as well as the ever-expanding landscape of resources that teachers call upon to implement their lessons. These materials include supplemental reading materials, open online educational resources (e.g., EngageNY), videos (e.g., Khan Academy), diagnostic materials, software (e.g., Google Apps for Education), and websites where teachers share lessons and resources (e.g., Pinterest, TeachersPayTeachers). Importantly, much of the research on textbooks was published before the widespread availability of these supplemental materials that are largely disseminated online. Therefore, we see the study of curriculum materials as being particularly relevant for future scholars seeking to understand the changing nature of classroom resources available to teachers.

We describe some methods that can be used for researching the adoption, implementation, and effects of curriculum materials, as well as the challenges associated with each method. We describe quantitative methods such as propensity score models, value-added models, and multilevel modeling for estimating achievement effects. We also describe how qualitative data can supplement quantitative data and provide a more complete understanding about the use of curriculum materials. These methods can be used to investigate a variety of policy-related questions about the adoption of curriculum materials, alignment of the classroom curriculum to the intended curriculum of the standards, and fidelity of implementation of curriculum materials. We suggest areas for future research based on the data that currently exist as well as the changing nature of the materials themselves.

REVIEW OF RELEVANT LITERATURE

Opportunity to Learn Much of the research on curricular materials is grounded in the opportunity to learn literature (see McDonnell 1995). Carroll (1963) introduced the concept of opportunity to learn (OTL) as one of several factors that affect student outcomes, along with aptitude, perseverance, and quality of instruction. Students' opportunity to learn a given topic is determined by teacher- and school-level decisions. Factors contributing to OTL include the length and content of classes, the order in which lessons are taught, the amount of time given to teach a specific topic, and the content to which students are exposed through curriculum materials (Kurz 2011).

Curriculum materials affect student learning through the content included in the materials (e.g., whether and how two textbooks differ in their coverage of multiplying fractions (Polikoff 2015)) and through teacher implementation of the materials (e.g., how teachers choose to implement the fraction multiplication lessons of the books (Remillard 2005)). OTL has been used as a central argument in court cases ruling that students from low socioeconomic backgrounds should have access to the same quality of curriculum materials as their higher-SES peers (e.g., *Eliezer Williams et al., v. the State of California et al.*, 2000). In these court decisions, curriculum materials have been cited as a key policy lever associated with student learning.

Opportunity to learn can be defined and measured across multiple levels of curriculum. Some research (e.g., Porter and Smithson 2001) differentiates between the *intended* curriculum and the *enacted* curriculum. The intended curriculum refers to the skills that students are expected to know as determined by factors such as state standards, district pacing guidelines, or the content of a teacher's guide. The enacted curriculum refers to the curriculum that is actually delivered to students in class via instructional materials and teaching methods. Under the OTL framework, curriculum materials contribute to student learning opportunities as they form a bridge between the intended and the enacted curriculum (Kurz 2011). For example, the Common Core State Standards (or a closely-related or renamed version) represent the current intended curriculum in many states and establish the content and skills that students must master in each grade. Publishers then produce new materials that are aligned to these standards and that help teachers interpret the standards and determine how they deliver instruction.

The Adoption of Curriculum Materials Curriculum materials can be adopted at the state, district, school, or even classroom level, depending on state and local laws. This means that to study curriculum selection, implementation, and effects, researchers need to have an understanding of the type of adoption practiced in the region being studied. Roughly half of the states publish a list of approved textbooks evaluated by the state department of education and determined to meet the needs and standards of the state. In some states, such as Mississippi, districts are required to adopt from the approved list of materials, while in states such as California and Texas, districts are encouraged to adopt materials from the list but are also free to make local decisions about curriculum material adoptions.

Previous research on the adoption of curriculum materials focused on the state-level processes related to evaluation and selection of textbooks (e.g., Finn and Ravitch 2004; Stein et al. 2001). Relatively little research exists on the adoption decisions made at the district or school level, especially if we only consider research from the standards era. The most comprehensive study (Zeringue et al. 2010) identified several factors that matter to district leaders when making adoption decisions, including perceived teacher support, evidence of quality of materials, and resources available for purchase and implementation. We recently conducted interviews with district leaders in California and found these factors to be important in district-level decision-making, with teacher voice being an especially influential factor. Other factors that seem to matter include the support of the county office of education, the cost of materials distributed over the length of the adoption cycle, and district characteristics such as the percentage of English language learners and quality of technology. Our interviews also revealed that the previous literature on textbook adoptions needs to be revisited as districts consider options such as digital or hybrid curricula and open educational resources available online.

The Implementation of Curriculum Materials Curriculum materials offer teachers suggestions regarding the sequence and pacing of lessons, the scope of the subject matter covered in a year, and the strategies to be used for instruction. The choices that teachers make regarding the implementation of the lessons shape the enacted curriculum—the curriculum that students actually experience in the classroom. In short, textbooks and other curriculum materials are an important factor in the relationship between the intended and the enacted curriculum, and therefore are a contributing factor to differences in student opportunity to learn (Schmidt et al. 2001).

Not every teacher uses traditional textbooks as their primary source of instruction, but it is clear that teachers do consider these materials in shaping their lessons. A recent national sample of teachers from the American Teacher Panel indicated that, while they use a variety of sources to plan and implement lessons, most teachers still use traditional textbooks for at least some, if not all, of their planning (Opfer et al. 2016). From this survey, we gain two important insights regarding teacher use of materials during the Common Core era. The first is that textbooks remain an influential factor in teacher planning and are thus a relevant policy lever. The second is that there is a proliferation of non-traditional materials that should be researched more extensively. These non-traditional materials include open online educational resources, interactive or hybrid textbooks (with features online and in print), and websites where teachers share materials with other teachers (either for free or for a charge).

Because textbooks remain widely used by teachers, they represent an important policy lever that can be used to improve student achievement. That is, improving the quality of textbooks or the processes by which textbooks get into teachers' hands could materially improve instruction and student learning due to the ubiquity of these materials. The textbook alone, however, cannot ensure high-quality instruction, as the teacher must interpret, design, and implement lessons. This process of teacher participation creates variations in resource implementation (Remillard 2005). Such variations occur as a teacher may implement the lessons with complete fidelity, may use the textbook as one of many available resources, or may interpret and interact with the text in a co-constructive relationship (for a review of teacher use of curriculum materials, see Remillard 2005). The relationship between teacher and curriculum materials is one that requires further research, especially in the context of the emergence of open online resources and other technological advances. Teacher implementation of materials is an important variable for researchers studying the effects of textbooks on student outcomes.

The Impact of Curriculum Materials In spite of the substantial discretion teachers have when implementing materials, recent rigorous quantitative evidence suggests that simply adopting one book over another can produce meaningful effects on student achievement. One recent large-scale study (Agodini et al. 2010) randomly assigned elementary mathematics textbooks to schools, finding statistically significant differences in their

effects on student learning. These differences were 0.10 to 0.15 standard deviations in magnitude, enough to move students from the 50th to the 57th percentile, depending on grade. Three recent studies use matching methods (which we describe below) to investigate the impact of elementary mathematics textbooks in California, Florida, and Indiana; again, each study found that at least one textbook outperformed the others (Bhatt and Koedel 2012; Bhatt et al. 2013; Koedel et al. 2017). Together, these four recent studies provide compelling evidence that textbooks matter for student learning.

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY

To date, there has been little comprehensive analysis of widespread patterns and effects of textbook usage due to a dearth of available data (Chingos and Whitehurst 2012). As mentioned above, there are three recent matching studies, but these are all in elementary mathematics, and they are only in three states. Clearly, there is a need for research in other subjects, other grades, and other states. Table 10.1 presents some suggested research questions and data sources aligned to the three principal areas of research related to curriculum materials: adoption, implementation, and effects of curriculum materials.

There are at least three main topics that merit continued scholarly investigation. First, regarding adoption, qualitative analysis of how adoption decisions are made at a district level can illuminate the factors prioritized in the selection process. Such insights could help researchers and policymakers to provide evidence to districts that might improve their selection decisions. Second, regarding implementation, qualitative or survey analysis of teacher implementation can provide insight into how the enacted curriculum may differ from the intended curriculum, which could enhance our understandings of students' access to the curriculum. Third, regarding impacts, collection, and analysis of textbook adoption data at the school or district level (i.e., cataloging which books are purchased in which districts and schools across entire states) can inform policymakers about which resources are associated with higher student achievement and illuminate how access to high-quality materials may differ between classes, schools, and districts. Ultimately, understanding how textbooks are adopted and implemented, and the impact of these decisions on student achievement outcomes can create a more nuanced understanding of student learning opportunities, as outlined under the OTL framework.

Table 10.1 Illustrative research questions and data sources related to the study of curriculum materials adoptions

<i>Research area</i>	<i>Suggested research questions</i>	<i>Suggested data sources</i>
Curriculum materials adoptions/ evaluations	What are the most important criteria for district leaders making adoption decisions? How do criteria and processes differ among privates, charters, magnets, and traditional public schools? How do differences in processes result in differences in adoption decisions? What role do administrative and teacher leaders play in the adoption process? How do textbook adoption processes differ in states that adopt at the state level versus states that adopt at the district or school levels?	Interviews, surveys
Curriculum materials implementation	To what extent do teachers implement a given curriculum with fidelity? How do district leaders affect teachers' implementation of materials? In what ways do teachers supplement their lessons with materials besides official school adoptions?	Teacher logs, classroom observations, interviews, surveys
Curriculum materials effects	What are the effects of textbook adoptions on student achievement in middle and high schools and in other subjects than mathematics? To what extent do curriculum materials effects vary by student demographic characteristics? In what ways does leadership mediate the effects of materials on student outcomes? To what extent do curriculum effects vary between types of curriculum (e.g., constructivist versus traditional; digital versus traditional)?	District- or school-level achievement data District- or school-level materials adoptions District- or school-level demographic characteristics

Analyzing Textbook Adoption Decisions An understanding of textbook adoption decision-making processes is important for situating studies of implementation and effects. For example, one might expect to find more variation in student outcomes in a school district where principals or teachers are allowed to make their own decisions about adopted materials than in a uniform-adopting district. Adoption decisions can take place at

the state, district, school, or even classroom levels, so it is essential for researchers to first identify the appropriate level in the area under study. This information can often be found on state or school district websites, and it can be confirmed through phone calls or emails.

In terms of the adoption processes themselves, researchers can qualitatively investigate the nuances of adoption processes through interviews with district leaders, teachers, principals, members of the state department of education, or any other parties involved in the selection of curriculum materials. A straightforward qualitative analysis of evaluation criteria and adoption decisions can provide comparative data across districts.

Additional scholarship might study textbook adoption decisions through ethnographic methods, such as participating in and observing textbook adoption committee meetings. A staple of ethnographic work is what Geertz (1973) referred to as “thick description,” which aims to uncover not only *what* happens but also how relevant actors and observers *interpret* what happens. Studying textbook adoption decisions through an ethnographic lens could thus provide insights into how individuals make meaning of curriculum materials, including how actors perceive their own curricular needs, how they evaluate materials based on these needs, and how they use these materials accordingly. Ethnography can also shed light on how institutional structures and cultures dictate curriculum evaluation and usage within the classroom. Ultimately, a deep and nuanced analysis of how decisions about curriculum materials are made would allow researchers and policymakers to provide targeted assistance to aid in this process.

Studying the Implementation of Textbooks There are two main questions of interest under the broad topic of textbook implementation. First, how do teachers actually use the materials that are adopted, and how do they supplement these materials? Second, what is the content of these materials, and how does it affect student opportunity to learn?

Analyzing Teacher Use of Materials We propose studying teacher use of materials through both qualitative case study research and large-scale survey research. Case studies provide extensive and holistic descriptions of a singular unit (Merriam 1998) and allow researchers to explore how occurrences in this unit are influenced by context (Stake 2005). Thus, studying teacher use of materials through case study would allow

researchers to explore how teachers are utilizing the materials they are given and how their pedagogical decisions are influenced by the political, social, and organizational culture within their school as well as their prior training and beliefs. For example, the degree to which teachers have autonomy over their classroom, their access to common planning, and their sharing of resources could profoundly impact their use of textbooks within the classroom. Understanding these social and organizational structures within districts, schools, and departments and how they impact teacher pedagogy requires in-depth analysis and experiential knowledge, which are well suited for a case study approach.

A limitation of case study research is its generalizability. Thus, to explore how common practices are in other schools, districts, and states, we propose using large-scale survey research. We have begun to explore these issues with state-representative samples of teachers, asking them detailed survey questions about their use of curriculum materials (see Kane et al. 2016, for an earlier version of this work). We will use these survey responses both to describe curriculum use at scale and to construct predictors to explore variation in textbook effects on student outcomes.

Additional research might focus how materials are used in states, districts, and schools from an equity perspective. For example, how does the use of curriculum materials differ within schools based on a student's prior achievement or status as an English Language Learner (ELL) or Special Education (SPED) student? Topics of consideration might include whether remedial students are equally likely to be exposed to materials that require high cognitive demands and whether ELL and SPED students are being provided materials that are appropriate for their learning needs. Ultimately, whether or not students with different background characteristics or special designations have access to high-quality and academically appropriate materials is an important policy question that we can begin to address by analyzing the implementation of the materials by teachers and schools.

Analyzing the Content of Textbooks In order to gauge how textbook adoption choices influence students' opportunity to learn, we can analyze the content of the materials. Using the Surveys of Enacted Curriculum (SEC) framework (Porter 2002), we can quantitatively code entire textbooks, creating an index of how thoroughly the book covers standard topics, and the level of rigor with which each topic is addressed (for an

example, see Polikoff 2015). Variants on the SEC methods have been in use for over 20 years (see Porter 2002 for a history; for other examples see Porter et al. 2007; Polikoff 2012).

There are existing SEC frameworks in mathematics and English language arts that have recently been revised to study implementation of Common Core and other content standards (see Porter et al. (2011) for an analysis of the Common Core using the SEC and www.c-sail.org for recent work to update the SEC languages). These define content at the intersection of specific topics and levels of cognitive demand. Independent coders use the frameworks to assign topics and levels of cognitive demand to sections of text. The results are then averaged across raters to arrive at a complete representation of the content in the textbook or other curriculum material.

Once the coding is complete, the data can be used to calculate alignment indices or other descriptive measures of textbook content. For example, we can calculate the alignment of a book with a set of content standards, indicating the book's overall coverage of the topics and cognitive demands emphasized in the standards. We can also report the proportion of each book's content on focal content strands or at lower and higher levels of cognitive demand. Finally, we can compare alignment and other descriptive indices across textbooks. To date, only mathematics materials have been studied using an SEC content analysis. It is possible that science, ELA, or history/social studies materials could be coded, though these are somewhat more complex than mathematics due to the nature of content in the subjects.

Recent research demonstrates that the content analysis of mathematics textbooks (Polikoff 2015) recommends specific strategies for simplifying the content analysis procedures (Polikoff et al. 2015). These papers use the SEC framework to measure the alignment of several popular math textbooks to the Common Core State Standards for math, on the principle that better aligned materials offer students a better opportunity to learn the standards. This work found that even the most popular textbooks claiming alignment to the Common Core math standards were not well aligned, particularly with regard to the cognitive demand required of the standards. Given the role of textbooks in influencing teachers' instruction, especially during the early years of a standards transition, these kinds of content analyses can shed important light on likely areas of alignment and misalignment in teachers' instruction. Furthermore, they can point

the way toward areas of needed supplementation. Finally, these measures of alignment or content coverage can provide a measure of the relative quality of textbooks—at least with regard to their coverage of the standards—that teachers and district leaders may wish to use when making adoption decisions.

Studying the Impact of Curriculum Materials Probably the question of greatest interest to policymakers is the impact of textbook choices on student learning. Here, we briefly outline methods for (a) gathering the necessary data and (b) conducting the actual impact analysis.

Collecting Data on Textbook Adoption Patterns The first step in identifying the impact of curriculum materials is collecting the necessary data to conduct a secondary data analysis. The preferred approach will vary based on the state and what data are available, as most states do not make textbook adoption data available in any form. California is among just a handful of states that do provide publicly available information on schools' adopted textbooks. Every school in the state is required to publish a yearly School Accountability Report Card (SARC) that includes information about the quality and availability of textbooks (typically this means that titles and adoption years are provided). For our work (Koedel et al. 2017), we have manually downloaded and recorded the SARC information for every school serving elementary and middle grades ($n \sim 7600$) for the years 2012–13 to 2015–16. This is a time-consuming process because there is no standardized format for schools to use when entering textbook information. The challenges of this process are described in detail elsewhere (Koedel et al. 2017).

A second option is to use state-level purchase records in the states that keep them. While we know of no definitive list of such states, we are aware that Louisiana, New Mexico, Tennessee, and Texas all track this information. For instance, the Texas Education Agency (the state department of education) holds records on all curricular materials purchased at each district site, and the data are updated daily. These data are recorded at the district level and are the most comprehensive records of curriculum materials used because they include everything from traditional textbooks to online supplemental programs to novels—anything that districts use money to purchase. Similar data are also available in New Mexico and Tennessee, and some other states we are not aware of may also track purchase data in this way.

A third option is to collect the data by contacting individual school districts. Surveys can be used, but these require incentives to obtain even moderate response rates. In our work, after attempting and failing at population surveys in Illinois, New York, and Florida, we found that districts have responded to a request for information filed under state Freedom of Information Acts/Laws (FOIA/FOIL). State-specific templates for FOIA letters are readily available online and can be mailed to school or district offices. This method is likely to yield a high response rate, as districts are required to provide any existing records containing the requested information (though they are permitted by law to charge for expenses associated with fulfilling the request, very few districts do this). However, even with clear instructions, there is a great deal of variability in the quality and completeness of information provided by individual schools and districts. The FOIA method could be used in states where data on purchases and adoptions are not readily available any other way, but it should be used sparingly as it is seen as intrusive and confrontational. The method is also time consuming because it requires careful tracking of contacts, and the information reported is not in any standardized format. All of these methods of data collection share the burden of being time consuming, though they have the potential to provide a nearly complete picture of the formal textbooks of record in schools and districts in a state.

Analyzing the Impact of Textbook Adoptions on Student Achievement Once the data on textbook adoptions have been collected, there are multiple analytic options that can be used to attempt to identify the causal effects. For a question such as “Which of the most common elementary school mathematics textbooks has the most positive effect on student achievement?” the kinds of matching methodologies used in Koedel et al. (2017) are appropriate. This research uses propensity score models and longitudinal school-level test data to match schools on a variety of demographic and achievement variables thought to be related to textbook adoption decisions (e.g., school-level demographics, poverty, and geographic variables) and track subsequent achievement trends. Koedel and colleagues’ most recent paper uses three matching techniques—kernel matching, restricted ordinary least squares, and remnant-based residualized matching (the methods are described in great detail in the paper). If evidence can be provided that the key assumption of conditional independence (that there are no unobserved variables related to both the textbook adoption decision and student achievement) is met, these methods can

produce causal estimates of textbooks on student outcomes. These conditional independence assumptions are generally explored by first demonstrating balance among the treatment groups on all available covariates and then by conducting falsification tests such as testing for math textbook “effects” in other subjects and looking for effects in years in which they should not exist. All of these methods are described clearly in Koedel et al. (2017), and they have also been applied in two other studies (Bhatt and Koedel 2012; Bhatt et al. 2013). These methods could also be used to investigate heterogeneous effects across student subgroups—Koedel et al. (2017) demonstrates this for student socioeconomic status.

Similar data could also be used to conduct something like a difference-in-differences or comparative interrupted time series analysis (see Murnane and Willet (2011) for a discussion of the DD and CITS methodologies), though we know of no instances where this has been done.

Another approach involving secondary data uses student-level data to conduct value-added analyses at either the school or teacher levels. For example, Kane et al. (2016) calculated value-added models to estimate the impacts of individual teachers on student achievement (any standard value-added model or student growth percentile could be used (see Koedel et al. (2015) for an overview of value-added models). Then, they related these value-added estimates to a variety of measures of curriculum materials used and curriculum implementation indicators. If these methods are similarly paired with efforts that control for pre-adoption differences in schools, such as by controlling for pre-adoption value-added, they can also identify causal impacts of curriculum materials.

An alternative approach to examining the impacts of textbooks with secondary data uses random assignment to generate unbiased causal effects. Agodini et al. (2010) recruited schools from geographically diverse regions of the country and randomly assigned them to investigate the impact of four of the most widely used math textbooks; they found significant achievement differences among the examined books. There are a number of other random assignment curriculum studies listed in the What Works Clearinghouse, but these studies suffer from many problems that substantially limit their utility. For instance, (a) the vast majority of them pre-date recent standards adoptions, (b) the control condition in many of the studies is underdescribed, (c) many of the studies are very small (just a few schools or classrooms), and (d) large proportions of the studies focus on small-scale curricula rather than core/basal curricula. Finally, there

may be a difference between schools that choose to adopt a curriculum and schools that participate in random assignment studies that may limit the external validity of random assignment studies.

One complication from the existing impact research is that Agodini et al. (2010) and Bhatt and Koedel (2012) found student achievement effects associated with the same textbook, but one found a positive and the other found a negative effect. While this may be seen as problematic (certainly it is problematic from a policy interpretation standpoint), these divergent findings do not imply any problem with methodologies. Rather, these differences may be attributable to differences in the outcome measures used (different state or study-administered tests) or samples (the Agodini et al. sample was more disadvantaged than the statewide sample in Bhatt and Koedel). Further research is necessary to determine how curriculum effectiveness may vary across different populations of students and outcome measures.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

Curriculum materials are a relevant topic for scholars of leadership and policy focused on instruction, given the primary role of these materials in shaping instruction. Thus, even if curriculum materials are not a primary area of a scholar's research, collecting data on the types of materials used, and the manner in which they are used, can be an important contribution. This section offers two broad sets of recommendations for scholars. First, we note the complications that have emerged as textbook markets have changed alongside teachers' use of supplementary and open source materials. Second, we offer thoughts about the most appropriate sets of methods to use, together and independently, to address important questions about the adoption, usage, and effects of curriculum materials. Overall, our work supports previous scholarship in arguing that it is essential to study the processes of policy implementation (in this case, standards implementation) at multiple levels in order to see meaningful improvements in teachers' practice (Knapp 1997).

If trends away from traditional textbooks continue, it will become increasingly important to study what resources are available to teachers, how leaders help teachers navigate these materials, and how these resources are shaping teachers' instructional practices. The methods proposed here give us a path forward for understanding textbook use and effects, but as materials change, so too will the research needs. The availability of resources

means that teachers are supplementing the traditionally adopted curriculum with materials they believe are specifically suited to their students' needs. We know little about widespread use of curriculum materials to begin with, and the degree of variation between classes is presumably affected by the expanding market for supplemental materials. We expect that future research in the area of curriculum materials can shed insight into the ways in which advances in technology are affecting the implementation of both traditional and non-traditional materials. This research can be helpful in creating a fuller picture of the enacted curriculum, including its alignment to the intended curriculum and its variation between classes and schools. However, these questions will be difficult to answer due to the even greater difficulties associated with trying to learn what teachers are doing on a day-to-day basis.

To gain a more comprehensive perspective of teaching and implementation, we recommend using both quantitative and qualitative measures of classroom instruction. Prior research has quantified classroom instruction using measures such as teacher logs (Rowan et al. 2004), and an analysis of the types of activities in which students were engaged in the classroom (Tarr et al. 2013). Any research tool that allows researchers to go into the classroom and analyze the content, duration, and quality of lessons would be helpful to the body of research on curriculum materials. It would also be useful to talk to teachers and understand why they are supplementing their traditional textbooks, how they find and evaluate supplemental materials (including the role of school leaders), and what effects supplementation has on the coherence and quality of their instruction. These questions are best answered qualitatively through rigorous case study methods.

There is also work to be done on the study of computer-based curriculum materials, about which we know little. Who adopts these materials, how do they differ from traditional textbooks, and what are their effects on instruction and learning? Which websites are the most popular, and how widely used are they? Are the materials on some sites better than others? Even within a website, there may be large variations in the quality and alignment of lessons—do these variations relate to teachers' likelihood of selecting the materials? Furthermore, there is the simple issue that it is no longer the case that traditional textbooks are “necessary,” since teachers or districts could pull together curricula from online sources. Would this be a good idea? Are districts that use created “units of study” better off? Would moving away from textbooks be a cost-effective solution? And do all schools have the infrastructure necessary to support these trends?

There are other aspects related to textbooks that also merit additional research. Pacing guides and other resources are frequently provided to teachers by schools and districts, but we know little about these materials. They can be analyzed using some of the same methods that we use to study curriculum materials. For example, content analysis could be used to measure alignment of district-provided pacing and implementation materials to state standards or to teachers' enacted curriculum. Again, pacing guides represent another mediating variable separating the standards policy from teachers' actual instruction.

Another important topic is the process by which textbooks get into schools. Researchers could add to the body of literature on materials adoption by looking at the processes by which schools and districts adopt materials and how these differ by key types of schools (e.g., charter schools and magnet schools). This is likely a key leverage point for getting better materials in the hands of teachers, but we know very little about how these decisions are made.

A final recommendation is to study the role of instructional leadership in helping teachers implement curriculum materials. Teachers likely need support to implement new materials, and leaders undoubtedly play a role in bringing effective professional learning opportunities to teachers. Leaders also play a key role in establishing a coherent instruction vision within a school or district. These recommendations are just a few starting points that researchers could pursue to enhance the scholarly literature on curriculum materials.

CHAPTER SUMMARY

Drawing on the OTL literature, we argue that curriculum materials are an important educational input that affects student learning. Yet, despite the importance of this work, research on the adoption, implementation, and effects of supplementary curriculum materials has been relatively minimal. In this chapter, we review the relevant literature, discuss potential topics and methods for future scholarship, and address the policy implications with respect to each of these categories.

Curriculum materials shape student learning opportunities by creating a bridge between the intended and enacted curricula, ultimately affecting the content to which a student is exposed. In reviewing existing literature on the adoption, implementation, and effects of these materials, we underscore three key findings: (1) districts generally have similar adoption

processes for selecting textbooks, but this is based on very limited evidence; (2) textbooks remain widely used by teachers, but they are increasingly being supplemented by additional materials and are implemented with varying degrees of fidelity; and (3) some books have larger effects on student achievement than others, as demonstrated through rigorous quantitative analysis.

We propose that additional scholarship is necessary to understand how textbooks are being adopted, used, and supplemented in the classroom and how these choices impact student learning. Correspondingly, we suggest: (1) qualitative analysis of textbook adoption decisions through methods such as interviews and ethnographic studies at the school and district level; (2) qualitative analysis of the utilization, supplementation, and content of curriculum materials through case study and survey methods; and (3) quantitative analysis of the effects of curriculum materials on student achievement, using matching, value-added, or experimental methods. Ultimately, we argue that curriculum materials have profound impacts on teacher practice and student exposure to content. As such, they are an important educational input that warrants further consideration by researchers and policymakers alike.

Recommended Readings

Bhatt R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis*, 34(4), 391–412.

Bhatt and Koedel use propensity score methods to estimate the effects of elementary mathematics textbooks on student achievement in Indiana. This paper is the source of the propensity score methods described in the paper and offers a guide for researchers who might want to replicate the work in other states, subjects, or grades.

Polikoff, M. S. (2015). How well aligned are textbooks to the Common Core Standards in mathematics? *American Educational Research Journal*, 52(6), 1185–1211.

Polikoff provides an introduction to and description of content analysis methods for analyzing the alignment of textbooks to content standards. The methods can also be used to describe and compare the content among several sets of curriculum materials.

Remillard, J. T. (2005). Key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246.

This review summarizes what is known about teachers' use of curriculum materials in mathematics. While the work pre-dates recent moves toward online and supplementary materials, the review has great relevance for understanding the relationship of curriculum materials with the enacted curriculum.

REFERENCES

- Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., & Novak, T. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders (NCEE 2011-4001)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education, Institute of Education Sciences.
- Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis, 34*, 391-412.
- Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review, 34*(1), 107-121.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.
- Chingos, M. M., & Whitehurst, G. J. (2012). *Choosing blindly: Instructional materials, teacher effectiveness, and the Common Core*. Washington, DC: Brookings Institution.
- Finn, C. E., & Ravitch, D. (2004). *The mad, mad world of textbook adoption*. Washington, DC: Thomas B. Fordham Institute.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In C. Geertz (Ed.), *The interpretation of cultures*. New York: Basic Books.
- Kane, T. J., Owens, A. M., Marinell, W. H., Thal, D. R. C., & Staiger, D. O. (2016). *Teaching higher: Educators' perspectives on Common Core implementation*. Cambridge, MA: Center for Education Policy Research, Harvard University.
- Knapp, M. S. (1997). Between systemic reforms and the mathematics and science classroom: The dynamics of innovation, implementation, and professional learning. *Review of Educational Research, 67*(2), 227-266.
- Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180-195.
- Koedel, C., Li, D., Polikoff, M. S., Hardaway, T., & Wrabel, S. L. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open, 3*(1), 1-22.
- Kurz, A. (2011). Access to what should be taught and will be tested: Students' opportunity to learn the intended curriculum. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all*

- students: Bridging the gaps between research, practice, and policy* (pp. 99–129). New York: Springer.
- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17, 305–322.
- Merriam, S. B. (1998). Case studies as qualitative research. In *Qualitative research and case study applications in education* (pp. 26–43). San Francisco: Jossey-Bass Inc.
- Murnane, R., & Willet, J. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York: Oxford University Press.
- Opfer, V. D., Kaufman, J. H., & Thompson, L. E. (2016). *Implementation of K–12 state standards for mathematics and English language arts and literacy*. Santa Monica: RAND.
- Polikoff, M. S. (2012). Instructional alignment under No Child Left Behind. *American Journal of Education*, 118(3), 341–368.
- Polikoff, M. S. (2015). How well aligned are textbooks to the Common Core Standards in mathematics? *American Educational Research Journal*, 52(6), 1185–1211.
- Polikoff, M. S., Zhou, N., & Campbell, S. E. (2015). Methodological choices in the content analysis of textbooks for measuring alignment with standards. *Educational Measurement: Issues and Practice*, 34(3), 10–17.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Porter, A., & Smithson, J. (2001). *Defining, developing, and using curriculum indicators*. Research Report Series RR-048. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Porter, A. C., Smithson, J. L., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education*, 20(1), 27–51.
- Porter, A. C., McMaken, J., Hwang, J., & Yang, R. (2011). Common Core Standards: The new U.S. intended curriculum. *Educational Researcher*, 40(3), 103–116.
- Remillard, J. T. (2005). Key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246.
- Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 105(1), 75–101.
- Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.
- Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 443–462). Thousand Oaks: Sage Publications.

- Stein, M., Stuen, C., Carnine, D., & Long, R. M. (2001). Textbook evaluation and adoption. *Reading & Writing Quarterly*, 17(1), 5–23.
- Tarr, J. E., Grouws, D. A., Chávez, Ó., & Soria, V. M. (2013). The effects of content organization and curriculum implementation on students' mathematics learning in second-year high school courses. *Journal for Research in Mathematics Education*, 44(4), 683–729.
- Williams v California*, No. 312236 (Cal. Super. Ct., S.F. County, May 17, 2000).
- Zeringue, J. K., Spencer, D., Mark, J., Schwinden, K., & Newton, M. A. (2010). Influences on mathematics textbook selection: What really matters. *NCTM Research Pre-session*.



Value-Added and Growth Models in Education Research

Cassandra M. Guarino

Value-added and growth models are statistical models that aim to identify and quantify the impact of programs, people, or environments on a specific outcome. In education research, the term typically refers to models that look at the impact of various inputs—generally schools, teachers, or educational programs—on student growth in achievement, in which achievement is generally measured by standardized test scores. In evaluating programs, educational researchers are generally seeking to understand “what works” to improve student learning. Evaluations of interventions, such as a specific type of reading curriculum, or characteristics of the learning environment, such as the presence or absence of a science lab, smaller class sizes, or teacher experience, would fall under the goal of “program evaluation”; value-added models have long been in use for this purpose (e.g., Monk 1994; Meyer 2000). Research aimed at identifying the effectiveness of individual teachers or schools, on the

C. M. Guarino (✉)
Graduate School of Education and School of Public Policy,
University of California Riverside, Riverside, CA, USA
e-mail: cassandra.guarino@ucr.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_11

other hand, is more typically used in policies related to accountability or performance evaluation, and this use of value-added or growth models presents a somewhat different set of statistical issues to address.

REVIEW OF RELEVANT LITERATURE

Value-added models for teacher evaluation have been widely used (Steinberg and Donaldson 2016), but this has been by far their most recent and controversial use and, perhaps for this reason, has generated the greatest amount of technical research on this specific methodological application (e.g., Aaronson, Barrow, and Sander, 2007; Ballou, Sanders, and Wright, 2004; Ballou and Springer, 2015; Braun, 2005; Chetty, Friedman, and Rockoff, 2014; Ehlert, Koedel, Parsons, and Podgursky, 2014; Goldhaber, Walch, and Gabele, 2014; Guarino, Reckase, and Wooldridge, 2015; Hanushek and Rivkin, 2010; Sass, Semykina, and Harris, 2014; Isenberg and Walsh, 2015; Johnson, Lipscomb and Gill, 2015; Kane and Staiger, 2008; McCaffrey, Lockwood, Koretz, Louis, and Hamilton, 2004; McCaffrey, Sass, Lockwood, and Mihaly, 2009; Papay, 2011; Reardon and Raudenbush, 2009; Koedel and Betts, 2011; to name only a few). As a result, in this chapter, I refer more frequently to studies that examine models for teacher evaluation than studies that examine models for school or program evaluation. This application is also particularly important given the fact that teacher compensation represents the lion's share of educational expenditures, and given the key link between student learning and effective instruction. However, skepticism has been voiced over the use of these measures for teacher evaluation purposes (see, for example, Amerin-Beardsley (2008), Darling-Hammond, Amerin-Beardsley, Haertel, and Rothstein (2012), and Rothstein (2010) among others), and virtually all value-added researchers urge that these models be applied carefully and cautiously in high-stakes policies.

The chapter describes commonly used value-added and growth models and discusses their strengths and weaknesses. Two basic types of value-added models and a basic type of growth model are discussed in detail. The value-added models covered are dynamic Ordinary Least Squares (OLS) model and empirical Bayes', both of which are commonly used in policy. The growth model covered is the student growth percentile model (Betebenner 2011), sometimes called the "Colorado Growth Model"—so named for the state in which it was first used for accountability purposes.

These models are discussed in relation to their ability to produce fair and error-free measures. In statistical parlance, these properties are called unbiasedness and precision. There are several challenges to estimating these models, and I cover key sources of bias and noise.

There are steps that can be taken to address both these potential problems. These are discussed in this chapter. As one might imagine, no one model is perfect, though some are clearly better than others for particular purposes. This chapter, through its discussion of the strengths and weaknesses of each approach, concludes with recommendations for modeling choices for each goal, whether it be the evaluation of teachers or schools for accountability purposes or the evaluation of particular programs or other types of inputs for school improvement.

The Theory Behind Value-Added and Growth Modeling The common sense theory behind value-added or growth modeling is that a student's learning growth over the course of a year is produced by the efforts of his or her teacher, school, family, and peers, as well as his or her family environment and individual traits, such as ability, motivation, or study habits. In educational research, we attempt to model this process in a way that allows us to determine the contribution of these efforts to learning in a quantifiable way using statistics. When we fit this common sense conceptual model into an equation that can then be transformed into a statistical model, it generally looks something like the following:

$$A_{it} - A_{i,t-1} = \theta E_{it} + \omega\beta b_i + c_{it} + \varepsilon x_{it}, \quad (11.1)$$

where A_{it} is the achievement of an individual student i at time t —with t signifying “time” at the end of the year and $t-1$ being the end of the prior year (thus $A_{it} - A_{i,t-1}$ represents the learning gain between t and $t-1$)—and where E_{it} represents the contribution of the educational inputs (e.g., the student's teacher or school or a particular educational program the student may be exposed to) in time period t , x_{it} is a set of student, family, and peer characteristics that apply in this time period, c_i represents all characteristics of the student or his or her family that are stable over time, such as ability or motivation, etc., and ε_{it} represents all other unobserved factors contributing to the difference in achievement between time periods t and $t-1$. A more complete and technical explanation of the structural foundations of the value-added model can be found in Guarino, Reckase, and Wooldridge (2015) and Todd and Wolpin (2003).

This type of theoretical framing of the learning process—sometimes referred to as an education production function—lends itself to statistical

analysis and enables us to estimate the separate contributions—i.e., the value added—of the various inputs into the learning process. One can estimate the equation directly using measures of achievement, teacher and school inputs, and family characteristics. We don't generally know how to measure the stable characteristics of the individual student—the c_i —but we can assume that variables that we can account for, such as special education status and gender, for example, proxy for these characteristics well enough. Additionally, we can use a longitudinal dataset of two or more periods to eliminate them, using techniques such as first-difference or fixed-effects specifications.

In this model, we can see that the dependent variable represents the difference between current and prior achievement—thus, the learning that has taken place over the year. This feature of accounting for prior achievement—in this case, subtracting it from current achievement—is the defining feature of a value-added model. Accounting for prior achievement in a value-added model can also be accomplished by controlling for prior achievement on the right-hand side of the model as in the following:

$$A_{it} = \lambda A_{i,t-1} + \theta E_{it} + \varpi \beta b_i + c_{it} + \varepsilon x_{it}. \quad (11.2)$$

Placing prior achievement on the right-hand side of the model has many advantages. First, it allows for the fact that students may have forgotten some of what they knew in the prior year. As such, the λ parameter represents the proportion of learning that was retained from the prior year. This flexibility allows for a more realistic model. The formulation in (11.1), with a simple difference score on the left-hand side, assumes that the difference between measured achievement at time t and time $t-1$ adequately captures the amount learned, whereas the formulation in (11.2) allows for the fact that not everything known at the time was carried forward into the next period. This type of model can be referred to as *dynamic OLS (DOLS)* to signify that the model is dynamic—i.e., the lagged (prior-year) dependent variable appears on the right-hand side.

A common refinement to value-added models consists of controlling for more than just one prior achievement score on the right-hand side. Sometimes, researchers include two or more prior achievement scores in the same subject as the dependent variable, and sometimes, they include other scores as well. For example, if the statistical model evaluates mathematics achievement, the dependent variable will be the mathematics achievement

score at the end of the year, and the right-hand-side could include not only the prior achievement score in mathematics but also prior achievement in reading and language arts. The addition of these controls can add more protection against sources of bias and contribute to the goal of better estimating the contribution of particular programs, teachers, or schools in the time period under consideration.

It should be noted that value-added models based on Eq. (11.2) have an important feature that distinguishes them from a different class of value-added models, the feature being the inclusion of the educational inputs E_{it} as observed components of the model. Having prior achievement on the right-hand-side of the model enables us to adjust the estimates of teacher and school effects for correlation with prior achievement and therefore reduce a potential source of bias. Such correlation can occur if students are tracked or matched to classrooms with particular teachers or assigned to particular programs based on their prior test scores. Thus, including these prior test scores on the right-hand side of the equation takes this “assignment mechanism” out of the error term (i.e., it makes explicit an otherwise unobserved factor relating to both the outcome variable and the independent variable of interest and therefore reduces the potential for an important source of omitted variable bias). This is an important feature of the model that is discussed further throughout the chapter.

Several commonly used value-added models, on the other hand, do not explicitly include the educational inputs in the model. Instead, they estimate the following type of equation:

$$A_{it} = \lambda A_{i,t-1} + \omega \beta b_i + c_{it} + \varepsilon x_{it}, \quad (11.3)$$

and, to compute, say, a teacher’s value added, one would take the residuals (i.e., the estimated student-level errors) and average them across students in the teacher’s class. Similarly, one could average these same residuals across students in a school to obtain an estimate of the school’s value added. I refer to this method as the *average residual (AR)* method. Because the observed variables E_{it} are left out of the equation, and the estimates of E_{it} are computed from residuals in a second step, they are not adjusted for observed factors included in the model that affect assignment. In other words, any partial correlation with other variables in the model is not taken out of the estimate.

A popular variant of the AR method is *empirical Bayes*⁷. There are two types of estimates that are generally referred to as empirical Bayes’ (EB)

estimates. One of these takes the averaged residuals computed as above and multiplies them by a “shrinkage” factor based on subsample sizes and particular variance components. The shrinkage process moves the estimates closer to the mean. An example of an EB model composed of shrunken residuals can be found in Kane and Staiger (2008). A more technically accurate use of the term empirical Bayes’ entails the use of Generalized Least Squares to compute the effectiveness estimates (see Guarino, Maxfield, Reckase, Thompson, and Wooldridge 2015, for a complete description and evaluation of the use of empirical Bayes’ techniques in computing teacher value added). Again, this method does not control for observed factors that might affect assignment to “treatment” from the treatment estimates. As such, the AR and EB models produce estimates that are more purely descriptive than DOLS, which aims to purge estimates from some of the correlation with factors tied to nonrandom assignment, such as prior year test scores. DOLS thus moves somewhat closer to a “causal” estimate than either AR or EB.

Growth models grow out of a similar concept to that of the value-added models described above. These models simply seek to measure a student’s growth in learning over a period of time and do not directly try to account for the influence of different programs, teachers, or schools. Conceptually speaking, they link up to the value-added conceptual model represented in Eq. (11.3) by modeling something like the following:

$$A_{it} = \lambda_1 A_{i,t-1} + \lambda_2 A_{i,t-2} + \dots + \lambda_{3k} A_{i,t-3k} + \varepsilon_{it}. \quad (11.4)$$

In other words, growth models control for several prior test scores but generally nothing else.

The growth models in common use, however, are more sophisticated than the representation in (11.4) in their functional form and use a number of statistical techniques to create what is aimed to be a more accurate measure of individual student growth. These models are generally associated with the work of Damian Betebenner (e.g., Betebenner 2011). A growth model, such as the Colorado Growth Model, is based on quantile regression and looks something like the following:

$$Q_A(\tau | A_{i,t-1}, A_{i,t-2}, \dots, A_{i1}) = \sum_{j=1}^{t-1} \sum_k \phi_{ik}(A_{ij}) \beta_{ik} \tau, \quad (11.5)$$

where Q_A represents a conditional quantile¹ of the current achievement distribution conditional on prior test scores and φ_{ik} is a flexible function (using splines that allow for nonlinearity in the functional form) of the prior test scores. A conditional quantile test score can be estimated for 100 quantiles, one for each percentile, although fractional intermediate percentiles can also be estimated. The 100 or more quantile regressions can be run for each unique combination of prior test scores, so students who have more prior test scores available can be included in more quantile regression models. If students are missing some prior test scores, they will have fewer estimated conditional quantiles to choose from. The quantile regressions are generally run for each grade, year, and subject separately. Fitted values for each student from each regression are generated from the above model. Each student's conditional percentile rank is then determined by counting the number of conditional percentiles that result in fitted test scores that are less than the student's actual current test score. So, if a student's actual test score is most similar to the fitted value for, say, the quantile regression at the 20th percentile, the growth percentile rank assigned to the student would be 20. Thus, the quantile regression results are used to place students on a percentile rank distribution corresponding to their growth. Once these student growth percentiles, generally referred to as SGPs, are computed for all students, they can be used in many ways. They can be averaged for all students in a teacher's class (sometimes the class median is also used), and that average can be assigned as the teacher's performance measure.

Although the growth-quantile regression approach was originally developed for the purpose of obtaining a good approximation of student growth using functional forms with a lot of flexibility built in, they are now in popular use for accountability purposes—for example, to compute teacher and school performance—due to the convenience of using them and their intuitive appeal.

The SGP-based measures, when averaged across students taught by particular teachers, for example, permit rankings of teachers by their estimated effectiveness. This is different from most value-added models in one important respect: the value-added models produce measures that tell us *how much* better or worse students performed in one teacher's class versus another teacher's class, not just how teachers rank against one

¹OLS regression estimates the conditional mean of a distribution. Quantile regression estimates conditional quantiles, such as the median or any other specific percentile.

another. This limitation of growth models (i.e., that they do not provide point estimates of the amount of value a teacher adds) makes it so that we also cannot tell how much of a spread there is in the distribution of teacher effectiveness, whereas most value-added models are capable of indicating this. Measures of spread can be informative if we wish to know how much variation there is across teachers—in other words, how much it matters in terms of student achievement to have a relatively effective versus a relatively ineffective teacher. Of course, value-added models can use percentile rank scores rather than scaled scores as the dependent variable, and, in this case, they would be similarly unable to indicate the magnitude of a teacher or school's value added.

*Two Primary Problems in Value-Added and Growth Modeling:
Bias and Error*

As mentioned above, value-added and growth models must attempt to overcome problems of bias and error. Bias can occur in value-added and growth models because students are not always randomly assigned to programs, teachers, or schools. Unless a model can account for nonrandom assignment to the program, teacher, or school it is trying to evaluate, it will likely produce biased estimates in scenarios in which this type of sorting occurs. It is fair to say that a nonnegligible amount of sorting does occur in schools, both public and private. On the one hand, sorting can be harmful, such as when less effective teachers are more likely to be teaching in schools serving low income or minority students with fewer educational resources in the home. On the other hand, sorting can be beneficial, such as when principals carefully match students to teachers based on the specific strengths of particular teachers and their ability to teach certain types of students. In both cases, regardless of the reasons for sorting, bias will result in estimates of teacher effectiveness if they do not somehow account for the sorting mechanism, such as by making sure that both teacher indicators and prior test scores are included in the regression model.

Further sources of inaccuracy are small sample sizes, random temporary factors that might influence student test performance on a given day (e.g., a disruptive noise or a contagious virus affecting the students in the class), and flaws in the quality of achievement measures. With regard to sample sizes, statistical models generally require fairly large “samples” to produce a fairly tight estimate of an effect. Imprecision in estimates can stem from insufficient amounts of information. To continue with the example of

teacher evaluation, it could be problematic to base an estimate of a particular teacher's effectiveness on the performance of a class of, say, 20 students; a well-known property of statistical estimates is that larger samples produce better estimates. To some extent, this can be addressed by incorporating test scores from more than one year at a time in the models—thus a teacher could be evaluated on the performance of two or more different cohorts of his or her students to raise the total number of students upon which a teacher is evaluated.

It is also possible that an unexpected temporary factor, such as a disturbance in the classroom as students are taking a test—the classic example being a disruptive “barking dog” outside the classroom window—can influence test scores in a random way (Schochet and Chiang 2013). Such phenomena introduce measurement error that is randomly distributed and uncorrelated with any of the variables in the model, since it is unlikely that certain types of students or schools are systematically more likely than others to encounter such temporary disturbances.

The third type of problem arises if student test scores are full of measurement error or sensitive to vertical scaling techniques and fail to adequately capture the true amount of learning that students gain from one period to the next. For example, Martineau (2006) showed that shifts in the learning constructs tested over time could distort models for school value added, and Briggs and Weeks (2009) showed that school value-added measures could be sensitive to the method used to create vertically scaled scores.

Comparing Value-Added and Growth Models As we have seen, the approach to calculating value-added and growth models is somewhat different. A key difference is that growth models simply try to quantify a student's growth, and these student growth percentiles can then be averaged across whatever person or entity one is interested in. If we are interested in comparing average percentile growth in a school, we can simply average all the student growth percentiles for the school. If we are interested in comparing average percentile growth for a teacher, we can simply average all the student growth percentiles for the teacher. In theory, researchers interested in evaluating the impact of a particular program could compare the averaged student growth percentiles for students exposed to the program with the average student growth percentiles for students who were not exposed to the program, although growth models are rarely used in program evaluation. In any case, the growth percentile

model produces a purely descriptive measure (e.g., how much change in students' growth, on average, occurred among the students in a particular school or a particular teacher's classroom), and it does not try to take into account any other unobserved differences across students in the school or in the teacher's classroom. These unobserved differences might include systematic nonrandom assignment to specific schools and teachers. In other words, it does not try to distinguish how much growth in a teacher's students is due to the teacher herself and how much is due to other factors. This is important if researchers are trying to evaluate teacher effectiveness for purposes that affect the teachers themselves. It would be unfair to reward or penalize a teacher based on effectiveness estimate that reflect many factors outside the teacher's control and not just what the teacher has been responsible for.

In a paper comparing value-added and growth models used to compute teacher effects, Guarino, Stacy, Reckase, and Wooldridge (2015) used both simulated and real administrative data to see if the two types of models ranked teachers differently or produced different amounts of error in classifying teachers in the bottom quartile of the teacher effectiveness distribution.

The study found that, under simulated conditions in which there was nonrandom sorting of students to teachers, the value-added model that included teacher-fixed effects generally outperformed the growth models on both metrics. They found a high correlation in results between typical value-added and growth models, but found a nontrivial amount of divergence in the ways in which individual teachers were classified by the different models. This difference in performance of the two types of estimators was corroborated in the administrative data, which was obtained from a large diverse state. When the administrative data were parsed into subsets—one that showed evidence of student tracking into classrooms with similarly high- or low-performing students and one that showed no evidence of nonrandom sorting—the study found that the two measures diverged more when there was tracking than when there was no tracking.

For example, they found that 16 to 18 percent of teachers classified in the bottom quartile of the effectiveness distribution by DOLS were classified in the top three quartiles of the distribution by the growth model—the percentage being highest for subset of teachers in schools that engaged in tracking (see Table 11.1 below adapted from the article). The study therefore suggested that, as statistical theory predicted, the choice between

Table 11.1 Fraction of teachers rated in the bottom 25 percent in the initial estimator who are not rated in the bottom 25 percent in another estimator for random and nonrandom grouping schools

	<i>DOLS</i>	<i>EB</i>	<i>SGP</i>
<i>Panel A: random grouping schools</i>			
DOLS	0	0.043	0.165
EB	0.043	0	0.157
SGP	0.165	0.157	0
Teacher/year obs	457	457	457
<i>Panel B: nonrandom grouping schools</i>			
DOLS	0	0.090	0.180
EB	0.090	0	0.203
SGP	0.222	0.218	0.090
Teacher/year obs	1061	1061	1061

Note: Information presented in this table has been adapted from Table 4 in Guarino, Reckase, Stacy, and Wooldridge (2015). Data are from a large, diverse state

Source: Guarino, Reckase, and Wooldridge (2015)

value-added and growth models is most consequential under nonrandom assignment of teachers to students, with the former being preferred in that case. If teachers are in any way to be evaluated on the basis of these measures, and if students are assigned to particular teachers based on their prior test scores (tracking offers some evidence of this), it is safest to use a method such as DOLS that adjusts the estimates for nonrandom assignment based on prior test scores.

Comparing Different Types of Value-Added Models As we have seen, value-added models have some advantages over growth models in reducing bias if the goal is to determine the effectiveness of a teacher, school, program, etc. Guarino, Reckase, Stacy, and Wooldridge (2015) study how well the dummy variable (or teacher-fixed effects) DOLS model performs relative to the average residual approach, as well as models that use gain scores (as in Eq. 11.1), student-fixed effects, and instrumental variables. They investigate how well the different modeling approaches approximate true teacher effects using simulated data constructed under various ways of assigning students to teachers—randomly and nonrandomly—intended to reflect realistic scenarios. They compare the estimates produced by each model to the true teacher effects using Spearman rank correlations. They also compute a measure of misclassification produced by each estimator.

Table 11.2 Rank correlations and misclassification error, under different scenarios, using different value-added modeling approaches

	<i>DOLS</i>	<i>AR</i>	<i>Gain scores</i>	<i>Student-fixed effects</i>	<i>Instrumental variables</i>
<i>Panel A: rank correlations of estimated teacher value added with true teacher value added</i>					
RG-RA	0.87	0.87	0.85	0.57	0.59
DG-PA	0.87	0.83	0.11	-0.44	-0.88
DG-NA	0.87	0.69	0.89	0.57	0.60
<i>Panel B: percentage of above average teachers who are misclassified as below average</i>					
RG-RA	15	15	16	28	27
DG-PA	15	18	47	61	51
DG-NA	15	26	14	21	20

Note: Information presented in this table has been adapted from an excerpt of Table 2 in Guarino, Reckase, and Wooldridge (2015)

Source: Guarino, Reckase, and Wooldridge (2015)

The first is the proportion of above average teachers who are mistakenly classified as below average. Table 11.2 reproduces a few of the main findings of the paper. The simulated assignment mechanisms are as follows: RG-RA represents random grouping of students into classrooms and random assignment of classrooms to teachers. DG-PA represents dynamic grouping of students into classrooms (i.e., based on their prior test score results (tracking)) and positive assignment to teachers (i.e., more effective teachers teach the higher performing classes). DG-NA also represents dynamic grouping of students into classrooms based on their prior test score results, but this time with negative assignment to teachers (i.e., more effective teachers teach the lower performing classes).

Although the particular numbers in Table 11.2 reflect specific assumptions used to model the generation of student test scores in the simulations,² the patterns are clear. The study finds that no one method accurately captures true teacher effects in all types of assignment scenarios and that all measures have some degree of misclassification error. However, the dynamic OLS estimator with teacher-fixed effects tends to be the most robust and bias-resistant estimator across the various scenarios considered. This is again due to the partialing out of the correlation between student

² Simulations represented in Table 11.2 assume geometric decay with $\lambda = 0.50$, as well as specific teacher effect sizes at 0.25 standard deviations in gain scores (see paper for details). 100 replications per scenario are used.

characteristics and their assignment to particular teachers—a strength of the model that comes into play in the nonrandom assignment scenarios. Certain value-added models, such as those that use gain scores, student-fixed effects, and instrumental variables, are particularly risky in the presence of tracking and can struggle to accurately capture a teacher’s value added even in the presence of random assignment.

While I have emphasized differences across models, it is important to acknowledge that the overall correlations across the EB, SGP, and DOLS can be very high. For example, Dieterle, Guarino, Reckase, and Wooldridge (2015) used administrative data from a large state to show that the correlation in teacher value-added estimates derived from a DOLS model and an empirical Bayes’ model can be as high as 98 percent. Thus for research or other non-accountability purposes, it is probably fairly safe to use either strategy, as each produces a similar pattern. However, if rewards are given or sanctions are placed on individual teachers, it is safer to use the model that is less subject to bias. Dieterle et al. also showed that substantial amounts of tracking exist in schools and that the probability of an individual teacher being rated in the top or bottom quintile of the estimated teacher effectiveness distribution will differ depending on the method used to compute value added—more so when tracking is present.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS AND CHAPTER SUMMARY

Given these types of discrepancies in the way different models produce classifications, a key takeaway point is that the choice of methods matters when trying to measure the effectiveness of teachers, schools, or educational programs based on outcomes like student test scores. The choice of a method matters the most when individuals or entities are being ranked and judged on the basis of model estimates and held accountable for the results. Some models, especially growth models, are more descriptive than others, particularly the strongest of the value-added models. Descriptions of growth in classrooms, schools, or other groups are useful, but may not convey as much information on the causal impact of the inputs of interest as certain value-added models are capable of doing. Multiple regression analysis, in its ability to adjust for correlation among observed variables and reduce the potential for omitted variable bias, is available to be used in estimating the effects of teachers, schools, and programs. Models that

include the teacher, school, or program indicators and the variables that might be used in assigning students to them (i.e., the more causal models that control for assignment like DOLS) would thus be a wise choice in policy evaluation, since they mitigate some of the risks of unfairly measuring effectiveness.

These considerations notwithstanding, many other factors make even the estimates from the more causal models far from perfect. Some important factors relating to test score outcomes may still be omitted from these models because they are not observed. The test scores themselves may be unreliable, particularly for certain groups of students, or tests may not accurately focus on what students have learned. Value-added or growth measures may be highly changeable over time, a phenomenon that has been documented for teacher effectiveness measures (Aaronson et al. 2007; Koedel and Betts 2009; McCaffrey et al. 2009; Goldhaber and Hansen 2013; Stacy, Guarino, and Wooldridge 2018).

Therefore, even with the best models, researchers should exert caution in using them and consider them as important pieces of information, but not necessarily a full or stable determination of the impact of these individuals or entities. As we have emphasized, the most controversial uses of the models are for the purpose of teacher accountability, in which issues relating to nonrandom assignment, measurement error, vertical scaling, and small sample sizes can create fairly important reasons to treat the results with caution. The use of the models for school accountability can also be somewhat controversial. While value-added models designed to evaluate schools do not suffer from the small sample size problem, they are similarly subject to issues relating to the nonrandom assignment of students to schools as well as issues relating to measurement error in test scores. A model of school value added that includes school-fixed effects would again be the preferred model and likely produces more credible estimates of school effectiveness than one based on residuals or SGP, but the other issues we have discussed can still affect these models.

The area in which researchers can feel most comfortable using value-added and growth models is in program evaluation, where programs are dispersed through numerous schools and classrooms or in research that uses teacher or school effects as variables in a regression—in other words, where there is no ranking or evaluation of individual teachers and schools. Studies have shown that a fair amount of signal, even in teacher value added, lending credibility to these estimates (e.g., Kane and Staiger 2008; Chetty et al. 2014). In these cases, the finer points of modeling may not

make as much of a difference as they do in accountability policies for teachers and schools because correlations across methods are generally high.

In summary, value-added and growth models of various types share a common feature in that they always account for prior performance in some way. Several commonly used modeling approaches, such as DOLS, EB, and SGP, can produce measures that are fairly highly correlated, but the optimal choice of a model can depend on their use in various policy and research applications.

Recommended Readings

Guarino, C. M., Reckase, M., & Wooldridge, J. (2015) Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10(1), 117–156.

Guarino et al. provide a step-by-step technical guide to the theory behind value-added modeling and an analysis of simulation data to show the expected performance of various value-added models under different assumptions.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003) *Evaluating value-added models for teacher accountability* (p. 191). RAND.

McCaffrey et al. provide a valuable primer on value-added models in language that is accessible to researchers, scholars, and policy-makers.

Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy*, 4(4), 319–350.

Harris carefully discusses the statistical properties and the policy implications and costs of policies to use value-added measurement in teacher evaluation.

REFERENCES

- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, 25(1), 95–135.
- Amerin-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system (EVAAS). *Educational Researcher*, 37(2), 65–75.

- Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the implementation of evaluation systems. *Educational Researcher*, *44*(2), 77–86.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*(1), 37–65.
- Betebenner, D. W. (2011). A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories. Technical report, The National Center for the Improvement of Educational Assessment.
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton: ETS.
- Briggs, D. C., & Weeks, J. P. (2009). The sensitivity of value-added modeling to the creation of a vertical score scale. *Education Finance and Policy*, *4*(4), 384–414.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *The American Economic Review*, *9*, 2593–2632.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan*, *93*(6), 8–15.
- Dieterle, S., Guarino, C., Reckase, M., & Wooldridge, J. (2015). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value-added. *Journal of Policy Analysis and Management*, *34*(1), 32–58.
- Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2014). The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in Missouri. *Statistics and Public Policy*, *1*(1), 19–27.
- Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, *80*(319), 589–612.
- Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. *Statistics and Public Policy*, *1*(1), 28–39.
- Guarino, C., Maxfield, M., Reckase, M., Thompson, P., & Wooldridge, J. (2015). An evaluation of empirical Bayes' estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics*, *40*, 190–222.
- Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). A comparison of growth percentile and value-added models of teacher performance. *Statistics and Public Policy*, *2*(1), e1034820. <https://doi.org/10.1080/2330443X.2015.1034820>.
- Guarino, C., Reckase, M., Stacy, B., & Wooldridge, J. (2015). Evaluating specification tests in the context of value-added models of teacher performance. *Journal of Research on Educational Effectiveness*, *8*(1), 35–59.

- Guarino, C., Reckase, M., & Wooldridge, J. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10(1), 117–156.
- Hanushek, E., & Rivkin, S. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review: Papers and Proceedings*, 100(2), 267–271.
- Isenberg, E., & Walsh, E. (2015). Accounting for co-teaching: A guide for policy-makers and developers of value-added models. *Journal of Research on Educational Effectiveness*, 8(1), 112–119.
- Johnson, M. T., Lipscomb, S., & Gill, B. (2015). Sensitivity of teacher value-added estimates to student and peer control variables. *Journal of Research on Educational Effectiveness*, 8(1), 60–83.
- Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.
- Koedel, C., & Betts, J. (2009). Value-added to what? How a ceiling in the testing instrument influences value-added estimation, Working Paper 14778, National Bureau of Economic Research.
- Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education*, 6(1), 18–42.
- Martineau, J. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31(1), 35–62.
- McCaffrey, D. F., Sass, T. R., Lockwood, J., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572–606.
- Meyer, R. (2000). Value-added indicators: A powerful tool for evaluating science and mathematics programs and policies. *NISE Brief*, 3(3), 1–8.
- Monk, D. (1994). Subject area preparation of secondary math and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193.
- Reardon, S., & Raudenbush, S. (2009). Key issues in value-added modeling. *Education Finance and Policy*, 4(4), 492–519.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175–214.
- Sass, T., Semykina, A., & Harris, D. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, 38, 9–23.
- Schochet, P. Z., & Chiang, H. S. (2013). What are error rates for classifying teacher and school performance using value-added models? *Journal of Educational and Behavioral Statistics*, 38(2), 142–171.

- Stacy, B., Guarino, C., & Wooldridge, J. (2018). Does the precision and stability of value-added estimates of teacher performance depend on the types of students they serve? *Economics of Education Review*, *64*, 50–74.
- Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, *11*, 340–359.
- Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3–F33.



Social Network Analysis Methods in Educational Policy Research

*Kara S. Finnigan, Daniela E. Luengo-Aravena,
and Kim M. Garrison*

Social network analysis (SNA) helps researchers to examine or uncover the underlying connections among people, behaviors, events, objects, and institutions within and across social systems that might not be obvious otherwise. SNA is a research methodology rooted in *network analysis* and *graph theory*. Some credit Moreno, as early as the 1930s, for focusing on these underlying network connections with his study of runaway girls, as he noticed that social links between girls influenced their behavior (Borgatti and Ofem 2010). Since then, SNA has played a pivotal role in paradigm shifts within and across diverse fields, including social science and epidemiology (Grunspan et al. 2014). Studies have relied on SNA methods to examine happiness and job satisfaction; health behaviors including obesity and drug use; and group behaviors such as community health access or the spread of innovative ideas throughout

K. S. Finnigan (✉) • D. E. Luengo-Aravena • K. M. Garrison
Warner School of Education and Human Development,
University of Rochester, Rochester, NY, USA
e-mail: kfinnigan@warner.rochester.edu; kim.garrison@warner.rochester.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_12

communities from business enterprises to farming. Researchers have embraced SNA in defining the structure of political, economic, and social environments (Wasserman and Faust 1994).¹ Connections explored using SNA range from similarities in affiliations (e.g., membership in the same club), to cognitive or emotional relations (e.g., liking someone), to work-related connections (e.g., giving advice), and flows of resources throughout systems (e.g., information) (Borgatti and Ofem 2010).

In this chapter, we focus on quantitative SNA because of its prominence within the field.² Traditional quantitative education research has focused on how the characteristics of an individual or an organization affect outcomes. Social network theory, on the other hand, is based upon an understanding that individuals in a social system are interdependent, and that these interdependencies shape opportunities and outcomes in ways that require distinct analytic techniques (Borgatti and Ofem 2010). SNA in education research allows researchers to measure and visually map characteristics and elements of social systems to explain or gain insight into a focal relationship of interest.

We begin the chapter by describing SNA with some basic displays of the underlying connections and data sets. Next, we discuss some common theoretical perspectives that frame SNA studies and have already or could inform policy-related work. We review the relevant literature on SNA in education research more broadly and then focus on the innovative application of SNA in educational policy research to strengthen our understanding of policy processes, decision making, and outcomes. Here, we focus on two particular areas—advocacy and implementation—to provide more detailed examples

¹ Note: Social network analysis studies may examine underlying connections through social media, but these are not synonymous.

² Though not the focus here, qualitative methods can be used for SNA and often supplement quantitative SNA methods. Qualitative analysis of intergroup relations can explain the social interpretation of one's position within a network and the meaning that emerges from the social construction of the network (Hollstein 2014). For example, see Cross, Dickmann, Newman-Gonchar, and Fagan's (2009) study of interagency collaboration that involved recorded discussions, reflections, and semi-structured interviews about intergroup relationships or Coburn and Russell's (2008) examination of district math reform policies that used observations and interviews to investigate the qualities of teachers' networks. Qualitative data can also provide information as to the organizational culture and climate that facilitate or hinder underlying relations (see Finnigan et al. 2013 and Finnigan and Daly 2012 for examples).

of the use of SNA in educational policy research. Finally, we offer recommendations to novice and emerging scholars including how to collect and analyze SNA data and useful resources to consult.

SOCIAL NETWORK ANALYSIS

A social network is a set of actors (also referred as nodes, vertices, or points) that can be connected to each other through relationships (also known as edges, links, arcs, lines, or ties). Actors can be a set of persons (e.g., students, principals, policy makers), organizations (e.g., firms, schools, school districts), objects (e.g., policies, documents), or even events (e.g., school meetings, political campaigns). As mentioned above, examples of relationships can include friendships, professional interactions, power structures, or the flow of resources between people or organizations.

The smallest possible social structure in which an actor can be embedded is a dyad, which has two actors that have a relationship or are connected through a “tie.” In turn, the smallest social structure in which a dyad can be embedded is a triad, defined as three actors and the possible relationships among them. Figure 12.1 illustrates the ways that two or three people can be connected through ties.

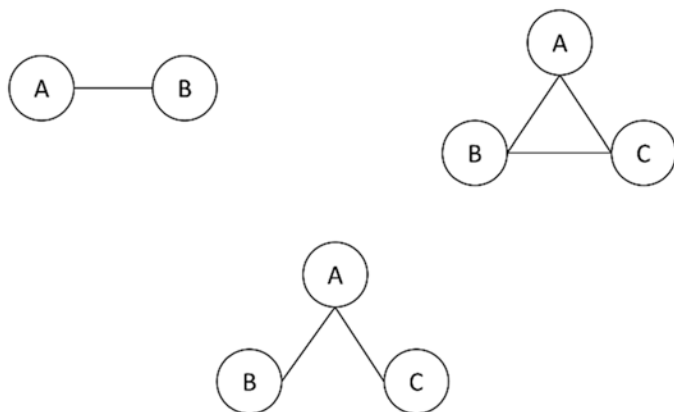


Fig. 12.1 Basic social structures: dyad and triad

Graphs or Sociograms SNA uses graphs and matrices to represent actors and summarize or present the patterns of social relations in an efficient and comprehensive manner. A graph of actors and their relationships is called a sociogram in the SNA literature. In a graph or sociogram, actors are visually represented by *nodes* and their relationships by *lines*. For example, let's suppose we are studying the relationship between four teachers: Lisa, John, Mary, and Paul. We have collected data through a survey that asked each teacher to indicate who they consult for professional advice at least once a month. Lisa selected John and Paul, John selected Paul, Mary also selected Paul, and Paul selected John. This information could be represented in an *undirected graph* such as Fig. 12.2, Panel A. This graph shows whether a relationship between two teachers exists or not, without considering directionality (i.e., the graph does not distinguish the sender or the receiver). We could further specify the information using a *directed graph* by drawing an arrow from the sender to the receiver (with sender in this case meaning who they turn to for advice and receiver meaning who gives the advice) as in Fig. 12.2, Panel B.

In addition, it is possible to add more information to a graph. For example, Fig. 12.2, Panel C shows a *valued graph* that indicates the strength of the relationship between actors. In this case, the thickness of the tie indicates the frequency of the relationship between two teachers (a thicker line indicates a more frequent relationship). Finally, we could add information about the attributes or characteristics of the actors by changing the shape, size, and color of the nodes as in Fig. 12.2, Panel D. In this case, the color of the node indicates the sex of the teacher (gray=female, black=male), the shape shows the subject they teach (circle=math, triangle=language), and the size of the node represents how well connected a particular teacher is to the rest of the teachers within the network (a bigger node size indicates more connections with other teachers).

The Adjacency Matrix Although graphs are useful to represent relational information, if the amount of data is too large (i.e., too many nodes and relationships among them), it may become difficult to interpret using simple visual inspection. Representing the network data in matrices allows computer programs to summarize the information efficiently and find patterns in the data. Usually, social network data is stored in an *adjacency matrix*. In most cases, this is a binary and square matrix with as many rows and columns as actors in the social network. This type of network data is called a one-mode network because it involves person-to-person connections. The matrix is filled with zeros or ones, with a 1 indicating

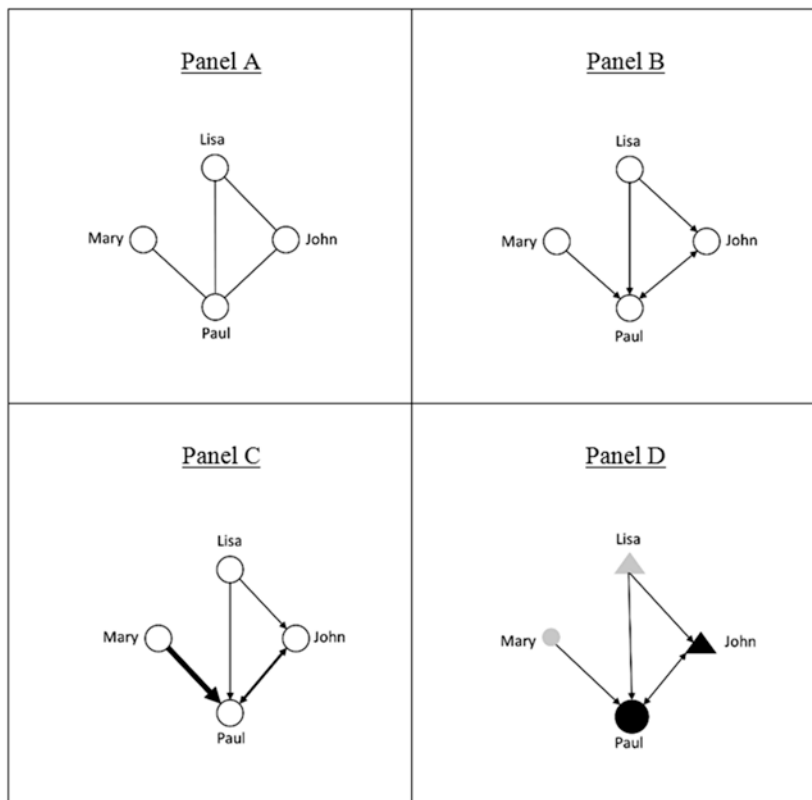


Fig. 12.2 Different ways of representing relationships using sociograms

that a relationship between two pairs of actors exists; while a 0 shows that this relationship is absent (see Table 12.1, Panel A). An adjacency matrix can be *symmetric* or *asymmetric*. In a symmetric adjacency matrix we do not distinguish who nominated whom, we only know whether a relationship between two actors exists. An asymmetric adjacency matrix represents directed relationships and as seen in Table 12.1, Panel B this results in a different set of data—e.g., you can see that Panel B is different from Panel A in the Lisa/John cells with Lisa getting advice from John (1), but John not getting advice from Lisa (0). In Panel B, the shaded gray numbers are those that changed when switching from a symmetric to asymmetric data representation.

Table 12.1 Representing relationships using matrices

Panel A: symmetric adjacency matrix

	Lisa	John	Paul	Mary
Lisa	–	1	1	0
John	1	–	1	0
Paul	1	1	–	1
Mary	0	0	1	–

Panel B: asymmetric adjacency matrix

	Lisa	John	Paul	Mary
Lisa	–	1	1	0
John	0	–	1	0
Paul	0	1	–	0
Mary	0	0	1	–

When relationships are studied with people across events or affiliations, SNA moves from the traditional *one-mode network* structure to a *two-mode network structure* (2MN). One often used data set, called *Deep South*, demonstrates how events can be used to understand how actors—in this case, 18 southern women—were connected through 14 social gatherings (see Davis et al. 1941). In this case, or others that look at affiliations in terms of attendance at events (e.g., memberships on boards, etc.), inferences are made about underlying patterns of ties or groupings based upon these affiliations. If two of the women in the Deep South data set attended the same social gathering, we now assume a relationship between these two women. Importantly, affiliations are considered broadly and thus two-mode networks may involve any connection between two different groups, such as researchers and journals (in terms of where they have published), donors and initiatives, voters and candidates, readers and magazines, etc. For two-mode networks, matrices are usually rectangular in shape because the number of people and events are no longer required to be the same.

Most of the discussion we include in this chapter focuses on the structure of whole networks, whether across people or agencies, because these seem particularly well-suited to educational policy research; but it is important to mention a different unit of analysis—the ego level or personal level of a network—which could also be considered. Ego network studies can provide rich detailed information about a policy in terms of knowledge around it, influence over decisions, or implementation of a policy on the ground, and is local to the person(s) versus examining more global patterns. For example, in considering how a particular policy became supported by a school board member one might consider the structure and quality of that board member’s ego network—meaning all of the connections that board member has to other board members,

higher education faculty, policy makers, or teachers. Additional details about differences in collecting and analyzing whole network versus ego network data can be found in Borgatti, Everett, and Johnson (2013).

SOCIAL NETWORK ANALYSIS IN EDUCATION RESEARCH

While SNA is not entirely new to education research it is still a burgeoning area of work.³ Within education, scholars have been concerned with the importance of connections, ties, and attributes of formal and informal networks, which can be important to building collaborative communities, student interest and support groups, and staff agency and efficacy (see Kezar 2014; Lubbers 2003; Lubbers and Snijders 2007; Siciliano 2016). Most of the scholarly work employing SNA in education focuses on three main areas: peer networks, teacher networks, and leader or administrator networks (predominantly at the K-12 level). In addition, some higher education research has employed SNA methods in studies of change in higher education, research collaboration, student activism, and campus social and cultural capital in peer and staff networks (see Kezar 2014). We briefly discuss some examples of this work to provide the reader with an understanding of the variety of uses of SNA at the K-12 and higher education levels before turning our attention to applying SNA in education policy research.

General Applications: Peers, Teachers, and Leader Networks Peer network studies have focused on the role of peers in students' educational outcomes (e.g., academic effort, dropout rate), health outcomes (e.g., weight control, alcohol and drug consumption), or socio-emotional outcomes (e.g., social integration, homophobic behavior). Friendship networks are thought to be a rich source of resources that students can accrue or exchange in order to shape their future opportunities and outcomes. One example of this type of research is Frank et al.'s (2008) examination of how high school students' math course-taking was influenced by friendship groups, finding that girls take into account the decisions of their friends in course selection. Similarly, Grunspan et al. (2014) used SNA to investigate whether and how learning outcomes were related to classroom networks in an undergraduate biology course.

³For more details of SNA in education, see Carolan (2014) and Daly (2010).

Networks among teachers have also been identified as a critical area of research. Work in this area has identified patterns relating to teachers' professional interactions, including the extent to which they exhibit collegial relationships, and the ways in which these impact social capital acquisition and teacher learning (e.g., Penuel et al. 2009). In addition, teacher networks have been found to be a leverage point for the diffusion of instructional expertise through professional development (e.g., Sun et al. 2013), impacting feelings of teacher efficacy (e.g., Siciliano 2016), influencing the reform-related attitudes of teachers (Cole and Weinbaum 2010), and impacting student achievement (Pils and Leana 2009; Siciliano 2015).

Recent scholarly work has focused on the leadership networks (including school and central office administrators) within the educational system as a crucial factor for educational change and improvement. For example, Daly and Finnigan's longitudinal study of social networks within and across low-performing schools and districts found that sparse networks across leaders limits access to research evidence as leaders undergo new strategies in response to policy sanctions (Finnigan et al. 2013) and that weak and uni-directional connections between principals and central office staff in low-performing districts are particularly problematic to district-wide reform efforts (Finnigan and Daly 2012).⁴ These authors also found high levels of leadership churn in low-performing districts, with the most sought after leaders for advice leaving the district (Finnigan et al. 2016), and called attention to the underlying politics that inhibited improvement given the network structure of leaders (Daly et al. 2014).

While there has been some attention to social networks in education research at the higher education level, this remains an underexplored area (Kezar 2014). Higher education researchers have focused on faculty networks and productivity and peer networks and student outcomes, but there is not yet research that studies university systems as a whole or higher education network actors as discussed by Biancani and McFarland (2013). A few studies that focus on college student peer networks include Thomas' (2000) study which examined ties among college freshman to predict college persistence; Rios-Aguilar and Deil-Amen's (2012) study of Latina/o college students' networks which found that ties that helped students to enroll were not as useful in supporting them during college and with post-college planning; and Gonzalez Canché and Rios-Aguilar's (2015) study of the influence of community college peers on credit attainment.

⁴For more results from this longitudinal study, see (Daly and Finnigan 2011, 2012; Finnigan and Daly 2014).

Applications to the Study of Education Policy Above, we provided examples of the limited attention to SNA in education research, to date. Perhaps more important has been the dearth of research using SNA methods at the educational policy level. Yet every stage of the policy cycle (problem definition and agenda setting, design, implementation, and evaluation) involves a social process. According to Knoke (2011), “policy network analysis seeks to identify the important actors – governmental and non-governmental organizations, interest groups, and persons – involved in policymaking institutions, to describe and explain the structure of their interactions during policymaking processes, and to explain and predict collective policy decisions and outcomes” (p. 210). Thus, SNA provides researchers a valuable tool especially well-suited to study policy (Hermans and Thissen 2009; Penuel et al. 2006; Song and Miskel 2005). Despite its potential, SNA has not been as commonly used in educational policy as it has been in other realms such as public health (Carolan 2013).

In this section, we illustrate the potential of SNA for studying policy issues in K-12 and higher education by offering two recent empirical examples. Each of these focuses on a specific stage of the policy cycle though it is important to note that there are many more ways to study policy using SNA.

Policy Advocacy SNA can be used to study policy issues in the early stages of the policy cycle. It is particularly well suited to studies that focus on underlying politics, including policy influences, agenda setting, and policy advocacy. For example, Au and Ferrare (2014) used SNA to examine how the network of relationships among policy actors influenced the passage of a charter school initiative in Washington. Although, voters had opposed charter school legislation in three previous referendums in 1996, 2000, and 2004, they approved I-1240 in 2012. Critical to understanding the passage of any policy is to recognize the influence that advocacy groups exercise over the public vision and political discourse. In this case, Au and Ferrare (2014) examined the influence of policy advocates within the context of the “Yes On 1240 WA Coalition for Public Charter Schools” campaign by using a social capital perspective where policy advocates transferred material (e.g., donations or volunteers) or symbolic resources (e.g., prestige) through their social connections to shift the public vision and political discourse.

Using various data sources—including tax returns, institutional reports, and public disclosures—Au and Ferrare (2014) uncovered connections among organizations and individuals and the Yes Campaign, and generated a binary adjacency matrix of the directed relationships among the policy actors. From this matrix, a directed graph was constructed that traces the transference, or flow, of resources among policy actors and organizations supporting the “Yes On 1240” campaign.

Using social network analysis methods, Au and Ferrare (2014) identified key influencers over this policy process, finding that the Bill & Melinda Gates Foundation was the most central transmitter of resources. Almost all the transferred resources were material instead of symbolic; and philanthropic foundations acted as channels directing resources toward the policy actors to influence the charter school policy adoption. SNA was an innovative way to trace the influence of private individuals and different types of organizations including philanthropic foundations and advocacy groups over the adoption of the charter school policy.

Policy Implementation Beyond policy advocacy, SNA can also be useful in studying policy implementation and its effects on education systems. To illustrate this application, we turn to a recent study by Hodge, Salloum, and Benko (2016). These researchers examined Common Core State Standards’ (CCSS) implementation by focusing on secondary ELA resources (e.g., professional development or curricular resources for teachers) that were sponsored and shared by 51 state educational agencies (SEAs) and other intermediary organizations (i.e., non-system actors such as research institutes, nonprofits, and policy or advocacy organizations, to name a few). The authors investigated the types of resources SEAs were recommending and which SEAs supported these resources—as well as how resources tied these groups together—to consider what CCSS messages were being spread throughout the system (and would ultimately reach ELA teachers). SNA allowed the researchers to examine the *structure* of the two-mode resource sharing network to understand CCSS implementation.

Using public information available on SEA homepages, the authors catalogued 2001 resources, which included 2644 ties or *edges* across 313 agencies (51 SEAs and 227 organizations) or *nodes*. The researchers examined the centrality of SEAs as well as their connectedness to other organizations (see Fig. 12.3). The results illustrate a *core-periphery network*, meaning certain agencies that are highly connected remain in the center,

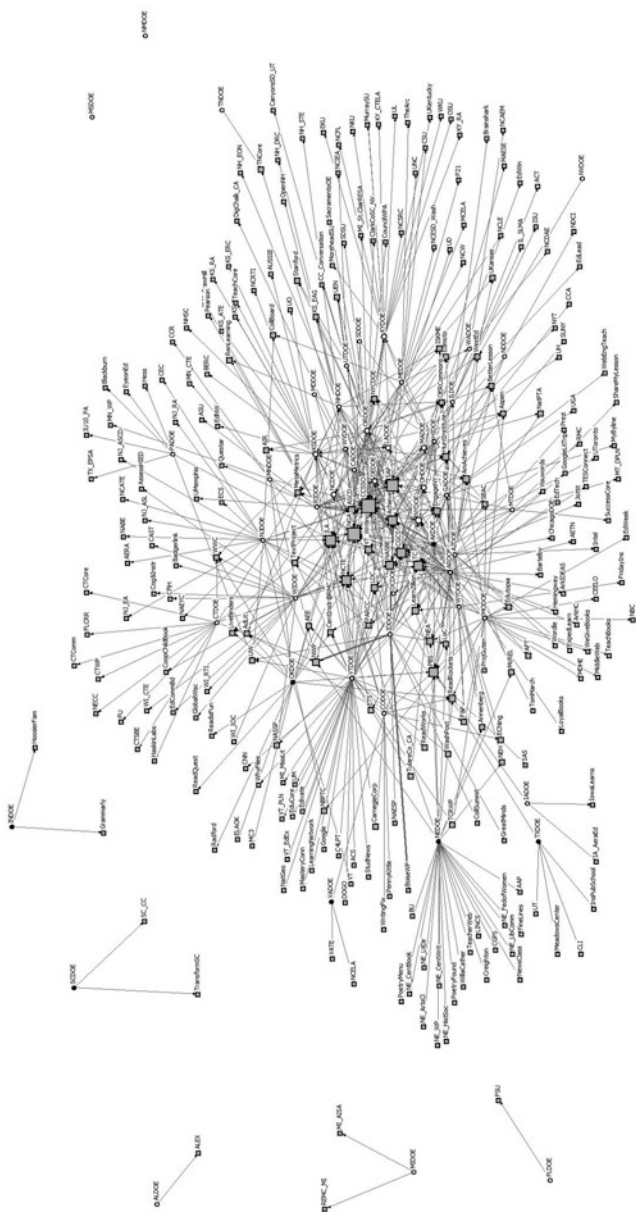


Fig. 12.3 Social network of state education agencies and intermediary agencies' connections around common core ELA resources. (Note: SEAs are circles; white are CCSS SEAs, and black are non-CCSS SEAs; grey squares are intermediary organizations, tie thickness represents number of resources used from an agency and node size represents a measure of the agencies' centrality in the network. Source: Hodge et al. (2016, p. 7))

or “core” (e.g., SEAs such as Massachusetts, North Carolina, Ohio, and New York as well as education policy membership organizations like CCSSO, general membership organizations like ACSD (formerly the Association for Supervision and Curriculum Development) and membership organizations focused on literacy like the National Council of Teachers of English known as NCTE). For-profit companies were also central in the network. Those with fewer connections were located toward the periphery (e.g., Iowa and Iowa Learns).

Social network analysis methods provided the tools for Hodge and her colleagues to uncover: (1) the more influential actors over Common Core implementation because of their provision of CCSS resources that were shared by a number of states; and (2) the multiple groups that may be sending conflicting messages about CCSS instruction, given their centrality in the network. They also found that CCSS states were more likely to connect with each other than non-CCSS states, but found uneven connectedness and isolation among some states, suggesting external resources may not be making their way into some states. Given the state department of education’s role in building capacity among teachers, *what* they provide teachers around ELA instruction is extremely important to CCSS implementation. The SNA suggested that while many states provided conceptual resources, more practical ones appeared desirable as many groups were seeking these out from particular agencies. SNA in this case provided important empirical data that can inform policymakers as to the assumptions and challenges relating to state CCSS implementation and the diffusion of resources.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

At this point we have described what SNA is and how it can be used to study educational policy. Here we provide some more specific recommendations to scholars as they consider whether SNA is appropriate for their work.

Step 1: Consider Whether Relational Questions Drive Your Analysis It is important to consider whether the relationships between individuals, organizations, or events are important to your area of study. Remember that social network studies consider questions like who is affiliated with a particular group or who seeks advice from whom? In the policy realm, questions that might drive the research are questions relating to which

advocacy groups may have influenced policy discourse, how connections to certain teachers or district events may have impacted implementation, or whether certain policy outcomes—e.g., increased collaboration among higher education units—were achieved. In this type of research, the connections or ties among actors are central to the analysis.

Let's consider a local policy that involves a new student assignment policy to reduce segregation in a particular district. In this case, questions relating to who (e.g., the media, groups representing communities of color, or parents) influenced the design of the policy would be particularly important and social network analysis could help to uncover the most influential actors.

Step 2: Consider What Theoretical Perspectives Drive Your Analysis Though SNA is not directly linked to any particular theoretical perspectives, we highlight a few common lenses here, including social capital, cognitive social structure theories, diffusion theories, and the advocacy coalition framework, to illustrate lenses that may be useful in policy research. Choice of perspective or lens will be closely linked to the questions asked and unit of analysis of the work.

Social Capital Network theories are central to the concept of social capital, as individuals are embedded within relationships, and these relationships are embedded in larger subgroups that eventually form a social network. This theory suggests that personal connections and interpersonal interactions are an investment just like other types of capital (e.g., human or cultural capital) (see Scott's (2017) discussion of how Bourdieu, Coleman, and Putnam contributed to these areas). Social capital is operationalized as the resources embedded in social systems and used by actors for action (Lin 2001), and these resources can vary from communication to information exchange, trust, and knowledge sharing (Scott 2017; Wasserman and Faust 1994). Relationships create a structure that determines opportunities for social capital transactions or access to these resources (Burt 1992; Coleman 1988, 1990; Granovetter 1973, 1983; Lin 2001; Putnam 1993, 2000).

Cognitive Social Structures Studies using cognitive social structure (CSS) theories (Krackhardt 1987) aim to integrate the role of cognition and meaning-making, so in this case it might be around making sense of or

interpreting policy. The most common frameworks shaping CSS studies emerge from social cognition and structuralist theories that position context and social relationships as central components to meaning-making. Based on these theories, social structures influence the individual's position and exposure within the network context, social interaction leads to expectations for future interactions, and individuals' social positions then impact how they see actors in the network (Casciaro 1998). CSS can uncover whether individuals' interpretations align with intended outcomes or predict future actions (see also Brands 2013 and Pierce et al. 2014).

Diffusion Theories Diffusion theory has roots in anthropology, sociology, epidemiology, geography, and marketing, among other areas, and describes the mechanism by which new ideas, opinions, attitudes, and behaviors spread throughout a community (Bailey 1975; Rogers 2003; Ryan and Gross 1943; Valente 1993, 1995; Valente and Rogers 1995). Initially described by Ryan and Gross (1943), the basic premise is that new ideas and practices spread through interpersonal contacts and communication (see Beal and Bohlen 1955; Hagerstrand 1968; Katz et al. 1963; Rogers 1995; Valente 1995; Valente and Rogers 1995). Diffusion modeling assumes a classic S-shaped curve whereby initial growth in adopting something occurs gradually at first, then accelerates, then decelerates (Rogers 2003). Because diffusion often occurs through personal networks, and these networks are shaped by many factors, including geography, ethnicity, age, and socioeconomic status (SES), there may be different diffusion trajectories for different subgroups (Valente and Fosados 2006). Knowledge diffusion is largely influenced by interactions, which serve as conduits (Moody 2004).

Advocacy Coalition Framework (ACF) This framework suggests that actors in a particular policy subsystem (defined as a policy issue/area, usually bounded geographically, that encompasses different policy stakeholders such as government, interest groups, research organizations, and media) structure themselves into coalitions of competing policy beliefs to shift policy toward their coalition's interests (Sabatier 1988; Sabatier and Jenkins-Smith 1993). Advocacy coalitions are stable social groups over time that coordinate and share beliefs and resources within but not across the boundaries of the coalition (Sabatier 1988). According to this framework, policy change can occur through administrative organizations,

which usually maintain a more moderate position regarding an issue, and thus, can act as brokers or mediators among coalitions. In addition, new scientific information can be used by coalitions to support their political views and produce policy learning. Finally, exogenous shocks or new information can also be the origin of policy change (Sabatier 1988; Sabatier and Jenkins-Smith 1993).

While these are by no means meant to be exhaustive, these four theoretical frameworks offer a few different conceptual lenses and show the variety of perspectives that might align with and inform SNA studies. Many other sociological, psychological, or political lenses could inform your work.

Step 3: Collect Network Data Collecting data for social network studies depends upon the type of analysis—one mode and two mode—and available data. Many studies involve surveys administered to individuals that involve questions relating to the existence or frequency of ties. As an example, if a researcher was concerned with the structure of organizational friendships and support, they may ask individuals to establish “Who would you consider a close friend?” In policy research, data may be collected around who someone asked advice from around a particular policy or who was at certain events when a policy was being discussed or designed. Using our example above again, we might consider two different types of studies. A one-mode study might ask all key stakeholders who they asked for advice about the policy—which would result in a matrix much like the one depicted in Table 12.1, panel B. Alternatively, a two-mode study might consider all of the individuals who were on the task force to develop the policy and what affiliations they had in the community to uncover the strength of the influence of various groups through these affiliation ties. In this case, the two-mode matrix would have individuals x community groups, as opposed to the one-mode matrix with actor x actor (for more on two-mode SNA, see Borgatti 2012; Borgatti and Everett 1997). As mentioned above, important to data collection is consideration of whether the study will examine a complete network (e.g., all of school board members and their relationship to each other), or ego networks of individuals (e.g., the network of advice for individual school board members which would include anyone they turn to for advice whether they are on the school board or not).

The use of SNA methods comes with certain ethical considerations relating to collecting this type of data (Borgatti and Molina 2003; Kadushin 2012). First, unlike traditional survey techniques and analysis, respondents' anonymity may be difficult to protect; this is especially true of intra-organizational and subunit analyses. However, inter-organizational or more nested methods move beyond the issues of anonymity. Practices like using a third party to process attribute data and other sensitive data such as value-added data that link teachers and student achievement to a unique identifier before releasing data to researchers for analysis can also help to maintain confidentiality.

Step 4: Prepare Data for Analysis In uncovering and understanding the actors and ties of social networks, researchers pay particular attention to metrics of *density*, *reciprocity*, *centrality*, and *homophily*. Density provides information about how well connected or sparse the relationships are in a social network, and it is defined as the proportion of actual ties to all possible ties within a network. Reciprocity is the proportion of mutual connections across the network and measures the strength of a relationship. Centrality aims to quantify the relevance or influence of a particular actor within a social network. Lastly, homophily measures the desire for individuals to establish relationships with others that share similar characteristics or beliefs to themselves.

Beyond descriptive analyses and visual presentation of sociograms, some social network studies in education involve regression analysis or estimation procedures (e.g., Daly and Finnigan 2012; Sun et al. 2013; Moolenaar et al. 2014), and multi-level modeling (e.g., Siciliano 2016; Spillane and Kim 2012).⁵ In these studies, the centrality of individuals for example, might be used to predict outcomes compared with more peripheral actors, or fidelity of implementation around a particular policy may predict higher levels of network centrality. In the case of our example around student assignment, more decentralized leadership networks in a community might predict more successful policy implementation (because of greater buy-in across diverse groups).

⁵ Additional relevant examples outside of education that might be useful include Yu, Hao, Dong, and Khalifa (2013) which investigated knowledge sharing behaviors of individuals and within teams using a multi-level nested model.

More advanced statistical methods can also be used to look at policy research which can overcome limitations of interdependence within and across ties. Using exponential random graph models (ERGMs)—also sometimes referred to as p^* models—SNA software like Simulation Investigation for Empirical Network Analysis (SIENA) (Ripley and Snijders 2010) and its R-package version RSiena (Ripley et al. 2017) can determine whether the formation of networks (e.g., voters and policies; organizations sponsoring projects; etc.) can help to inform policy formation, implementation, and outcomes by comparing actual networks with simulated stochastic models of networks with the same characteristics to establish whether network structure is based on chance or not. In this manner, Berardo (2014) was able to determine the structure of organizations as they were linked to projects and the role that governmental actors might play in *brokering* and *bridging* inter-organizational collaboration on projects through actor by organization network analysis. Though not an educational policy-specific study, it sheds light on the complexity of SNA methods that can be applied to educational policy research agendas by considering agency and structure of networks within specific educational contexts.

CONCLUSION

Social network analysis (SNA) is a unique methodology, allowing researchers to examine and uncover the underlying connections among people, behaviors, events, objects, and institutions within and across social systems. Its focus on connections and relationships makes SNA ideal for studying policy—which involves a social process at every stage. As emerging and experienced researchers consider the theories and methods that best explain, uncover, and advance understandings relating to policy advocacy, policy design, policy implementation, and policy outcomes, it is worthwhile to consider the ways that SNA might expand our knowledge base in these critical areas. We hope this chapter has contributed to the larger conversation around how policy research can be advanced by innovative methodological approaches to meet the complex needs of the field and to produce rigorous results that can improve policy and practice and ultimately the outcomes and opportunities for youth.

Recommended Readings

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. London: Sage Publications.

This practical book walks readers through all aspects of the research process from designing a study to interpreting the results. The book includes chapters on data collection and management, visualization, and analytical approaches including analyses particular to SNA such as related to subgroups, centrality, ego networks, etc. Readers are also introduced to the software developed by the first two authors for analysis of social network data, UCINET, and Netdraw.

Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. New York: Oxford University Press.

This book covers fundamental concepts in SNA, presenting core themes, constructs, and applications. It is especially useful for researchers who are new to the social network field and particularly interested in the psychological and sociological underpinnings of SNA. The book calls attention to ethical considerations in collecting and using social network data.

Scott, J., & Carrington, P. J. (Eds.) (2011). *The Sage handbook of social network analysis*. London: Sage Publications.

This is a comprehensive text that introduces readers to SNA by systematically reviewing the concepts, theories, methods, principal topics, and discussions within the field. While it can provide introductory material to a newcomer it also will be useful to more seasoned researchers who are interested in developing stronger grounding in the underlying theories, mathematical models, and variety of applications of SNA.

REFERENCES

- Au, W., & Ferrare, J. J. (2014). Sponsors of policy: A network analysis of wealthy elites, their affiliated philanthropies, and charter school reform in Washington State. *Teachers College Record*, 116(8), 1–24.
- Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications* (2nd ed.). London: Griffin.
- Beal, G. M., & Bohlen, J. M. (1955). *How farm people accept new ideas, Cooperative Extension Service Report* (Vol. 15). Ames: U.S. Department of Agriculture.
- Berardo, R. (2014). Bridging and bonding capital in two-mode collaboration networks. *The Policy Studies Journal*, 42(2), 197–225.
- Biancani, S., & McFarland, D. A. (2013). Social networks research in higher education. In M. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 151–215). Dordrecht: Springer.
- Borgatti, S. P. (2012). Social network analysis, two-mode concepts. In R. A. Meyers (Ed.), *Computational complexity* (pp. 2912–2924). New York: Springer.
- Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, 19(3), 243–269.

- Borgatti, S. P., & Molina, J. L. (2003). Ethical and strategic issues in organizational social network analysis. *The Journal of Applied Behavioral Science*, 39(3), 337–349.
- Borgatti, S. P., & Ofem, B. (2010). Overview: Social network theory and analysis. In A. J. Daly (Ed.), *Social network theory and educational change* (pp. 17–30). Cambridge, MA: Harvard Education Press.
- Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. Thousand Oaks: Sage Publications.
- Brands, R. A. (2013). Cognitive social structures in social network research: A review. *Journal of Organizational Behavior*, 34, S82–S103.
- Burt, R. (1992). *Structural holes: The social structure of competition*. Cambridge: Harvard University Press.
- Carolan, B. V. (2013). *Social network analysis and education: Theory, methods and applications*. Thousand Oaks: Sage Publications.
- Carolan, B. V. (2014). *Social network analysis and education: Theory, methods, & applications*. Thousand Oaks, CA: SAGE Publications.
- Casciaro, T. (1998). Seeing things clearly: Social structure, personality, and accuracy in social network position. *Social Networks*, 20, 331–351.
- Coburn, C. E., & Russell, J. L. (2008). District policy and teacher's social networks. *Education Evaluation and Policy Analysis*, 30(3), 203–235.
- Cole, R. P., and Weinbaum, E. H. (2010). Changes in attitude: Peer influence in high school reform. In A. J. Daly (Ed.), *Social network theory and educational change* (pp. 77–96). Cambridge, MA: Harvard University Press.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120.
- Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.
- Cross, J. E., Dickmann, E., Newman-Gonchar, R., & Fagan, J. M. (2009). Using mixed-methods design and network analysis to measure development of inter-agency collaboration. *American Journal of Evaluation*, 30(3), 310–329.
- Daly, A. J. (2010). *Social network theory and educational change*. Cambridge, MA: Harvard Education Press.
- Daly, A. J., & Finnigan, K. S. (2011). The ebb and flow of social network ties between district leaders under high-stakes accountability. *American Educational Research Journal*, 48(1), 39–79.
- Daly, A., & Finnigan, K. (2012). Exploring the space between: Social networks, trust, and urban school district leaders. *Journal of School Leadership*, 22(3), 493–530.
- Daly, A. J., Finnigan, K. S., Jordan, S., Moolenaar, N. M., & Che, J. (2014). Misalignment and perverse incentives: Examining the role of district leaders as brokers in the use of research evidence. *Educational Policy*, 28(2), 145–174.
- Davis, A., Gardner, B. B., & Gardner, M. R. (1941). *Deep south: A social anthropological study of caste and class*. Chicago: University of Chicago Press.

- Finnigan, K. S., & Daly, A. J. (2012). Mind the gap: Organizational learning and improvement in an underperforming urban system. *American Journal of Education*, 119(1), 41–71.
- Finnigan, K. S., & Daly, A. J. (Eds.) (2014). *Research Evidence in Education: From the Schoolhouse Door to Capitol Hill*. Switzerland: Springer.
- Finnigan, K. S., Daly, A. J., & Che, J. (2013). Systemwide reform in districts under pressure: The role of social networks in defining, acquiring, and diffusing research evidence. *Journal of Educational Administration*, 51(4), 476–497.
- Finnigan, K. S., Daly, A. J., & Liou, Y. (2016). How leadership churn undermines learning and improvement in low-performing school districts. In A. J. Daly & K. S. Finnigan (Eds.), *Thinking and acting systemically: Improving school districts under pressure* (pp. 183–208). Washington, DC: American Educational Research Association.
- Frank, K. A., Muller, C., Schiller, K. S., Riegle-Crumb, C., Mueller, A. S., Crosnoe, R., & Pearson, J. (2008). The social dynamics of mathematics coursetaking in high school. *American Journal of Sociology*, 113(6), 1645–1696.
- González Canché, M. S., & Ríos-Aguilar, C. (2015). Critical social network analysis in community colleges: Peer effects and credit attainment. *New Directions for Institutional Research*, 2014(163), 75–91.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Granovetter, M. S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1, 201–233.
- Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE Life Sciences Education*, 13(2), 167–178.
- Hagerstrand, T. (1968). *Innovation diffusion as a spatial process*. Chicago: University of Chicago Press.
- Hermans, L. M., & Thissen, W. A. (2009). Actor analysis methods and their use for public policy analysts. *European Journal of Operational Research*, 196(2), 808–818.
- Hodge, E. M., Salloum, S. J., & Benko, S. L. (2016). (Un)commonly connected: A social network analysis of state standards resources for English/Language Arts. *AERA Open*, 2(4), 1–19.
- Hollstein, B. (2014). Qualitative approaches. In J. Scott & P. J. Carrington (Eds.), *The SAGE handbook of social network analysis* (pp. 404–416). London: Sage Publications.
- Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. New York: Oxford University Press.
- Katz, E., Levin, M. L., & Hamilton, H. (1963). Traditions of research on the diffusion of innovation. *American Sociological Review*, 28, 237–252.
- Kezar, A. (2014). Higher education change and social networks: A review of research. *The Journal of Higher Education*, 85(1), 91–125.

- Knoke, D. (2011). Policy networks. In J. Scott & J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 210–222). London: Sage.
- Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9(2), 109–134.
- Lin, N. (2001). *Social capital: A theory of structure and action*. London: Cambridge University Press.
- Lubbers, M. J. (2003). Group composition and network structure in school classes: A multilevel application of the p* model. *Social Networks*, 25(4), 309–332.
- Lubbers, M. J., & Snijders, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4), 489–507.
- Moody, J. (2004). The structure of a social science collaboration network. *American Sociological Review*, 69, 213–238.
- Moolenaar, N. M., Daly, A. J., Cornelissen, F., Liou, Y., Caillier, S., Riordan, R., Wilson, K., & Cohen, N. A. (2014). Linked to innovation: Shaping an innovative climate through network intentionality and educators' social network position. *Journal of Educational Change*, 15(2), 99–123.
- Penuel, W. R., Sussex, W., Korbak, C., & Hoadley, C. (2006). Investigating the potential of using social network analysis in educational evaluation. *American Journal of Evaluation*, 27(4), 437–451.
- Penuel, W. R., Riel, M., Krause, A., & Frank, K. A. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record*, 111(1), 124–163.
- Pierce, J. J., Siddiki, S., Jones, M. D., Schumacher, K., Pattison, A., & Peterson, H. (2014). Social construction and policy design: A review of past applications. *Policy Studies Journal*, 42(1), 1–29.
- Pils, F. K., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *The Academy of Management Journal*, 52(6), 1101–1124.
- Putnam, R. D. (1993). *Making democracy work*. Princeton: Princeton University Press.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.
- Rios-Aguilar, C., & Deil-Amen, R. (2012). Beyond getting in and fitting in: An examination of social network and professionally relevant social capital among Latina/o university students. *Journal of Hispanic Higher Education*, 11(2), 179–196.
- Ripley, R. M., & Snijders, T. A. B. (2010). *Manual for SIENA version 4.0*. Oxford: University of Oxford.
- Ripley, R. M., Snijders, T. A. B., Boda, Z., Voros, A., & Preciado, P. (2017, September 9). *Manual for RSiena*. Oxford: University of Oxford.
- Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Simon & Schuster.

- Ryan, B., & Gross, N. C. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8(1), 15.
- Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences*, 21(2), 129–168.
- Sabatier, P. A., & Jenkins-Smith, H. C. (1993). *Policy change and learning: An advocacy coalition approach*. Boulder: Westview Press.
- Scott, J. (2017). *Social network analysis* (4th ed.). Thousand Oaks: Sage Publications.
- Siciliano, M. D. (2015). Professional networks and street-level performance: Advice networks influence student performance. *American Review of Public Administration*, 47(1), 79–101.
- Siciliano, M. D. (2016). It's the quality not the quantity of ties that matters: Social networks and self-efficacy beliefs. *American Educational Research Journal*, 53(2), 227–262.
- Song, M., & Miskel, C. G. (2005). Who are the influentials? A cross-state social network analysis of the reading policy domain. *Education Administration Quarterly*, 41(1), 7–48.
- Spillane, J. P. and Kim, C. M. (2012) An exploratory analysis of formal school leaders' positioning in instructional advice and information networks in elementary schools. *American Journal of Education*, 119(1), 73–102.
- Sun, M., Penuel, W. R., Frank, K. A., Gallagher, H. A., & Youngs, P. (2013). Shaping Professional Development to Promote the Diffusion of Instructional Expertise Among Teachers. *Educational Evaluation and Policy Analysis*, 35(3), 344–369.
- Thomas, S. L. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The Journal of Higher Education*, 71(5), 591–615.
- Valente, T. W. (1993). Diffusion of innovations and policy decision-making. *Journal of Communication*, 43(1), 30–45.
- Valente, T. W. (1995). *Network models of the diffusion of innovations*. Cresskill: Hampton Press.
- Valente, T. W., & Fosados, R. (2006). Diffusion of innovations and network segmentation: The part played by people in promoting health. *Sexually Transmitted Diseases*, 33(7), S23–S31.
- Valente, T. W., & Rogers, E. M. (1995). The origins and development of the diffusion of innovations paradigm as an example of scientific growth. *Science Communication*, 16(3), 242–273.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Yu, Y., Hao, J., Dong, X., & Khalifa, M. (2013). A multilevel model for effects of social capital and knowledge sharing in knowledge-intensive work teams. *International Journal of Information Management*, 33(5), 780–790.



Essential Steps to Assessing a School System's Fiscal Health

*Joshua R. Zender, Kenneth A. Smith,
and John R. Kurpierz*

School district financial management focuses on resource generation and expenditures related to the accomplishment of specific educational goals. The financial management field thus explores sound school district budgeting techniques, effective management of public resources, and thorough monitoring of budgetary progress through accounting, auditing, and reporting activities. Although related to the study of school finance and fiscal policy, such research does not generally adhere to the tenets of these fields. Rather, as a management discipline, the goal of financial research is to benchmark subjects' performance against their peers, identify potential

J. R. Zender (✉)

School of Business, Humboldt State University, Arcata, CA, USA

e-mail: joshua.zender@humboldt.edu

K. A. Smith

College of Business, Central Washington University, Ellensburg, WA, USA

e-mail: kenneth.smith@cwu.edu

J. R. Kurpierz

Schulich School of Business, York University, Toronto, ON, Canada

© The Author(s) 2018

C. R. Lochmiller (ed.), *Complementary Research Methods for Educational Leadership and Policy Studies*,

https://doi.org/10.1007/978-3-319-93539-3_13

gaps in performance, and propose corrective action. In this regard, financial management research resembles program evaluation studies, and other forms of improvement research, where the ultimate goal of the research activity is to inform localized educational practices and specific stakeholder needs.

Evaluating the financial activities of public education institutions, including school districts, is not easy. Unlike examining the operations of a business franchise, such as McDonald's, where all locations have a singular objective of generating a profit, school district goals and objectives can vary by region, school level, and the focus of each educational program. Given this reality, a "one-size fits all" research approach often does not apply when scholars seek to examine financial management practices within schools. Adding to the complexity is the fact that a scholar must be familiar with the local political landscape, technical laws and regulations that govern financial operations, and the budgeting process used by the school district. Of the 14,000 school districts across the United States, over 91% are independent of general-purpose governments, such as city or county municipalities (U.S. Census Bureau 2012). Despite being common public organizations, there are many underexplored areas in this field. Thus, there are many opportunities for scholars to make a contribution using the analytic techniques found in financial management research. The purpose of this chapter is to introduce scholars to some of the fundamental resources and methods that they might use to undertake a study of this field. Within this chapter, we explore common techniques for assessing the accountability and financial performance of individual school systems by addressing:

- *Who are the users of school financial condition studies?*
- *What key research contributions have occurred within the field?*
- *How can one execute a research study using financial data?*

The first section of the text focuses on the different stakeholders in the educational policy landscape and their respective financial reporting needs. In the next section, we introduce the reader to seminal works and resources in the school financial management field. The heart of the chapter addresses research strategies for conducting fiscal research. We reinforce the introduction to these research techniques using practical examples drawn from school district data. We conclude the chapter by sharing practical insights drawn from our own experiences.

USERS OF SCHOOL FINANCE STUDIES

Educational institutions operate within a multilayered, yet locally empowered policy landscape. Under the Tenth Amendment of the U.S. Constitution, the responsibility of coordinating and operating public schools is reserved to the states and their local conduits. When designing a research plan, emerging and seasoned scholars alike should have an appreciation of the role of stakeholders in this context, as well as their financial information needs.

Federal Government Although the federal government plays a limited decision-making role on local district-level policy decisions, federal policy mandates can have a profound effect on school system budgets. Odden and Picus (2013) note that the federal government can influence the school district by “mandating changes in the way local services are provided, or it can use intergovernmental grants” (p. 154). Examples of national laws and programs affecting local resource allocation decisions include the Elementary and Secondary Education Act, Head Start, and Special Education programs. In other cases, the federal government may offer competitive grants to encourage systems to experiment with pedagogical models or school choice programs. Often, these laws impose performance standards which require sufficient program funding levels. School districts are likely to have to collect financial and performance data to demonstrate compliance with standards or achievement of equity to the U.S. Department of Education.

The U.S. justice system can also affect school systems’ budgets by dictating mandatory spending levels. Court decisions have changed requirements around sports programs, identified inequalities due to variations in taxable wealth, and even compelled districts to change their busing routes as a remedial technique for achieving desegregation (e.g., *Brown v. Board of Education*, 1954; *Rodriguez v. San Antonio Independent School District*, 1972; *Serrano v. Priest*, 1973; *McDaniel v. Thomas*, 1981; *Hertzell v. Connell*, 1984).

State Government In contrast to the federal government, states perform a dominant policymaking role at the local level. For instance, Wong (2007) highlights the important role state legislatures play in establishing an integrated governance model involving city mayors and other civic leaders

(p. 12). State Board of Education members, appointed by a Governor or legislative committees, supervise the state superintendent of public instruction in most states, as well as the administrative staff responsible for policy implementation. Furthermore, State Superintendents may establish accountability over such matters as the ideal pupil-teacher ratios, length of school years, teacher certification, testing standards, and other important technical matters with budgetary significance.

Outside of imposing mandatory spending levels, states also establish limits on school systems revenue sources and other financing activities. The flexibility afforded to school systems to cultivate taxpayer resources varies significantly from state to state. Some states allow administrators to issue debt without voter authorization, others do not. In states where voter authorization is required, voting thresholds necessary to approve certain forms of debt can vary significantly from state to state. For instance, most states place limits on the size of operational property tax levies, as well as caps on school systems' long-term indebtedness outstanding in relation to the assessed value of property.

In an effort to promote accountability of taxpayer resources, states also impose uniform financial accounting and reporting systems for schools. A few states go so far as to require government auditors to examine districts' compliance with laws and regulations. Others charge oversight bodies with the responsibility of monitoring the overall financial condition of the entity. State entities often intervene into a school system's affairs when the state deems the district underperforming, insolvent, or in violation of the law. Simply put, we do not recommend that a researcher examine a school system's financial condition without first familiarizing themselves with the external state-level constraints.

Local Decision-Makers Day-to-day operating authority, such as the selection and dismissal of teachers or operating decisions, are functions traditionally carried out at the local level. School boards, for instance, establish school district policy. Board members also hire district leadership, approve management appointments, identify maintenance and capital financing needs, and assess instructional demands relative to local resources. Most boards also hold significant financial planning and oversight responsibilities. For instance, boards may establish minimum reserve policies to adjust to unexpected emergencies or untimely tax collections.

The superintendent of a given district is charged with executing the policy directives of the board and managing the administrative apparatus of the school system, including planning and control of the budget. Administrators and board members alike must balance the interests of many stakeholders, including teacher unions, businesses, and parents. As yet another stakeholder, teachers' bargaining units often use financial studies to assess the adequacy of their pay. Watchdog groups rely on studies to assess how effectively districts utilize taxpayer resources.

Capital Markets The financial market, which includes individuals inside and outside of the jurisdiction, also demands financial data. Investors purchase school district debt, not necessarily for altruistic purposes, but to earn a financial return. School system debt is tax-exempt to investors thereby making these instruments quite attractive; however, these stakeholders want reasonable assurance that school systems will meet future payment obligations. Reliance is often placed upon bond rating agencies, like Moody's, Fitch, and Standard & Poor's, to assess default risk.

National, state, and local stakeholders of school systems demand financial studies to make important choices. Whether the decision is to buy bonds of a district, send children to a given school, or award a federal grant, stakeholders rely on financial information to take action. In short, financial management studies empower decision-makers with information that enables them to assess the organizations' near-term and long-term fiscal health.

KEY RESEARCH CONTRIBUTIONS

School finance, as an independent area of research in the United States, emerged in earnest within the progressive movement of the 1920s. During this era, people demanded higher accountability from their public institutions. As such, new researchers may benefit from not only surveying the school finance literature, but may also find value in drawing from diverse outside literatures—including economics, political science, accounting, and public administration—to most effectively develop a research plan. With methodological innovations in the field and data sources, new findings are emerging as researchers make sense of the increasingly complex relationship between finance and student achievement. We suggest that one way to categorize scholarship related to financial management is to

think about three primary school business management activities: planning, control, and decision analysis.

Planning Of the three domains, school district budgetary planning research is the most prolific. As the research community has sought to address the information needs of principals and other administrators, researchers have produced a plethora practical budgeting guides. Several books have specifically focused on how to guide the budgeting process, from inception to completion (see Hartman 2003; Mestry and Bisschoff 2009). Researchers can also find planning research through national or state professional associations, such as the Center for Priority-Based Budgeting (CPBB), California School Board Organization (CASBO), or Government Finance Officers Association (GFOA). GFOA, as one example, has an entire series of step-by-step guides on appropriate school budgeting best practices.

Empirical research can also be found in journals, such as *Education Finance and Policy*, *Public Budgeting & Finance*, and especially the *Journal of Education Finance*. Planning research includes evaluating strategic decisions like whether a school should privatize or consolidate. School district reorganizations, as one example, were a major topic of research in the 1950s (Brewton and Tyler 1951) with cost-benefit studies reaching their heyday in the late 1980s (Rogers 1987; Strang 1987). Today, these studies have re-emerged as cost-effectiveness studies (Levin et al. 2018). Other recent studies have found consolidations are less common as they often fail to deliver the cost savings and performance outcomes anticipated (Berry and West 2010; Duncombe and Yinger 2010; or DeLuca 2013).

Control Research within the control domain bridges the gap between applied versus theoretical challenges confronting the field. As an example, Allison and Johnson's (2015) textbook focuses on the unique financial accounting issues of a school system from a practical perspective. Mead (2012) highlights common financial analytics unique to school systems and how best to interpret financial condition assessments. Researchers can also find meaningful cost accounting guidance, of a practitioner-focus, from the FASB (2005) (see SFFAC Concept #4), Office of Management and Budget, and Government Accountability Office (see GAO (2009)).

Academics have also carried out internal control research, which focuses on the study of the processes and systems used to assure operational effectiveness and efficiency. Examples of such research include examining school district cash management practices (Dembowski 1980; Neu et al. 2002; Trussel and Patrick 2012) and perfecting financial trend monitoring techniques (Bruck and Miltenberger 2013). Other cost control research can be found in the *Journal of Education Finance* (see Ingle et al. 2011) and *State & Local Government Review* (see Rivenbark 2005). Within this research, researchers note how school districts manage and account for their resources. This work offers insights into the day-to-day financial management practices of school districts.

Decision Analysis Decision analysis research focuses on the use of financial data to empower stakeholders—especially administrators—to make sound and practical judgments about setting a climate to enhance instructional outcomes. School business managers must make determinations about whether to maintain or replace buses, drop or expand food service programs, centralize or decentralize the storage of materials, associate teacher pay with performance, as just a few examples. Decision-making research often intersects with the rich pool of economics and policy research (Simon 1976, etc.). The vast majority of research in this area focuses on appropriate resource allocation at the local level.

Empirical research across this domain is varied, with most studies focused at the state and district-level. For instance, Verstegen (2013) explores the different methods used to meet constitutional requirements of funding equity within the State of Nevada. Sonsteli (2008) performed a direct simulation using California educators and administrators to see what resource allocation choices individuals make. Baker et al. (2008) find that the No Child Left Behind (NCLB) act punished schools for having higher standards, which caused some schools to re-organize their resource allocation accordingly. Other research has focused on school district capital investment activities. Gronberg et al. (2011) explore whether school districts are overcapitalized or need to modernize facilities to achieve higher education outcomes. Across these studies, many researchers point to the importance of external factors in shaping district-level responses to resource allocation. As such, these studies demonstrate the utility of decision analysis studies in K-12 education.

Although not a complete review of the literature, the sources highlighted within this section represent the diversity of research in the public financial management field. Given all these different avenues, the diffusion of knowledge across multiple disciplines can stymie new researchers. Hence, a straightforward approach toward conducting fiscal research studies holds vast benefits for many individuals.

STRATEGIES FOR PLANNING AND EXECUTING FINANCIAL RESEARCH

Research Design A research design provides a roadmap for the researcher. Specifically, it allows the researcher to respond to the following questions: What research objectives will be achieved? What sources of data will be used? What methods will be deployed to meet the research objectives? How will the data be analyzed once it is collected? Before embarking on the research design, a researcher should survey the literature to identify areas of research need. Reviewing past works helps bolster one's confidence in the subject matter and can offer new approaches to use in your own study. We have already provided some examples of how a researcher might undertake this within the field of financial management research.

We place emphasis on the most common tasks a researcher is likely to deploy when designing a financial management study. We frame our discussion of financial research tasks around five fundamental steps, which can be remembered using the acronym "RADAR":

Research objectives clearly specified;
 Appropriate data sources leveraged;
 Dataset screened and readied for analysis;
 Analytical techniques applied appropriately; and,
 Results interpreted and professionally summarized.

Step One: Research Objectives Clearly Specified Based on a study's specific research objectives, a researcher will encounter a series of choices about the design of their study. The initial step of any fiscal research project should entail identifying the intended audience of the report and establishing research objectives. The literature review should help identify research questions that the study will address and dictate the appropriate

analytic to apply. Researchers should also attempt to connect their topic to a theoretical framework. From the three general domains previously discussed, a researcher must also specify their research objective by considering potential gaps in knowledge. At this early stage in the process, researchers need to ensure that the project has been properly scoped and verify that the research problem is indeed salient enough to warrant further investigation.

Step Two: Appropriate Data Sources Leveraged Fiscal analysis studies can be performed using either publicly available data, private financial information obtained from the school district, or a combination of the two. There are advantages and disadvantages to both data sources. Publicly available data is accessible and, in some cases, has been verified by an independent third party, such as an auditor. However, private datasets, such as budgets, forecasts, primary source documents, can be advantageous for certain research objectives as they can add insight into the organization's performance and identify how the information links to the organization's strategic performance objectives. Both types of data can provide an in-depth, detailed assessment of the fiscal condition and capacity of an organization. Before making use of internal school district data, the researcher should establish client-data sharing and confidentiality agreements. There is some additional sensitivity in K-12 research given some data may involve information about individual children. Minors and their families are entitled to certain legal protections, including but not limited to Family Educational Rights and Privacy Act (FERPA) rights. Since the availability of private datasets varies by organization and project, we focus on public data sources for our examples.

As public entities, a researcher will find no shortage of suitable data sources for school-related research. Financial data is made readily available at the federal and state level, as well as many research organizations and universities. Some of the leading organizations focused on education research include the Consortium for Policy Research in Education (CPRE), Center on Reinventing Public Education (CRPE), American Institutes for Research (AIR) and Policy Analysis for California Education (PACE). Within higher education, there are several leaders in this field such as Wisconsin University's Center for Education Research (WCER), Stanford University's Center for Education Policy Analysis (CEPA), and Indiana University's Center for Evaluation & Education Policy (CEEP).

Researchers will find a plethora of fiscal research of an empirical nature through most of these outlets.

At a federal level, the National Center for Educational Statistics (NCES) publishes a financial survey for each state. While there are currently no comprehensive benchmark values based on up-to-date financial data, researchers can extract data from this source and perform their own analysis. School districts who have issued bonds also have annual filing requirements with the Securities and Exchange Commission (SEC). The Electronic Municipal Market Access (EMMA) system is an online service that investors and others can use to evaluate the features and risks of particular investment securities, such as the price, amount, and credit rating of particular debt offerings. Contained within this system is the organization's "official statement" that offers rich insight into the financial profile of a given school system.

If a researcher desires more detailed information, he or she might also draw upon data published by the school district itself. Most school systems will make public two documents that will be of relevance to emerging researchers: (i) budgets, (ii) financial reports based on historical spending activities. Often times, these documents are posted to the financial section of the district's website. Operational budgets are prepared by school district administrators to forecast near-term resource inflows (e.g., taxes, grants, etc.) and resource outflows (salaries, fuel, maintenance, etc.). Capital budgets are prepared and published to highlight major, long-term spending related to property, plant, and equipment. These budgets typically have a five-to-ten-year time horizon. Budgets are based on assumptions and intentions to spend money, but do not reflect actual spending activities.

Financial annual reports, often published on State education and finance websites, are among the most reliable sources of information a researcher might use. School financial reports prepared in accordance with Government Accounting Standards Board (GASB) standards typically include certain minimal disclosures, such as a management's discussion and analysis (MD&A) section, basic district-wide and fund financial statements and related notes, and other supplementary statistical information. These reports can be used to answer questions, such as: What is the tax burden placed on the community of operating this school? How much does the school own? To what extent does the district leverage debt? Are there any potential going concern issues? Significant violations of finance-related contractual provisions are also typically disclosed in the

notes to the financial statements. A central value of these reports is the fact that independent auditors have reviewed reports and attest to the fact that the reports have been prepared in accordance with generally accepted accounting conventions.

Step Three: Dataset Screened and Readied for Analysis Before embarking upon data analysis, researchers must be familiar with the accounting standards being applied as they can vary significantly from state to state and even among school districts in the same state of a similar size. In order to make “apple to apple,” as opposed to “apple to orange” comparisons; researchers should select entities using the same accounting methodology for benchmarking activities. Most public school financial statements are reported in a similar format to other governments applying GASB or another closely related Other Comprehensive Basis of Accounting (OCBOA). Under GASB standards, districts apply full-accrual accounting standards at the district-wide level and modified-accrual accounting at the fund-level. The modified-accrual concept recognizes revenues when they are measurable and available, not necessarily when the money is received and paid.

However, not every school district uses full-accrual or modified-accrual accounting. OCBOA standards, as one example, can range idiosyncratically from almost Generally Accepted Accounting Principles (GAAP)-equivalent to cash-based accounting systems. Public school systems often follow specialized accounting and reporting procedures prescribed by state oversight agencies, sometimes referred to as a cash or budgetary basis of accounting. Differences between accrual vs. a budgetary basis of accounting can distort analysis and should be avoided.

Most schools use fund accounting to segregate activities for the purpose of complying with special regulations, restrictions, or limitations of funding sources. Examples of governmental funds include: (1) general fund—used as the main operating fund for the district; (2) special revenue fund—used to account for financial resources restricted by external resource providers for specific purposes; (3) capital project fund—used to account for proceeds earmarked for a capital project or improvements; (4) debt service fund—used to account for financial resources segregated for making principal and interest payments on general long-term debt; and (5) permanent fund—used to account for endowments in which the principal cannot be depleted. The NCES (2007) prescribes revenue and expenditure classifications for public school districts in the United States.

Once a researcher has collected, screened, and classified his or her data, she may consider adjusting the data for the effects of inflation. The consumer price index (CPI) found at <http://www.bls.gov/cpi> is the most common index used to make economic adjustments because the source serves as a proxy for measuring the effects of inflation and regional cost of living differences. Using the CPI to adjust financial data essentially involves adding inflation to or subtracting the measure from all data so that the amounts for the earlier years are stated on the same basis as a selected base year.

Step Four: Analytical Techniques Applied Appropriately The appropriate measurements to apply depend upon one's research objectives, as well as the unique economic, demographic, and political characteristics of a given district. Most financial analysis can be conducted in Microsoft Excel, though some researchers may wish to use a more sophisticated software package. The following section highlights some of the most common methodological choices one may choose to apply when studying school finance across the three major domains of planning, control, or decision-making:

Planning: Equity Measures Planning involves ensuring future cash inflows and outflows are sufficient to meet service demands and in compliance with applicable state law. For instance, many school districts will forecast their property tax revenues given their heavy reliance on this source. A diminished growth rate in property taxes could be indicative of decline in the local housing market; thus, requiring schools to cut back on certain services.

A unique consideration when engaging in budget plans for school systems is whether the funding is distributed equitably. Determining appropriate levels of public education funding is grounded on normative decisions of policymakers, as well as legal precedent. Several notable court cases have established a legal litmus test for the allocation of equitable resources, most commonly when examining state-level distributions.

Various analytic techniques may prove useful when attempting to assess fiscal equity. Berne and Stiefel (1984) identify three common measures of school funding equity: Gini Coefficient, coefficient of variation, and the McLoone index. The Gini coefficient utilizes the Lorenz curve to compare a school district's per pupil spending activities with other districts in similar jurisdictions. Meanwhile, the McLoone index focuses on contrast-

ing expenses the school system incurred in relation to the median of all districts across a jurisdiction. However, some scholars have challenged the applicability of these measures (Toutkoushian and Michael 2005); therefore, one is encouraged to closely study each method before adapting to your study.

Research on Fiscal Control–Fiscal Ratio Measures and Revenue & Expenditure Analysis Once school districts have allocated their resources, the district must effectively manage these resources to maintain sufficient levels of cash and cash equivalents to finance future obligations. Understanding the cost structure of the entity is also an essential step to maintaining fiscal control of the educational enterprise (see Mort et al. 1960). Fiscal and service capacity studies typically examine a district's ability to meet long-term financial obligations to creditors, employees, suppliers, as well as meet service commitments to families and taxpayers. Several authoritative bodies, including GASB, NCES, and major rating agencies, have disseminated best practices in determining whether a system is meeting its control objectives. We examine the three most common analytical techniques applied across this domain: horizontal, vertical, and ratio analysis.

Horizontal analysis entails comparing changes in current period spending to previous years. Significant variations from year to year may highlight changes in funding priorities. To determine percentage change, the year over year dollar change is divided by a given base year. When performing a horizontal analysis, analyze data over a five to ten-year period, a timeframe that is likely to reveal emerging trends or perhaps a change in the national and regional economy. Also try to obtain a holistic picture of significant changes to revenues or expenditures, as well as the potential factors leading to account changes like higher tax rates, population shifts, or a loss of grant funding. Due to the sheer number of accounts one is likely to encounter, focus on those with relatively stable or permanent sources of funding, as opposed to one-time events.

Another control research technique focusing on the relations of accounts is called *vertical analysis*. When using this technique, the researcher will common-size accounts by expressing each as a percentage relative to some base, typically total revenue or total assets. Under this technique, try to identify material accounts (i.e., major sources of revenue or expense, largest assets or liabilities). Assess what impact a change in a particular account might have on the ability of the entity to meet future service demands.

As a final strategy, a researcher may consider conducting a *financial ratio analysis*, a method that entails inter-relating various accounts from the Statement of Revenues Expenditures and Changes in Fund Balance (SRECFB), Balance Sheet, and Statement of Cash Flows. Traditionally, four dimensions of the financial condition are examined: liquidity, solvency, asset management, debt management. While analysts commonly apply financial ratio methods to single entities, this technique can also be helpful in benchmarking performance. For example, one may want to determine industry standards, such as a safe level of cash reserves or proportion of debt to assets.

Research on Fiscal Decision-Making Although the vast body of published academic research has focused on planning and control techniques, researchers have increasingly focused on the outcomes school systems derive given their financial resources (see Chap. 14 of this volume for a further discussion). Researchers who specialize in this area attempt to correlate spending with non-financial performance metrics, such as graduation rates, dropout rates, or test results. The process of correlating expenditures with performance data typically entails isolating operating expenses adjusted for inflation and adjusting non-financial data horizontally across years to allow comparison over time by standardizing data. Often, raw data will be converted to a relative or percentage basis. For instance, one may calculate the z-score so measures can be compared on a “like by like” basis. When engaging in this form of research, researchers need to ensure that adequate evidence exists relating to the appropriate standard to be targeted, perhaps using state or national criteria.

Step Five: Results Interpreted and Professionally Summarized The final step of the research process entails delivering an appropriate conclusion in a clear and logical manner, as well as figures/tables that are understandable. The final product of your research endeavor will be a report that sets forth clearly and precisely what you have accomplished. While the length of the document can vary, readers must be able to grasp what you have discovered (e.g., actual findings versus opinions). It is important that researchers present the data as accurately as possible, which includes addressing the research implications and any limitations to your study.

So that the reader of this report does not get lost in the data presentation, we recommend incorporating a brief discussion of tables and figure

structures and the corresponding results before presenting the information. If the data involved in the report is extensive, we recommend presenting a summary in the main body of the report and attach supplemental tables as an appendix.

CASE STUDY: AN APPLIED ILLUSTRATION OF PERFORMING RADAR STEPS

To highlight the fundamental fiscal research phases, we have selected Puyallup School District (PSD), located in Washington State. According to PSD's website, the district had approximately 23,000 students attending 21 elementary schools, seven junior high schools, three high schools, an alternative school, and an online academy in the 2017–18 academic year. The district operates on approximately \$260 million annually within the general fund budget. Currently, there are about 3000 employees with over half of the workforce being certificated teachers. While the school district is the eighth largest in Washington state, PSD would be considered representative of an average-sized school district in larger states, such as California, Texas, and New York.

Step 1: Research Objectives Clearly Specified The first step a researcher might take is to scan the district's external policy environment. In the case of PSD, judicial mandates and popular initiatives have disrupted the traditional incremental budgetary process. For example, in 2012, the Washington State Supreme Court found the State had violated a paramount duty to educate all children in the *McCleary v. State of Washington* decision. Around the same time, Washington voters also approved *Initiative 1351*, which effectively mandated much smaller classes at the secondary level (Zender 2015). Consequently, the Court recently held State legislature in contempt for not submitting a plan to fund education at "ample" funding levels, arguably a subjective benchmark. These state-level events have both direct and indirect consequences for the district. In 2016, the State Superintendent sued PSD and six other school districts for unlawfully relying on local property tax levies to fund basic education, including supplemental salaries for employees; thereby, enabling the State legislature to continue to underfund the school systems. While PSD cannot predict what the outcome of this lawsuit will be, these external pressures place uncertainty on the district's future financial situation.

As detailed in Washington state law, the school system does not have unlimited independent taxing or borrowing authority. For instance, under State statutes, non-voted debt cannot exceed three-eighths of 1% of the assessed value of taxable property within a school district. Meanwhile, voter approved debt cannot exceed 5%. Bond levies are required to be approved by 60% of those voting and the number of “yes” votes must equal or exceed 40% of those voting in the last general election. School districts in Washington can submit special levies for maintenance and operation for up to four years by gaining only a simple majority, but these funds are also limited to the excess of 24% of General Fund revenue. Currently, the outstanding debt of the district is approximately 2.81% of total assessed value (of which 0.01% is non-voter approved debt outstanding).

Running operational deficits is also not an option for the district. Pursuant to law, the school district’s budgeted ending fund balance cannot be negative. Many portions of PSD’s existing fund balance are set aside based on external restrictions or Board intent; therefore, management discretion is limited. For instance, the School Board has established a minimum fund balance reserve policy that includes a commitment of at least 13 to 15 days of the current year’s operating expenses. Given this restrictive budgetary environment, our research objective is to determine whether the district is capable of meeting future service demands at existing resource levels. Our research question could be stated as follows: *Do recent financial indicators suggest the district has a positive financial outlook?*

Step 2: Appropriate Data Sources Leveraged Like many districts, information related to the business and financial operations is made available on the district’s website (<https://www.puyallup.k12.wa.us/>), as well as other public sources. For instance, the district publishes “budget to actuals” reports on their home page. These reports provide a higher level of detail of financial account activities, offer variance analysis, and other important context behind the numbers of the district. A researcher can also find detailed information relating to debt instruments of the district on the Electronic Municipal Market Access website (<https://emma.msrb.org/>) by doing a keyword search on “Puyallup School District No. 3, Washington.” For the purpose of conducting our simple financial analysis, we rely on the official “audited” financial statements of the district, which are housed on the Washington State Auditor’s Office (<http://portal.sao.wa.gov/reportsearch>).

Step 3: Dataset Screened and Readied for Analysis Once financial statements have been downloaded, the researcher will discover a document approximately 40 pages in length containing the audit opinion and financial statements and related notes for the entity. All school districts within the State of Washington are required to adhere to the uniform Accounting Manual for Public School Districts (AMPSD), which provides the option to prepare full-accrual GAAP-based, OCBOA, or cash-basis (non-GAAP) financials. PSD uses OCBOA or modified-accrual accounting in accordance with AMPSD, the financial framework differs from GAAP in several key respects. For instance, PSD does not prepare district-wide financial statements, debt obligations are reported in the notes to the financial statements, and the management discussion and analysis or budget to actual schedule is not published.

To overcome this limitation, we rely on the total fund activities, a summation of the general fund, capital projects fund, student transportation vehicle fund, debt service fund, and Associated Student Body (ASB) special revenue fund. In conducting our analysis, we will need to be aware that this format does not eliminate inter-fund transfers between government and enterprise funds and represents financial data on a modified-accrual basis. This may have the effect of overstating or understating accounts if they were to be converted to a full-accrual basis.

Step 4: Analytical Techniques Applied Appropriately As illustrated in Table 13.1, we begin by conducting a horizontal analysis. In our case, we use 2012 as a base year for evaluating changes in a financial account over time on the SRECFB (essentially, the income statement of PSD). We examine data over this time horizon to detect periods of stress for PSD. For instance, in FY2012 the district ran \$5 million operating deficits, a scenario where operating expenditures exceed revenues. Although the district ran in the red, this does not mean the district failed to pay its bills. Merely, that reserves from prior years had to be used to cover the difference. This practice is not sustainable. However, in recent years, the district has turned these deficits into operating surpluses.

To examine what caused the operating deficit or surplus in a given period, we might examine which accounts have changed materially. The vertical analysis helps highlight significant changes over a five-year period (e.g., state revenue at 36% and support service expenses at 24%); mean-

Table 13.1 Horizontal analysis—statement of revenues, expenditures, and changes in fund balance

Revenues:	2016	2015	2014	2013	2012	Variance (2016-2015)	%	Variance (2016-2012)	%
Local	\$88,891,225	\$81,702,277	\$74,506,157	\$76,352,069	\$70,897,768	\$7,188,948	9	\$17,993,457	25
State	\$182,318,445	\$159,059,572	\$148,291,301	\$135,073,563	\$133,874,451	\$23,258,873	15	\$48,443,994	36
Federal	\$10,581,385	\$10,319,585	\$9,382,771	\$10,345,620	\$12,944,045	\$261,799	3	-\$2,362,661	-18
Other	\$1,044,925	\$1,101,325	\$970,723	\$1,151,786	\$1,059,701	-\$56,400	-5	-\$14,776	-1
Total Revenues	\$282,835,980	\$252,182,760	\$233,150,952	\$222,923,038	\$218,775,966	\$30,653,221	12	\$64,060,015	29
Expenditures:									
Current:									
Regular instruction	\$130,204,550	\$119,695,084	\$114,942,478	\$109,157,190	\$110,353,429	\$10,509,466	9	\$19,851,121	18
Special education	\$29,538,560	\$27,048,937	\$26,180,804	\$25,307,017	\$25,811,406	\$2,489,622	9	\$3,727,153	14
Vocational education	\$9,127,890	\$8,696,250	\$8,345,784	\$8,066,482	\$7,910,699	\$431,640	5	\$1,217,191	15
Compensatory programs	\$9,172,110	\$8,630,036	\$7,549,411	\$6,208,228	\$7,713,472	\$542,074	6	\$1,458,637	19
Other instructional programs	\$437,407	\$392,272	\$429,974	\$416,408	\$398,508	\$45,135	12	\$38,899	10
Community services	\$1,018,556	\$1,030,797	\$961,118	\$827,534	\$819,466	-\$12,241	-1	\$199,089	24
Support services	\$46,252,598	\$43,566,284	\$44,642,769	\$41,052,656	\$39,854,020	\$2,686,314	6	\$6,398,577	16
Student activities	\$2,280,261	\$2,738,589	\$3,016,133	\$2,872,489	\$2,919,948	-\$458,328	-17	-\$639,687	-22
Capital outlay:	\$15,170,186	\$10,599,666	\$11,556,225	\$6,343,251	\$7,633,558	\$4,570,519	43	\$7,536,627	99
Debt service:	\$25,877,686	\$19,338,898	\$20,604,713	\$20,343,540	\$20,465,345	\$6,538,788	34	\$5,412,342	26
Total expenditures	\$269,079,803	\$241,736,813	\$238,229,410	\$220,594,795	\$223,879,853	\$27,342,989	11	\$45,199,949	20
Revenues over (under) expenditures	\$13,756,178	\$10,445,947	-\$5,078,458	\$2,328,243	-\$5,103,888	\$3,310,231	32	\$18,860,066	-370
Total other financing sources (OFS)	\$80,276,595	\$63,063	\$39,529	\$468,501	\$436,233	\$80,213,532		\$79,840,362	
Excess of revenues/OFS	\$94,032,773	\$10,509,010	-\$5,038,929	\$2,796,744	-\$4,667,655	\$83,523,763		\$98,700,428	
Fund balance, beginning	\$54,590,388	\$44,081,378	\$49,120,307	\$46,323,563	\$50,991,218	\$10,509,010	24	\$3,599,170	7
Fund balance, ending	\$148,623,161	\$54,590,388	\$44,081,378	\$49,120,307	\$46,323,563	\$94,032,773	172	\$102,299,597	221

Note: All calculations derived by the chapter authors using public information Source: Washington State Auditor's Office. *Financial Statements and Federal Single Audit Report: Puyallup School District No. 3: 2016-2012.* Olympia, WA: Washington State Auditor's Office

while, the percentage changes highlight the changes that are likely special or one-time transactions (e.g., transportation equipment and bond/levy issuance). Trend percentage analysis suggests rapid growth in building spending and declines in land acquisitions, suggesting the district is nearing completion on construction of new facilities.

Once a researcher has conducted a general analysis, he or she may want to drill-down into the details. For instance, what specific sources of federal revenue have declined in relation to total revenue and what programs are likely impacted? When performing this analysis of the budget, we find the district receives its major federal funding from Special Education, Title I, Title II, and the National School Lunch Program, as well as various other special purpose programs. As State monies have increased, what is driving this trend? State funding is based primarily on average full-time equivalent student enrollment which has been increasing within the district. In 2016, PSD received about \$9 million in Local Effort Assistance Program (LEA) monies, which are used to equalize the tax burden in districts with low property values.

When shifting one's attention to expenditures, we notice that virtually every category of operating expense is increasing, except for student activities. Expenditures are a rough measure of a school district's service output. The costliest operating line item of any school is salary and benefits to employees. Consequently, PSD has entered into 11 different collective bargaining agreements. If a researcher was given access to internal PSD documentation, these agreements could be carefully screened for recent increases to salaries, vacation, sick leave, and medical benefits.

As highlighted within the Table 13.2, we examine each account in relation to total assets; thereby, highlighting the relative importance of investments to the district (78% of total assets). When a district issues a bond, monies are temporarily invested over the period of time buildings are being constructed. Given the size of this account, we may want to survey whether the district is engaged in prudent investment management practices. For example, Chapter 39.59 Revised Code of Washington (RCW) limits the investment of public funds by local governments to municipal bonds, warrants of local governments, certificates of deposit, and any investments authorized by law for the State Treasurer like U.S. Treasury Bills, utility revenue bonds, and federal agency instruments.

Table 13.2 Vertical or common-size analysis—balance sheet (FY 2012 & 2016)

	2016	%	2012	%
Assets:				
Cash and cash equivalents	\$4,810,493.34	2.5	\$7,963,486.95	9.7
Minus warrants outstanding	\$(3,859,150.25)	-2.0	\$(6,347,238.49)	-7.8
Taxes receivable	\$38,332,844.30	19.6	\$30,919,576.16	37.8
Due from other Funds	\$331,793.86	0.2	\$44,074.62	0.1
Due from other governmental units	\$1,702,337.67	0.9	\$1,614,526.86	2.0
Accounts receivable	\$50,182.05	0.0	\$14,635.94	0.0
Inventory	\$642,815.48	0.3	\$1,423,356.83	1.7
Prepaid items	\$928,549.32	0.5	\$773,511.71	0.9
Investments	\$151,616,000.00	77.7	\$45,455,000.00	55.5
Investments/Cash with Trustee	\$594,886.71	0.3	\$-	0.0
Total Assets and Deferred Outflows	\$195,150,752.48	100	\$81,860,930.58	100
Liabilities:				
Accounts payable	\$6,195,168.24	3.2	\$3,525,807.84	4.3
Accrued salaries	\$-	0.0	\$527,257.80	0.6
Payroll deductions and taxes	\$1,060,259.53	0.5	\$250,986.61	0.3
Due to other governmental units	\$24,052.17	0.0	\$-	0.0
Due to other funds	\$331,793.86	0.2	\$44,074.62	0.1
Unavailable revenue- Taxes receivable	\$38,826,317.97	19.6	\$31,189,240.44	38.1
Total Liabilities and Deferred Inflows	\$38,826,317.97	19.9	\$31,189,240.44	38.1
Fund balance:				
Nonspendable fund balance	\$1,571,365.32	0.8	\$2,196,868.54	2.7
Restricted fund balance	\$97,758,728.82	50.1	\$21,794,090.18	26.6
Committed fund balance	\$39,340,128.34	20.2	\$9,905,377.00	12.1
Assigned fund balance	\$5,588,970.00	2.9	\$5,239,918.23	6.4
Unassigned fund balance	\$4,453,968.23	2.3	\$7,187,309.32	8.8
Total fund balance	\$148,713,160.71	76.2	\$46,323,563.27	56.6
Total liabilities, deferred inflows, and fund balance	\$195,150,752.48	100	\$81,860,930.58	100
Assessed Property Tax Value	\$12,484,865,009		\$11,361,019,610	
Students	-21,409		-20,377	

Note: All calculations derived by the chapter authors using public information

Source: Washington State Auditor's Office. *Financial Statements and Federal Single Audit Report: Puyallup School District No. 3: 2016-2012*. Olympia, WA: Washington State Auditor's Office

With respect to liabilities, two significant long-term liabilities are not reflected on the district's financial statements under OCBOA standards: long-term debt and pension obligations. Under modified-accrual accounting, only the current year payments are reflected (e.g., bond principal and interest and immediate term pension contributions). While at its face this practice may appear misleading, the State of Washington is ultimately the guarantor of long-term liabilities of the district and, therefore, the district only needs to focus on disclosing current year obligations. Article VIII, Section 1 of the Constitution of the State, allows the State to guarantee any voted general obligation bonds issued by a school district and the \$17 million pension contribution made by PSD in FY 2016 represented the full liability due to the State Department of Retirement System. Therefore, long-term liabilities do not need to be factored into our short-term analysis.

We further see that a substantial portion of the district's residual equity is either deferred (e.g., tax receivables) or restricted (e.g., designated for specific activities by legislature or grantors). A restricted fund balance is legally earmarked based on externally dictated terms, usually referenced in state law, grants, or bond contracts. For example, a state may require that gas tax revenues to be used only for transportation equipment acquisitions by schools. After accounting for these restrictions, unassigned fund balance in FY 2016 was \$4.5 million in FY 2016.

Before concluding our vertical and horizontal analysis, we may want to consider what effects, if any, inflation had on the changes of our account values. For example, while the change in State revenues over the five-year period appears significant, at 36%, once we take into effect the changes in the consumer price index, the uptick in State funding is less dramatic. When our statements are expressed in nominal format, we find State funding only increased 28%; meanwhile, some expenditures like Special Education only slightly increased at 8%. Adjusting for the effects of inflation can dramatically change our outlook in periods of rising prices.

Table 13.3 contains a summary of key school district measures according to the NCES (2007) and the respective formulas for each ratio. Some stakeholders may be more focused on certain measures than others. For instance, school business managers interested in ensuring they can meet payroll obligations may focus on liquidity measures; whereas, bond investors, interested in the district's ability to repay debt, may focus on solvency and debt management indicators. With the exception of the debt management metrics, a higher ratio is generally considered to be more favorable.

The first two indicators, which focus on the district's solvency or the ability of the school district to withstand future financial emergencies,

Table 13.3 Financial condition indicators

<i>Indicator</i>	<i>Formula</i>	<i>2016</i>	<i>2015</i>	<i>2014</i>	<i>2013</i>	<i>2012</i>
Financial position	$(\text{Assets}-\text{Liabilities})/\text{Total revenues}$	0.66	0.36	0.33	0.37	0.35
Change in financial position	$(\text{Fund balance, end}-\text{Fund balance, beg.})/\text{Total revenue}$	0.33	0.04	-0.02	0.01	-0.02
Current ratio	Current assets/Current liabilities	5.52	7.12	4.49	8.76	8.19
Quick ratio	$(\text{Cash and short-term investments})/\text{Current liabilities}$	0.63	1.03	0.88	2.03	1.83
Debt to assets ratio	Total liabilities/Total assets	0.04	0.06	0.10	0.05	0.05
Debt to fund balance ratio	Total liabilities/Fund balance	0.05	0.10	0.18	0.09	0.09
Taxable property per student	Total taxable property value/Total full-time equivalent students	\$583 k	\$556 k	\$527 k	\$518 k	\$558 k
Property tax revenues per property value	$(\text{Total local property tax revenues}\times 100)/(\text{Total assessed property value})$	0.71	0.70	0.69	0.73	0.62
Taxes per student	Total local tax revenues/Total full-time equivalent students	\$4.2 k	\$3.9 k	\$3.6 k	\$3.8 k	\$3.5 k
Debt per \$100 assessed property value	Total liabilities $\times 100$ /Total assessed property value	0.06	0.05	0.07	0.04	0.04
Debt per student	Total liabilities/Total full-time equivalent students	\$355	\$262	\$393	\$208	\$213

Note: All calculations derived by the chapter authors using public information

Source: Washington State Auditor's Office. *Financial Statements and Federal Single Audit Report: Puyallup School District No. 3: 2016-2012*. Olympia, WA: Washington State Auditor's Office

appear to be increasing. As such, we view this as a favorable trend. The current and quick ratios reflect the short-term cash position of the district or liquidity. The trend for these two factors has been downward sloping or negative since 2012. If this trend continues, the district may experience difficulties paying their bills. With respect to debt management, the district appears to have reached an apex in debt reliance in 2014 and has since reduced its short-term debt obligations. While these are general trends of the district, a researcher would be well advised to benchmark this data

with districts of a similar size and demographic profile to compare PSD with their peer group.

Step 5: Results Interpreted and Professionally Summarized While this case study is not intended to represent a comprehensive research report, we provide a snapshot of our aggregate findings in proper form (see Tables 13.1, 13.2, 13.3). The suggested analytical results are structured on single page tables that are easy for the reader to interpret. Before presenting each table, we offer a brief introduction to the accompanying results. When evaluating trends as a whole, there are more positive signs than negative. Revenues are increasing, financial condition indicators are improving, and the district maintains a healthy fund balance reserve. In light of these facts, we can conclude PSD is financially sound. This view appears to be reinforced by credit rating agencies Standard and Poor, and Moody's, who recently critically assessed the financial health of the enterprise. When the district went to market with a \$212 million dollar bond issue, the district received an "Aa1" and "AA+" score, which is considered a high-quality, safe investment grade (Official Statement 2017).

RECOMMENDATIONS FOR NOVICE RESEARCHERS

We conclude by offering a few pieces of advice for researchers who are new to the field of financial management research. First, although we describe data for school districts that is largely available in electronic form, we caution the reader that this information is not always digitized and thus may require some effort to convert. Indeed, we think it is important that researchers think about the time and cost associated with the activity of data collection when conducting such an analysis. Most states have some form of centralized database via the State Department of Education on both finances, as well as education inputs, outputs, and outcomes. If one experiences trouble finding information, he or she should connect with an individual who has completed similar research projects, published in a relevant journal, or is involved with a research grant. Consider inquiring into the specific tools and resources this individual uses.

Second, researchers should exercise caution when undertaking studies across different states. State context matters. The accounting and budgeting systems may be significantly different making comparability across state contexts infeasible. However, we also note that a creative researcher could look at the types of variations across states to investigate how the

various regulations or practices might impact finances—or how finances might impact educational outcomes. If not, many studies can be structured within a single state.

Finally, and most importantly, keep in mind that financial statement analysis is more of an art than a science. There are relatively few absolutes when one assesses the financial status of a school district because different people are interested in different aspects of financial health. While some may be concerned with the cost of the system, others may be focused on a school's ability to repay long-term debt. Even if two people are interested in the same financial issue, their judgments remain subjective. For instance, is a 1.25 current ratio appropriate? The answer likely varies by person. Fortunately, there are ways to bolster the validity of the conclusions if you focus on clarifying definitions, context, and scope.

CHAPTER SUMMARY AND CONCLUSION

This chapter explored the unique finance and accounting considerations an emerging researcher must consider when evaluating school district performance, as well as common analytical research techniques one is likely to utilize when preparing fiscal research studies. Researchers must be confident in not only reading financial statements and budgets, but in interpreting and analyzing accounting data. An emerging researcher interested in exploring this subject in greater depth is encouraged to study the GASB codification for a greater understanding of the major financial statement elements and reporting requirements of school districts. Cost accounting textbooks may offer additional background into equity and accountability issues, as well as best practices in improving internal controls over the use of public resources.

Recommended Readings

Hartman, W. T. (2003). *School district budgeting*. Lanham: Scarecrow Education.

In this volume, Bill Hartman provides an overview of school district budgeting practices. This textbook is helpful for scholars without previous experience in school district financial management. The textbook addresses issues related to budgeting, accounting, and fiscal management.

Odden, A. R., & Picus, L. O. (2013). *School finance: A policy perspective*. (5th Ed.). New York: McGraw-Hill.

In this textbook, Odden and Picus provide an overview of school finance from a policy perspective. The volume addresses key concepts in school finance practice and policy, including those related to education law, adequate funding, connecting resources, and student learning, and efforts to redesign school finance system. This textbook may be especially helpful for novice scholars and those without an in-depth understanding of school finance.

REFERENCES

- Allison, G. H., & Johnson, F. (2015). *Financial accounting for local and state school systems: 2014 edition* (NCES 2015-347). Washington, DC: U.S. Department of Education National Center for Education Statistics (NCES).
- Baker, B., Taylor, L., & Vedlitz, A. (2008). Adequacy estimates and the implications of common standards for the cost of instruction. *National Research Council, 9*(2), 24-38.
- Berne, R., & Stiefel, L. (1984). *The measurement of equity in School Finance: Conceptual, methodological and empirical dimensions*. Baltimore: John Hopkins University Press.
- Berry, C. R., & West, M. R. (2010). Growing pains: The school consolidation movement and student outcomes. *Journal of Law, Economics, & Organization, 26*(1), 1-29.
- Bruck, E., & Miltenberger, L. (2013). A school district condition assessment system and its application to Pennsylvania school districts, *University of Illinois Press, 39*(2), 115-131.
- DeLuca. (2013). K-12 non-instructional service consolidation: Spending changes and scale economies. *Journal of Education Finance, 39*(2), 150-173.
- Dembowski, F. L. (1980). School district cash management programs. *Journal of Education Finance, 6*(1), 51-67.
- Duncombe, W. D., & Yinger, J. M. (2010). School district consolidation: The benefits and costs. *School Administrator, 67*(5), 10-17.
- Federal Accounting Standards Advisory Board. (2005). *Statement of federal financial accounting standards #4* (2005). Washington, DC: U.S. Printing Office.
- Government Accountability Office. (2009). *GAO cost estimating and assessment guide: Best practices for developing and managing capital program costs*. Washington, DC: U.S. Government Accountability Office (GAO).
- Gronberg, T. J., Jansen, D. W., & Taylor, L. L. (2011). The impact of facilities on the cost of education. *National Tax Journal, 64*(1), 193-218.
- Hartman, W. T. (2003). *School district budgeting* (2nd ed.). New York: R&L Education.

- Ingle, W. K., Petroff, R. A., & Johnson, P. A. (2011). Estimating resource costs of levy campaigns in five Ohio school districts. *Journal of Education Finance*, 37(1), 52–71.
- Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2018). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). Thousand Oaks: SAGE.
- Mead, D. M. (2012). *What you should know about your school district's finances: A guide to financial statements*. Norwalk: Government Accounting Standards Board.
- Mestry, R., & Bisschoff, T. (2009). *School financial management explained* (3rd ed.). Cape Town: Pearson.
- Mort, P., Reusser, W. C., & Polly, J. W. (1960). *Public school finance*. New York: McGraw-Hill.
- National Forum on Education Statistics. (2007). *Forum guide to core finance data elements*. Washington, DC: U.S. Department of Education National Center for Education Statistics (NCES).
- Neu, D., Peters, F., Taylor, A., Neu, B. D., Peters, F., & Taylor, A. (2002). Financial reforms in Alberta?: The impact on School Districts. *Journal of Education Finance*, 27(4), 1067–1083.
- Odden, A. R., & Picus, L. O. (2013). *School finance: A policy perspective*. New York: McGraw-Hill.
- Official Statement. (2017). *Puyallup school district's unlimited tax general obligation bond: 2017*. Puyallup: Puyallup School District.
- Rivenbark, W. C. (2005). A historical overview of cost accounting in local government. *State & Local Government Review*, 37(3), 217–227.
- Rogers, R. G. (1987). Is Big Better? Fact or Fad Concerning School District Organization. *ERS Spectrum*, 5(4), 36–39.
- Simon, H. (1976). *Administrative behavior*. New York: The Free Press.
- Sonstelie, J. (2008). Resource needs of California public schools: Results from a survey of teachers, principals, and superintendents. *Education Finance and Policy*, 3(1), 58–89.
- Strang, D. (1987). The administrative transformation of American education: School district consolidation, 1938–1980. *Administrative Science Quarterly*, 32(3), 352–366.
- Toutkoushian, R. K., & Michael, R. S. (2005). Demystifying school funding in Indiana. *Education Policy Brief*, 3(2), 1–17.
- Trussel, J. M., & Patrick, P. A. (2012). Predicting significant reductions in instructional expenditures by school districts. *Journal of Education Finance*, 37(3), 205–233.
- U.S. Census Bureau. (2012). *Population of interest-school district governments and public school systems*. Retrieved August 8, 2017, from https://www.census.gov/govs/go/school_govs.html

- Verstegen. (2013). Leaving equity behind? A quantitative analysis of fiscal equity in Nevada's public education finance system. *Journal of Education Finance*, 39(2), 132–149.
- Washington State Auditor's Office. (2016). *Financial statements and federal single audit report: Puyallup school district No. 3: 2016-2012*. Olympia: Washington State Auditor's Office.
- Wong, K. K., Shen, F. X., & Anagnostopoulos, D. (2007). *The education mayor: Improving America's schools*. Washington, DC: Georgetown Press.
- Zender, J. (2015). Outside pressure: How educational financing reforms Circumvented Washington State Legislature. *National Social Science Association Conference Proceedings*, 59(1), San Diego.



Evaluating Education Policy & Program Costs

Tammy Kolbe and Rachel C. Feldman

A critical issue facing educational leaders and policymakers is how to invest scarce resources to improve student learning and educational experiences. Decision makers not only require information on potential costs so that they can ensure sufficient resources are in place to support reforms, they also need to understand how their decisions might influence both the types and amounts of resources needed to effectively implement reforms. Comparing costs with program effects is useful for choosing among different reforms to ensure that selected alternatives either provide the best results given available resources or achieve similar results at a lower cost.

Given the importance of costs in decision making, there have been increased calls to apply cost analysis to understanding the resources, costs,

T. Kolbe (✉)
Educational Leadership & Policy Studies, University of Vermont,
Burlington, VT, USA
e-mail: tkolbe@uvm.edu

R. C. Feldman
School of Education and Social Policy, Northwestern University,
Evanston, IL, USA
e-mail: rachel.feldman@northwestern.edu

and relative efficiency of education reforms (Ingle and Cramer 2013; Levin and Belfield 2015; Rice 1997). Yet, despite the important role cost analysis plays in decision making, efforts to systematically analyze costs occur less frequently than other types of evaluation (Rice 2002). The extent of its use stands in stark contrast to the considerable number of evaluations that examine the impact of educational policies and programs and the sophisticated methodological literature base on which that work is based (Belfield 2015). This discrepancy between the availability of cost and impact studies has resulted in decision makers having a better sense of which education reforms are more and less effective, but without crucial information about what these reforms cost or their relative efficiency in producing desired results (Belfield 2015; Rice 1997).

In part, cost analysis may be underutilized because of researchers' lack of understanding about the methods available for estimating costs and how to design and implement education cost studies (Levin and McEwan 2001, 2002; Rice 1997). The purpose of this chapter is to introduce researchers to education cost analysis and strategies for incorporating cost-related research in their work. We begin by defining education cost analysis as a methodological approach and describing the types of research questions it is best suited to answer. We then introduce the methods used to estimate costs and provide an example demonstrating how cost analysis was integrated with an evaluation of a school reform model. The chapter concludes with recommendations for conducting cost studies in educational contexts.

REVIEW OF RELEVANT LITERATURE

Conceptualizing Cost Education cost analysis identifies and values the resources required to implement educational policies, programs, and systems (Belfield 2015). Resources can be thought of as packages, comprised of different types and quantities of ingredients that go into producing observed outcomes (Rice 1997). For instance, the resources involved in delivering a school-wide literacy intervention may consist of the time spent by literacy specialists, classroom teachers, and even parents and students, as well as other tangible resources such as curricular materials, technology, and dedicated space within the school building. Each resource has a value that can be articulated as a dollar amount, or program cost.

This process of accounting for the value of the full package of resources involved in implementing education policies and programs is grounded in the economic concept of opportunity cost (Belfield 2015). Opportunity

costs account for potential alternate uses of the limited resources, necessitating choice between competing priorities. From an economic perspective, decisions about how to allocate resources are predicated on the idea that resources have a potential alternative use, invoking the ideas of scarcity and choice in how decisions are made to allocate resources (Friedman 2002). For instance, teachers could allocate their contract hours to developing and delivering a literacy intervention, *or* they could use this same time for other instructional activities; new programs may require the school to hire additional staff, *but* the job candidates have alternative employment opportunities; classroom space could be used for grade level instruction *or* it could be reserved as a dedicated space for a new school-initiative (e.g., “maker space”; science lab; resource room for students with disabilities).

The cost of a particular resource is the value forgone by not using that same resource for its next best alternative purpose. This cost represents the payment necessary to maintain a resource in its current use (Friedman 2002). To measure this cost, consider the valuation of a teacher’s time as the dollars given up by not “selling” the teacher’s time to the highest bidder. In practical terms, presuming labor market competition, a teacher’s compensation (salary and benefits) would be the price paid for the corresponding value of her time. The same logic applies to valuing classroom space. Assuming the classroom holds the same value in the real estate market as space of similar size and quality, its value constitutes the prevailing cost of renting similar space at another location.

Conceptualizing educational costs from the “bottom up”—i.e., by first identifying the quantity and quality of resources necessary to implement a policy or program and subsequently assigning value to these resources—is theoretically and methodologically distinct from how accountants view cost (Chambers 1999; Hartman et al. 2001). The accounting approach focuses on what organizations *spend* to implement a program or deliver goods and services during a given period. Here, cost means the sum of accumulated expenditures represented in day-to-day financial transactions. Accounting data become de facto historical “accounts” that chronicle the flow of money into and out of an organization. At their most useful, accounting data track spending and revenues and hold organizations accountable for their spending. Yet, these accounting requirements, because they lack the structure to measure cost by specific activities or programs, cannot answer fundamental questions about a program’s value or explain how it deploys resources to accomplish its goals.

The distinction between how economists and accountants view the concepts of opportunity cost and expense has implications for framing and addressing research questions (Table 14.1). Researchers adopting an economic perspective frame questions in terms of the cost of the resources consumed and the distribution of the cost burden across stakeholder groups. In contrast, the accounting perspective prioritizes questions about expenditure allocation and sources of revenue.

Limitations of Using Expenditure Data to Estimate Education Costs Issues arise when using expenditures to estimate education policy and program costs because financial data may not resemble actual program delivery (Chambers 1999). How expenditure information, derived from state, district, and school accounting systems, are categorized and aggregated make it difficult to link spending with particular school-based activities or interventions (Chambers 1999; Hartman et al. 2001; Levin and McEwan 2001). Financial cost accounting chronicles expenditures on core functions (e.g., instruction, administration) and objects (e.g., personnel and facilities). This accounting framework, however, frequently does not mirror the structure of education policies and programs. Thus, the difficulty lies in identifying expenditure information in ways that reflect an intervention's component parts (Chambers 1999; Hartman et al. 2001; Levin and McEwan 2001). For example, in their cost study of Massachusetts' Expanded Learning Time (MA ELT) Initiative, O'Reilly and Kolbe (2011) found that the state could not reliably parse out what it spent to implement the program. Spending on schools' extended school day programs cut across function and object classifications used in their accounting system, making it impossible to establish links between school expenditures and certain activities (e.g., new academic and non-academic enrichment activities) that comprised these programs.

Expenditure information also may not account for the full range of resources schools use to implement policies and programs (Belfield 2015; Levin and McEwan 2001). For instance, financial accounting systems fail to capture time and materials donated or paid for by outside providers. In this instance, programmatic expenditures would underestimate real resource costs. In their companion cost study of the MA ELT Initiative, Kolbe and O'Reilly (2016) showed in one school that the time added to the school day was staffed by AmeriCorps volunteer teachers, making these hours essentially "off the books." Accordingly, the resource-based

Table 14.1 Comparison between accounting and resource-based cost estimation

	<i>Accounting approach</i>	<i>Resource-based approach</i>
Focus	Actual expenditures	Resource consumption
Question asked and answered	How much was spent to implement a policy or program? On what types of things were dollars spent? What sources of revenue were used to pay for a policy or program? On what were revenues spent?	What does a policy or program cost to implement? What types, quantities, and quality of resources were used in implementation? What resources would be needed to replicate the program elsewhere? Who contributed resources to the effort? How was the cost burden distributed across stakeholder groups?
Unit of analysis	Entity covered by financial statement	Program or service delivery system
Data used	Actual expenditures	Physical resources utilized to deliver program or service
Information collected and reported	Dollars spent, by accounting categories	Types, amounts, and qualities of physical ingredients and their values
Data sources	Financial accounting systems	Program documents, personnel files, interviews, direct observations to construct inventory of ingredients
Applications	Summarizing expenditures Determining variation across interventions or implementation sites in types of type and timing of spending Accounting for/monitoring how revenues from certain sources were applied	Estimating total and marginal costs Creating a resource inventory used in implementation Understanding who contributes resources and the distribution of costs across stakeholder groups
Limitations	Accounting conventions limit the ability to perfectly ascribe expenditures with programs and services that cut across fund, function, object, and activity expenditure categories Unable to separate the effects of price, quantity, capacity, and mix of resources used to achieve results	Absence of correspondence with program budgets and balance sheets Depending on program scope and scale, creating resource inventories can be cumbersome and time consuming

Source: Adapted from Hartman et al. (2001)

cost of implementing the school's extended day program revealed the program cost nearly twice what the school reported on its balance sheet. As this example shows, in the absence of detailed information about *both* resources and costs, policymakers and schools may misjudge the resources required and the corresponding costs associated with implementing educational policies and programs. Accounting systems also frequently fail to disaggregate expenditure information at the school-level; rather, this information is aggregated according to school district or other larger educational agency to which the school reports or belongs (Chambers 1999). This presents difficulties for unpacking site-specific differences in how schools use resources to implement a particular program or intervention as well as developing site-specific cost estimates that can be compared to other school-level outcome measures (Levin and Belfield 2013, 2015).

The Ingredients Method The “ingredients method” for evaluating education costs is a formalized process for identifying all the ingredients that collectively create a program and then systematically assigns costs to those ingredients (Clive Belfield 2018). The cost of the intervention is “defined as the value of all the resources that it [the intervention] utilizes had they been assigned to their most valuable alternative use” (Levin and McEwan 2001, p. 44). Enumerating ingredients accounts for the resources represented by program expenditures *and* other resources, such as donated personnel time, not captured by these sources (Levin and McEwan 2001). The ingredients method is also an effective tool for characterizing the package of resources used by a program to produce observed outcomes, contributing to an understanding of the details of how an intervention has been implemented (Levin and Belfield 2013). Unlike expenditure information from accounting systems, the ingredients method enables linkages between resources and various programmatic elements and initiatives that comprise a program (Chambers 1999; Levin and McEwan 2001; Rice 1997). Not only does this type of analysis calculate simple cost estimates, it also explains how resources are used and distributed (Rice 2001).

Enumerating the full complement of resources that contribute to producing an intervention's observed outcomes allows comparisons with program outputs (Table 14.2). In education, comparisons frequently occur between costs and program effects—or, cost effectiveness analysis (CEA). This approach compares alternatives that offer a given level of effectiveness

Table 14.2 Nested cost template

<i>AVID/TOPS program components</i>	<i>Activity</i>	<i>Resource category</i>	<i>Ingredients</i>	<i>Amount</i>	<i>Unit value</i>	<i>Period</i>	<i>Shared</i>	<i>Total</i>	<i>Reallocated</i>	<i>Marginal</i>
	[Col 2]	[Col 3]	[Col 4]	[Col 5]	[Col 6]	[Col 7]	[Col 8]	[Col 9]	[Col 10]	[Col 11]
<i>Program administration</i>	General administration	<i>Personnel</i>								
		<i>Supplies & materials</i>								
		<i>Contracted services</i>								
		<i>Other program inputs</i>								
	Scheduling									
	Student monitoring									
	Parent advisory board									
	Program meetings									
	AVID coordinator meetings (district)									

(continued)

Training & PO
 Coordinator-lead trainings at PLCs
Student
recruitment

Student recruitment & selection
 Parent nights
Subtotal

Facilities

Classrooms
Subtotal

for the lowest possible cost, or the highest level of effectiveness for a given cost (Levin and McEwan 2001). Cost-effectiveness estimates are calculated as a ratio of program costs to some standardized measure of effect. Necessarily comparative, cost-effectiveness ratios can be both applied to compare interventions on the same outcome and those that are similar in scale, as well as to benchmark policies against accepted efficiency metrics (Harris 2009).

While similar to cost effectiveness, cost benefit analysis (CBA) evaluates interventions according to their costs and benefits as measured in monetary terms as opposed to program effects (Levin and McEwan 2001). Unlike CEA, CBA relaxes the assumption that the interventions being compared have similar objectives. Rather, monetized benefits become the standardized objectives, supporting comparisons in investments, for example, between education and non-education policies. CBA also enables comparison within a single intervention—i.e., do the program’s benefits exceed its costs? A third type of cost analysis, cost utility (CU) analysis extends the benefit concept to evaluate costs in terms of “utility,” or the satisfaction derived by individuals as a result of one or more outcomes. CU relies on subjective information about individuals’ satisfaction with one or more measures of effectiveness (Levin and McEwan 2001). Analyzing costs relative to utility has been most frequently applied in health-related research (Rice 2002).

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY

Designing and Implementing Cost Studies Implementing the ingredients method involves three steps:

1. Developing detailed resource profiles that enumerate the full complement of resources used by a school to deliver programs and services;
2. Assigning a value to these resources, using market prices; and
3. Combining resource quantities with corresponding prices to calculate costs (Chambers 1999; Levin and McEwan 2001).

In practice, executing these steps involves an iterative process of collecting and analyzing data from administrative and financial data systems, institutional documents such as budgets and program descriptions, and first-hand information from stakeholders within the organization (Levin and

Belfield 2013; Levin and McEwan 2001). This information is used to construct detailed resource profiles that facilitate assigning monetary values to the identified resources. Programmatic costs reflect the sum of these resources.

Cost Templates The cost template approach serves as a tool for enumerating and valuing ingredients (see Table 14.2). Specifically, cost templates create the analytic frameworks that itemize ingredients used by educational programs, assign prices, and compute costs (Rice 1997, 2001). Within a common policy framework, templates become a tool that policymakers, practitioners, and evaluators can use to better understand how reforms differ across sites and how these differences affect program costs. A nested template structure further organizes ingredients according to multiple levels of categorization—e.g., according to program components (e.g., program administration, instruction), then by activities (e.g., teacher collaboration and planning), and properties (e.g., personnel, supplies and materials) (Kolbe and O’Reilly 2016). Nested templates are a particularly useful tool for developing cost estimates that support “apples-to-apples” comparisons among sites implementing a similar intervention or reform (Kolbe and O’Reilly 2016; Rice and Hall 2008).

Assigning Prices The cost template guides the process of valuing resources and constructing program cost estimates. A dollar value is assigned to each ingredient listed in the cost template by multiplying the unit price, for a given time period, by the number of units and the time over which the unit was used by the program. The market price for a given resource determines the unit prices. Prices can be either unstandardized or standardized. Unstandardized prices represent the actual, location-specific, price paid for a good or service. On the other hand, standardized prices reflect average prices across a geographic jurisdiction (e.g., state or nation). For example, teacher time can be valued using standardized, statewide average salaries, rather than unstandardized, local wages. In this case, standardized prices allow for statewide comparisons, independent of cost of living differences across localities and overrides district-specific variations in teacher salary schedules.

In cases where individuals or organizations donate, or provide at a reduced expense, personnel or other materials, resources are valued according to their use. For instance, where outside personnel serve in instructional roles similar to teachers, their time can be valued in terms of the statewide average for teacher compensation. Similarly, volunteers and consultants who serve as instructional assistants or provide enrichment opportunities could be valued at the state rate for teacher assistants and specialists. With regards to physical space, its cost stems from determining if the space is rented (in which case a market value exists) or owned. Alternatively, the cost of the space could be determined either through local realtor estimates or by measuring facility depreciation (Levin and McEwan 2001). In the latter instance, calculating costs depends upon the knowledge of the building's lifespan, the cost to replace the building, and the interest rate that represents the forgone costs of continual building investment (Levin and McEwan 2001).

Calculating Costs Once prices have been assigned to resources, the total program cost can be identified. Total program costs account for the full package of resources used by an intervention. First, the total cost for each ingredient is calculated. If an ingredient (e.g. teacher or administrator) contributes to the program in multiple ways, we prorate the cost in each instance and then aggregate across activities. In the case of a static ingredient—the classroom that is used in multiple ways—to avoid double counting the ingredient's value, the cost template prompts the use of a “shared” resource column to adjust for the ingredient's shared use (Rice 2001). The cost of each ingredient is calculated as follows:

$$\text{Ingredient cost} = \left[(\text{Amount Shared}) * \text{Unit Value} \right] / \text{Period}$$

Total costs are the sum of the costs of the individual ingredients. Total costs can be further standardized according to a cost per unit (e.g., per pupil cost).

Additionally, the researcher may be interested in calculating marginal costs. This focus on the additional or incremental change in resources associated with implementation occurs when stakeholders want to know what *additional* resources, over-and-above “business-as-usual,” were used

to implement a program. For example, a school-initiative may require five full time teachers, three of which the school already employs and whose time can be reassigned to the intervention. In this case, the additional resources needed to implement the program will be two full time teachers. The corresponding marginal costs is equivalent to the value of two full time teachers. Like total cost, marginal cost estimates also can be standardized as the marginal cost per pupil.

Imperfect data and different assumptions about the types, quantities, qualities, and value of inputs work together to introduce uncertainty in cost estimates (Levin and McEwan 2001). As a result, cost estimates should be tested against different assumptions. Rather than a single value, differences in values should be reported to establish a possible range for costs.

Applying Cost Analysis to Decision Making In this section, we discuss four potential applications for incorporating cost analysis into decision making: (1) characterizing resource requirements for implementing or replicating an educational program or intervention; (2) understanding the bottom line costs to determine its cost feasibility; (3) distributing the cost burden across stakeholders; and (4) as a tool for selecting among alternative policies and programs, by comparing costs with estimates of the programs effects and benefits.

We use findings from a cost analysis of a school district's efforts to implement the AVID/TOPS college readiness program in its four comprehensive high schools to illustrate each application. Put in place in Madison, Wisconsin in 2008 by the Madison Metropolitan School District (MMSD), AVID/TOPS combines the national AVID college-access program model with a local initiative, the Teens of Promise (TOPS) program. AVID is a national school reform model that prepares students for success in high school, college, and careers. Many AVID students are the first in their families to attend college and come from groups traditionally under-represented in higher education. Selected students are drawn from the "academic middle" and are viewed as capable of attending college, but at risk of falling short of their potential. The national AVID model requires students to enroll in a rigorous college-preparatory curriculum, as well as participate in specialized AVID programming. Operated by the Boys and Girls Club of Dane County (BGCDC), TOPS expands AVID to incorporate community-based mentoring, a summer internship experience, and

college transition support. The two programs are tightly integrated and operated through a long-standing partnership between MMSD and BGDCD.

Resource Requirements Education policies frequently establish broad guidelines for program implementation while also allowing for site-specific differences in program design to account for local context and needs. The result of this policymaking strategy, however, is that a stipulated policy or intervention can be implemented in very different ways across schools. In fact, decades of research examining education policy and program implementation points to the important role played by context and capacity in shaping local responses to education reforms (Honig 2006). More recently, particularly at the federal and state levels, education policies increasingly embed assumptions that allow for, and even encourage, local variation in program implementation. Identifying “ingredients” that go into producing observed effects sheds light on how different resource packages might explain differences in program impacts across sites, as well as highlight different strategies for deploying resources to attain stated goals.

Policymakers and educational leaders are also increasingly interested in replicating and scaling up effective educational programs. Frequently, however, they encounter a critical road block in doing so: insufficient information about the resources and corresponding costs associated with implementing a policy or program. While knowledge about which policies work and for whom is critical for decision makers, summative evidence of a policy’s or program’s effects falls short of providing that information. This is especially true when it comes to information about the package of tangible ingredients that were put in place to attain observed effects. Itemizing these ingredients is not only important for understanding what might be required for program implementation, they also form the basis for evaluating the feasibility—both in terms of available resources and potential costs—of undertaking or expanding a particular policy or program.

As an initial step in our efforts to evaluate the cost of implementing the AVID/TOPS program, we profiled the package of resources each school site used to implement its program. Table 14.3 presents the resource summary for implementing the AVID-only program components in schools, broken down by site. The ingredients method generated an inventory of

Table 14.3 AVID program delivery resource summary

		<i>East</i>	<i>North</i>	<i>South</i>	<i>West</i>
	<i>Units</i>	<i>Time</i>			
Personnel					
School administrator(s) (MMSD)	FTE	0.10	0.03	0.17	0.11
TOPS administrative assistant	FTE	0.50	0.50	0.50	0.50
AVID coordinators (<i>average effort per coordinator; 2 coordinators per school</i>) (MMSD)	FTE	0.53	0.39	0.44	0.38
TOPS coordinators (<i>average effort per coordinator; 2 coordinators per school</i>) (BGC)	FTE	0.78	0.68	0.78	0.73
AVID elective teacher (<i>average effort per course section</i>) (MMSD)	FTE	0.38	0.43	0.36	0.30
Tutor hours (BGC)	Hours	3262	2448	1562	1871
Guidance counselor(s) (MMSD)	FTE	0.61	0.12	0.03	0.07
Outside speakers	#	36	36	36	28
Substitute teacher days (MMSD)	Days	46	27	30	39
School site team (MMSD/BGC)	Hours / year	47	32	56	29

Source: Kolbe and Feldman (2016)

the resources used by a program and generated site-specific resource profiles that describe these resources. In this example, comparisons among sites illustrate the different staffing models (e.g., time spent on program administration, the extent to which guidance counselors participated in college and career coaching). Not only does the list of ingredients provide decision makers with an organized stock of information that guides them through considering resource needs, it also offers a framework for considering whether resources can be reallocated or if new resources might be necessary.

Cost Feasibility At the most basic level, decision makers need to understand whether they can afford to implement and operate a program. Estimating cost to determine feasibility is an essential consideration for program development, replication, and expansion. While the resource summary can be used to answer the question, “do we have the capacity to implement the program?” Cost feasibility asks, “can we afford to implement the program?” As a second step in our evaluation, we applied the cost template approach to generate bottom line estimates of AVID/TOPS program costs, by school site (see Table 14.4). We also disaggregated

Table 14.4 AVID/TOPS program costs

	<i>Cost (\$)</i>	<i>% total cost</i>
<i>General program oversight/operations</i>		
AVID		
Program administration	185,283	6.14
AVID-related training and professional development	493,094	16.34
Experiential learning opportunities	10,000	0.33
Tutor management	63,967	2.12
<i>Subtotal</i>	752,345	24.93
TOPS		
Program administration	131,446	4.36
TOPS mentoring	79,539	2.64
Summer experience	132,479	4.39
College continuation	147,578	4.89
<i>Subtotal</i>	491,042	16.27
<i>Program delivery at school sites</i>		
Program administration	460,275	15.25
AVID elective	933,294	30.93
AVID tutoring	287,129	9.52
Site-specific training and professional development	8172	0.27
Student recruitment	50,167	1.66
Facilities	35,145	1.16
<i>Subtotal</i>	1,774,182	58.80
Total	3,017,569	

Source: Kolbe and Feldman (2016)

resources and corresponding resource costs according to common program components. Nesting resources and costs in this way within a program sheds light on how different aspects of a program contribute to its costs. Additionally, in cases when a program is implemented across multiple sites, resource disaggregation according to a common structure can be used to identify differences in how resources are used as well as help to develop rationale for site-specific differences in program costs.

Distributing Cost Burden Distributional analysis supports efforts to go beyond simple cost estimates to also understand what resources were contributed by various stakeholder groups and how costs are distributed (Levin and McEwan 2001; Rice 1997). For example, while BCBG is solely responsible for implementing TOPS, MMSD, and BGCDC share

Table 14.5 Distribution of costs among stakeholder groups

	<i>Actual program cost</i>		<i>Budgeted expenses</i>		<i>Cost vs. budgeted expense (\$)</i>
	<i>Total (\$)</i>	<i>% total</i>	<i>Total (\$)</i>	<i>% Total</i>	
Total	3,017,569		2,202,338		815,231
<i>Cost distribution</i>					
MMSD	1,736,617	57.6	1,171,088	53.2	565,529
BGCDC	1,134,262	37.6	1,031,250	46.8	103,012
Unpaid MMSD instructional staff time	105,779	3.5			
Community members' donated time	40,911	1.4			

Source: Kolbe and Feldman (2016)

responsibility for Advancement Via Individual Determination (AVID). The organizations jointly staff the program's instructional and tutoring components. BGCDC organizes weekly experiential learning opportunities. MMSD oversees school site operations for AVID, including annual certification, student recruitment and selection, and AVID-related training. Both entities jointly fund AVID program activities. Table 14.6 shows the distribution of the cost burden by the two organizations. The cost burden can also be understood as the share of costs covered by the program's budget. In the case of the combined AVID/TOPS program, we found that that the amount of time teachers spent on program activities far exceeded the district's budgeted FTE allocation and that several schools also partnered with community volunteers to deliver supplemental tutoring and mentoring support for students (Kolbe and Feldman 2016). Altogether, about 5% of the program's costs were attributable to professional time that was not included in the program's budget (Table 14.5).

Selecting Among Alternative Policies Educational policymakers and leaders need to identify not only the most effective programs, but also those that achieve desired outcomes at the least possible cost. One approach combines measures of program effects and costs into cost-effectiveness ratios. Cost-effectiveness ratios can then be compared among programs with similar goals and objectives. For the AVID/TOPS program, we

compared multiple measures of program costs and effects to those available for the federally-funded Talent Search program (Hollands et al. 2014). Like AVID/TOPS, Talent Search shares common goals for improving rates of high school completion and transitions to college among at-risk youth by providing tutoring, financial awareness, and career selection training, college tours, and assistance with applying and enrolling in college. Unlike AVID/TOPS, however, the Talent Search program is typically less intensive in its interactions with students and places less emphasis on academic support and student course taking.

Table 14.6 presents the cost-effectiveness indicators for the two programs. Overall, we find that the AVID/TOPS program resulted in about 6.3% additional students graduating on time from high school than would have occurred without the program. By comparison, Talent Search saw a 10.8% gain in high school completion, although it is unclear whether students completed within four years or over a longer time period. These gains were quantified in terms of a cost per outcome (number of extra

Table 14.6 Comparison of AVID/TOPS and Talent Search program costs and effects

	# <i>participating students</i>	% <i>gain</i>	<i>Yield of extra grad/ enrollment</i>	<i>Cost per participant</i>	<i>Cost per outcome</i>	<i>Yield of extra graduates per \$100 k</i>
<i>AVID/TOPS: On time high school graduation</i>						
All students	225	6.3	14	\$16,990	\$271,492	0.37
White	47	1.6	1	\$16,990	\$1,061,875	0.09
Underrepresented minorities	178	5.0	9	\$16,990	\$339,800	0.29
Low income	151	8.0	12	\$16,990	\$212,412	0.47
First generation	91	10.5	10	\$16,990	\$162,362	0.62
<i>Talent Search: high school completion</i>						
Talent Search	3930	10.8	423	\$3290	\$30,520	3.3

Source: Kolbe and Feldman (2016)

graduates multiplied by the average cost per participant). Here, we see that the cost per outcome, for all students participating in the AVID/TOPS program, was \$271,492, whereas the cost per outcome for the Talent Search program was \$30,290.

On the one hand, taken together, these findings suggest that the AVID/TOPS program may be less “cost effective” than the Talent Search program. However, this example, also provides an opportunity to consider why educational leaders need to be careful when drawing such conclusions without further consideration. For instance, by breaking down the data by student subgroups, we can see that the AVID/TOPS program may be a more cost-effective intervention for some student subgroups than others. Given the already high graduation rates for White students in the district, the net gain in high school graduates for the AVID/TOPS program was just one White student. This stands in contrast to the gains for low-income and first-generation students. Similar data are unavailable for the Talent Search program (Bowden and Belfield 2015).

Another consideration are differences in what is known about the programs’ outcomes. Existing evaluations of the Talent Search program have largely evaluated the program’s effectiveness in terms of high school graduate rates. The AVID/TOPS evaluation also evaluated the program’s impact on whether students enrolled in a two- or four-year college immediately following high school graduation. When comparing college attendance rates with program costs, we found that the AVID/TOPS program was considerably more cost effective at producing college attendees than it was high school graduates (Kolbe and Feldman 2016). For instance, the cost per college attendee was about \$58,000 per White student, \$61,500 for first generation students, and \$70,000 for first-generation college student. These findings might prompt educational leaders to consider the program’s merits in light of alternative outcome measures, in addition to high school graduation.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

Cost analysis is a powerful tool for evaluating and analyzing education policies, programs, resources, and communicating cost information in ways that support decision making. Its usefulness as an evaluation and analytic technique, however, is predicated on how the researcher conceptualizes the research. We recommend that researchers begin by asking—and answering—three questions at the outset.

- **What is the education policy, program, or intervention that I am evaluating?**

To enumerate resources, the researcher must first be able to clearly articulate what constitutes the initiative under evaluation. At times, this articulation is simply a discrete intervention. In these instances, there exists a clear logic model linking program inputs, services delivered, and short-, medium- and long-term outcomes. In other cases, the reform under consideration is systemic or large in scope and scale, comprised of many discrete and interrelated activities. The researcher's task is to clearly articulate what comprises the service delivery system being evaluated, its component parts, and how these parts are linked to expected outcomes—recognizing that, at times, the lines delineating this system are porous with shared resources and jointly produced outcomes. In these instances, the researcher needs to define what constitutes the intervention's scope for the purposes of evaluating costs and, where questions remain, clarify the decision process and its potential impact on resultant cost estimates.

Researchers should also be prepared to “get familiar” with the details of how the program, intervention, or service works in practice. In many respects, this familiarization process resembles high quality implementation evaluation. To effectively identify and categorize the full complement of resources used in implementation, the researcher must truly understand how the program is being implemented. Oftentimes, this involves extensive interviews with program personnel, careful review of program documents, and triangulation across multiple data sources.

- **What are the questions I want to answer about this initiative's resources and costs?**

Cost studies should be designed to answer the questions posed by the researcher and stakeholders. Will the study estimate both total and marginal costs? How might cost estimates be combined with program effectiveness data? What information is needed about how resources are distributed among sites, participants, or funders? What types of sensitivity analyses should be built into the design to explore meaningful variation in estimates? Defining the questions ahead of time will help determine the necessary data sources and how the study's cost templates are structured.

- **Who is the audience for my work?**

Although a seemingly straightforward statement, at the outset, researchers need to be mindful about the desired contributions of their work and how the research can be best designed to meet stakeholders' information needs. Are stakeholders focused on whether a program can be implemented within existing budget constraints? Are they interested in understanding differences in program implementation across sites? Is it important to understand the relative resource contributions made by different groups, including those that might be donated or paid for by outside organizations?

CHAPTER SUMMARY

Resource-based cost estimation using the ingredients method applies an economic lens to estimating education program costs by identifying the package of resources used to implement a program, and subsequently assigning dollar values to these resources. Relying on the economic perspective of resource allocation, cost analysis seeks to identify and quantify all resources that contribute to program implementation. This bottom-up approach differs from the accounting model, which typically cannot capture or convey the recipe of resources necessary to implement or replicate an educational policy or program.

Implementing the ingredients method relies upon multiple sources of information, including interviews with main stakeholders, financial reports, and budget sheets. These data sources provide a detailed picture of a program's ingredients as implemented rather than as budgeted. Then, costs are applied to these ingredients. This creates a more detailed and comprehensive picture of both the resources and the costs necessary to implement the program than would otherwise be available. By valuing the resources necessary, and combining them with program effectiveness measures, researchers can expand the decision-making information educational leaders and policymakers have when choosing between multiple potentially beneficial interventions.

Recommended Readings

Levin, H. M., McEwan, P. J., Belfield, C. R., Bowden, A. B., & Shand, R. D. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis* (3rd ed.). Thousand Oaks: Sage Publications.

Levin, McEwan and colleagues outline the rationale for carrying out a cost analysis study. While the book does not provide step-by-step guidance, it details the different types of cost analysis and the rationale behind each one and explains the criteria a researcher must take into account in order to undertake a cost study. A helpful overview and reference guide.

Belfield, C. R. (2015). Cost-benefit analysis and cost-effectiveness analysis. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy* (2nd ed., pp. 141–156). New York: Routledge.

Provides a conceptual and methodological overview to cost analysis and its application to evaluating educational policies and programs.

Hollands, F. M., Hanisch-Cerda, B., Levin, H. M., Belfield, C. R., Menon, A., Shand, R., & Cheng, H. (2015). *CostOut: The CBCSE cost tool kit*. Retrieved from www.cbccosttoolkit.org

The CostOut Tool provides national lists of prices for school, district, and university personnel, resources, technology, and school facilities.

Kolbe, T., & O'Reilly, F. (2016). The cost of increasing in-school time: Evidence from the Massachusetts Expanded Learning Time initiative. *Leadership and Policy in Schools*, 1–39.

Rice, J., & Hall, L. J. (2008). National Board certification for teachers: What does it cost and how does it compare? *Education Finance and Policy*, 3(3), 339–373.

These articles provide useful examples of a fully enacted cost studies, using the ingredients method and cost template approach. Both illustrate the methods' usefulness in supporting policy decision making.

REFERENCES

- Belfield, C. R. (2015). Cost-benefit analysis and cost-effectiveness analysis. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy* (2nd ed., pp. 141–156). New York: Routledge.
- Bowden, A. B., & Belfield, C. (2015). Evaluating the Talent Search TRIO program: A benefit-cost analysis and cost-effectiveness analysis. *Journal of Benefit-Cost Analysis*, 6(03), 572–602. <https://doi.org/10.1017/bca.2015.48>.
- Chambers, J. G. (1999). *Measuring resources in education: From Accounting to the resource cost model approach*. Retrieved from Washington, DC: <https://nces.ed.gov/pubs99/199916.pdf>

- Clive Belfield & A. Brooks Bowden & Henry M. Levin, (2018). *Cost estimation in education: The ingredients method*. Chapters, in: Teaching Benefit-Cost Analysis, chapter 16, pages 200–207 Edward Elgar Publishing.
- Friedman, L. S. (2002). *The microeconomics of public policy analysis*. Princeton: Princeton University Press.
- Harris, D. N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3–29. <https://doi.org/10.3102/0162373708327524>.
- Hartman, W. T., Bolton, D., & Monk, D. (2001). *A synthesis of two approaches to school-level financial data: The accounting and resource cost model approaches*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2001/2001378_2.pdf
- Hollands, F., Bowden, A. B., Belfield, C. R., Levin, H., Cheng, H., Shand, R., et al. (2014). Cost-effectiveness in practice: Interventions to improve high school completion. *Educational Evaluation and Policy Analysis*, 36(3), 307–326.
- Honig, M. (2006). *New directions in policy implementation: Confronting complexity*. Albany: SUNY Press.
- Ingle, W. K., & Cramer, T. (2013). Not just accountability: A cost-effectiveness analysis of third-grade reading diagnostic tools. In B. G. Barnett, A. R. Shoho, & A. Bowers (Eds.), *School and district leadership in an era of accountability* (pp. 143–172). Charlotte: Information Age Publishing Inc.
- Kolbe, T., & Feldman, R. (2016). *Evaluation of the AVID/TOPS program's resources & costs*. Retrieved from University of Wisconsin-Madison, School of Education, New Hope Lab website: <http://wihopelab.com/publications/AVID-TOPS-Cost-Resource-Analysis.pdf>
- Kolbe, T., & O'Reilly, F. (2016). Expanded learning time in public schools: A cost effectiveness analysis of Massachusetts' Expanded Learning Time Initiative. *Leadership & Policy in Schools*. <https://doi.org/10.1080/15700763.2016.1232832>.
- Levin, H., & Belfield, C. R. (2013). *Guiding the development and use of cost-effectiveness analysis in education* Retrieved from Center for Benefit-Cost Studies: http://cbcse.org/wordpress/wp-content/uploads/2014/11/Guidance_on_cost_effectiveness_analysis.pdf
- Levin, H., & Belfield, C. R. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, 8(3), 400–418.
- Levin, H., & McEwan, P. (2001). *Cost-effectiveness analysis: Methods and applications* (2nd ed.). Thousand Oaks: Sage Publications.

- Levin, H., & McEwan, P. (2002). Cost-effectiveness and educational policy. In H. Levin & P. McEwan (Eds.), *Cost-effectiveness and educational policy* (pp. 1–17). Larchmont: Eye on Education.
- O'Reilly, F., & Kolbe, T. (2011). *Where does the money go? Expenditures for the Massachusetts Expanded Learning Time (MA ELT) Initiative*. Retrieved from <https://www.doe.mass.edu/research/reports/2011/12ELT-Expenditures.pdf>
- Rice, J. (1997). Cost analysis in education: Paradox and possibility. *Educational Evaluation and Policy Analysis*, 19(4), 309–317.
- Rice, J. (2001). The cost of working together: A framework for estimating costs of comprehensive support systems for children. *Administration & Society*, 33(4), 455–479.
- Rice, J. (2002). Cost analysis in education policy research: A comparative analysis across fields of public policy. In H. Levin & P. McEwan (Eds.), *Cost-effectiveness and educational policy* (pp. 21–35). Larchmont: Eye on Education.
- Rice, J., & Hall, L. J. (2008). National Board certification for teachers: What does it cost and how does it compare? *Education Finance and Policy*, 3(3), 339–373. <https://doi.org/10.1162/edfp.2008.3.3.339>.

PART III

Mixed, Applied, and Collaborative Approaches to the Study of Educational Leadership and Policy

Part III of this volume presents mixed, applied, and collaborative approaches to the study of educational leadership and policy. Specifically, this part of the volume addresses mixed methods research, program evaluation, and the design of international research collaborations.



Using Mixed Methods to Inform Education Policy Research

*Colleen E. Chesnut, John H. Hitchcock,
and Anthony J. Onwuegbuzie*

The purpose of this chapter is to describe the philosophical grounds and practical applications of using mixed methods in education policy research. Through review of relevant scholarship on education policy and research design, we will establish not only the importance of attention to rigorous research in the field of education policy, but also the need for robust mixed methods design options. We define mixed methods research as that which involves the researcher(s) combining elements of both qualitative and quantitative approaches (e.g., methodologies, paradigms, methods, analyses) for

C. E. Chesnut (✉)
Center for Evaluation & Education Policy, Indiana University,
Bloomington, IN, USA
e-mail: cechesnu@indiana.edu

J. H. Hitchcock
Abt Associates, Inc., Rockville, MD, USA
e-mail: jhhitch@indiana.edu

A. J. Onwuegbuzie
College of Education, Sam Houston State University, Huntsville, TX, USA
University of Johannesburg, Johannesburg, South Africa

a particular study (Johnson et al. 2007). This does not imply a predetermined sequence of applying both qualitative and quantitative data collection and/or analysis methods but rather a context-specific utilization of multiple approaches to research, interwoven into the inquiry process so as to address a policy or research problem and to build a cohesive narrative either to describe or to begin to solve it. The broad definition of mixed methods research applied here allows for a variety of approaches to the process, recognizing the complexity of education policy problems. In particular, mixed methods research is suited to address *wicked problems*, which refer to “problems involving multiple interacting systems, replete with social and institutional uncertainties, for which there is no certainty about their nature and solutions, and for which time is running out to find solutions” (Mertens et al. 2016, p. 225). As noted by Mertens et al. (2016), “Additional concepts related to researching wicked problems include the need for researchers to address power inequities, violations of human rights and impediments to social justice, and strategizing for action in the form of policies and behaviors” (p. 225). Thus, the goal of proposing a mixed methods perspective to education policy research is not to proscribe a singular technique, but to present examples of and options for mixed methods research design that will inform the work of education leaders, policy analysts, scholars, and other stakeholders as they decide how to tackle the complex policy questions in their particular contexts.

This chapter proceeds with a review of relevant scholarship in two areas: (1) policy research, analysis, and evaluation; and (2) theoretical frameworks for mixed methods research. Following these reviews of the literature, recent examples are provided to illustrate applications of mixed methods to the study of education policy. This chapter concludes with recommendations for novice and emerging scholars of education policy to apply mixed methods research to their own works. General guidance and an overview of a mixed methods concept called *legitimation* (akin to validity) are provided in terms of making choices on methodological approaches throughout the inquiry process, with a hypothetical example woven throughout to illustrate the various options.

REVIEW OF RELEVANT POLICY ANALYSIS AND MIXED METHODS PARADIGM LITERATURE

One of the stated purposes of this chapter is to provide guidance on applying mixed methods research to education policy scholars and practitioners. To lay the foundation for this guidance, it is necessary first to define policy and policy research and then to outline a theoretical framework for using mixed methods research in the education policy context.

Policy and policy research are neither consistently nor clearly defined among scholars; as Guba (1984) pointed out, policy scholars often assume that their readers will understand the term *policy* and, thus, do not flesh it out. Whereas Guba (1984) addressed this need by defining policy in several different ways (e.g., goals or intents, sanctioned behavior, problem-solving strategy) and outlining a variety of approaches to policy analysis, Kerr (1976) broadly defined policy as some person or agency planning to do something specific under certain conditions, in order to achieve some purpose. Moving from this expansive definition of policy, we now consider how to conceptualize policy research, policy evaluation, and policy analysis, operating from the assumption that these concepts depend on contextual factors. To clarify, the purpose of *policy research* is to address a policy problem, and the outcomes may be either enhanced knowledge about the problem or action toward addressing it. The purpose of *policy evaluation* is to determine the value of a policy, and outcomes may be either program improvement or assessment of program impacts. Finally, the purpose of *policy analysis* is to provide evidence toward informing policy choices and results in enhanced knowledge among stakeholders who make those policy choices (Lincoln and Guba 1986). As Sabatier (1988) suggested, policy actors may use the results of policy analysis to advocate for certain policy alternatives, and more robust analyses may result in greater “policy-oriented learning” among stakeholders. To these definitions of policy analysis, Nagel (1990) added an emphasis on “systematic analytic methods which can be quantitative or qualitative” (p. 276). While Nagel underscored the importance of rigorous methodology in policy analysis, DeLeon (1994) asserted that interdisciplinary and democratic approaches to policy inquiry should prevail in order to suit the complexity of policy problems and contexts. In summary, scholars of policy science, which broadly encompasses policy research, evaluation, and analysis, have called attention to both the purposes of policy inquiry (e.g., to inform policy decisions) and the processes of policy inquiry (e.g., systematic, interdisciplinary methods).

We highlight the scholarship on the varied purposes of policy inquiry to emphasize that mixed methods, in which applications of research processes are driven by the goals of the investigation, may often be the most appropriate methodological approach to thoroughly examine policy problems. For this chapter, we rely on Kerr’s (1976) broad description of policy and follow Lincoln and Guba’s (1986) classifications of policy research, evaluation, and analysis. Additionally, we embrace DeLeon’s (1994)

reframing of policy inquiry toward a paradigm with a greater multidisciplinary focus and emphasis on democratic discourse and participation. Finally, because the focus here is on utilizing mixed methods in policy inquiry, we agree with Nagel's (1990) emphasis on the need for "systemic analytic methods", but underscore that these may be *both* quantitative *and* qualitative, rather than either/or, and if both classes of methodologies may be used, then they can be mixed so as potentially to conceptualize and to address policy problems in a way that can be difficult to achieve using a singular approach.

As the scholarship distinguishing among policy analysis, research, and evaluation suggests, the type of inquiry conducted depends upon the purpose of the study, the intended outcomes, and the target audience. For this reason, mixed methods approaches to policy inquiry are quite useful, due to the variety of data that can be collected and multiple options for the application of analytic techniques. In defining and applying mixed methods research, education leadership and policy scholars have employed several different philosophical frameworks. For example, Mertens (2010) relied on a transformative paradigm for mixed methods research on the effectiveness of a teacher preparation program serving the deaf community. The transformative paradigm is grounded in the assumption that research should seek to enhance social justice and uphold human rights while engaging diverse stakeholders and respecting cultural norms. She cited several more examples of research across a variety of disciplines, asserting that the transformative paradigm is applicable for all types of social inquiry that seek to engage stakeholders in alleviating social injustices (Mertens 2010).

Other mixed methods researchers have advocated for pragmatism as a driving philosophy for mixed methods approaches, a framework that suggests that methodological decisions should be driven by research objectives rather than by ideological preferences (Johnson and Onwuegbuzie 2004). More recently, scholars have called for mixed methods inquiry to be grounded in dialectical pluralism. Specifically, dialectical pluralism is a *metaparadigm* that enables application of different and even competing theoretical perspectives within and throughout the research process (Johnson 2017). Working within the paradigm of dialectical pluralism allows for mixed methods research that provides (approximately) equal status to both qualitative and quantitative methodologies and invites greater dialogue between researchers and various stakeholders (e.g., education leaders, policy analysts, scholars), as well as continual reflexivity in

the research process. The research purpose and focus should inform both methodological and paradigmatic choices throughout all phases of inquiry. Furthermore, it has been suggested that framing mixed methods research designs through dialectical pluralism can result in research findings that have the desired impact in their fields, due to the importance placed on stakeholder perspectives and participation (Johnson 2017). Moving forward, we see dialectical pluralism as an appropriate framework for conceptualizing mixed methods research in education because education policy research encompasses a wide range of contexts, stakeholders, and content areas. Thinking about mixed methods research in this way allows for researchers and practitioners to select not only the methods that will best address their research problems, but also the theoretical paradigms that reflect their own and their research participants' contexts. With that focus in mind, readers are encouraged to explore alternatives such as Mertens's transformative paradigm; indeed, Morgan (2007) presents a good introductory source of broader paradigmatic discussion.

APPLICATIONS TO THE STUDY OF EDUCATION POLICY

Proceeding from the definitions and broad theoretical underpinnings of mixed methods research, policy, and policy inquiry, we move now to outlining the practical applications of mixed methods research to the study of education policy. While we recognize that this edited volume focuses broadly on research methods in educational leadership and policy studies, this chapter emphasizes the use of mixed methods in education policy research. This is not meant to exclude education leadership as an important field of research. On the contrary, we conceptualize education policy research as inclusive of many policy issues applicable to the realm of education leadership and the work of both researchers and practitioners in the field. The following section provides examples from education researchers regarding the use of mixed methods in their fields, including education leadership, as well as a review of recent mixed methods research in education.

Recent work by education policy scholars illustrates that mixed methods research approaches have gained a foothold in policy realms. Asserting that the complex nature of education research warrants the flexibility offered by mixed methods research designs, Creswell and Garrett (2008) suggested not only that education researchers have become increasingly interested in using mixed methods research approaches but also that they should be credited with helping to shape the growth of the field of mixed

methods research. They predicted that education is ripe for mixed methods approaches to inquiry, due to an “openness to experiment with research methodologies and ways of thinking about research” (Creswell and Garrett 2008, p. 329). Similarly, policy scholars have argued that mixed methods approaches can be used effectively to explore policy questions, particularly as related to complex social issues (Brannen and Moss 2012). The benefit of mixed methods research designs is that researchers are encouraged to consider multiple perspectives, which can facilitate productive research around policies that affect a variety of stakeholders. Furthermore, as the amount and scope of quantitative data surrounding policy initiatives increase, there might be an enhanced need for “qualitative and mixed methods research that can better interrogate the limits and possibilities of the policy questions these data sets are used to explore” (Brannen and Moss 2012, p. 792). Thus, policy research that incorporates mixed methods approaches likely will better inform both stakeholders and policymakers in search of answers to complicated policy questions.

Numerous peer-reviewed publications demonstrate applications of mixed methods research in education policy, showing a variety of mixed methods designs for exploring complex questions. Table 15.1 depicts the studies cited in the following section, with corresponding research foci and brief descriptions of their mixed methods designs.

Because education research questions often represent problems that are multifaceted, mixing methods can be an effective way to examine an issue from several angles, particularly if there are multiple policies targeting a certain outcome. Consider the goal of ensuring that “highly qualified” teachers teach all students and the range of policies aimed at this goal, from requirements for advanced coursework to teacher evaluation based on student achievement. Employing an *across-stage* mixed methods research design, wherein answers to research questions at one stage lead to another set of questions requiring different methods (see Johnson and Onwuegbuzie 2004), Rutledge et al. (2010) explored how policies around teacher quality and test-based accountability impacted principals’ hiring decisions. They aligned their selection of methods across the stages of their study with the complexity of the problem, and the qualitative data collection and analysis led to the testing of hypotheses via quantitative methods to probe the findings further (Rutledge et al. 2010). Among the implications of their research, Rutledge et al. (2010) highlighted the value of using mixed methods to gain a comprehensive understanding of the policy context; using mixed methods helped to mitigate some of the limitations inherent in employing a singular methodology.

Table 15.1 Applications of mixed methodology in education policy research

<i>Citation</i>	<i>Research focus</i>	<i>Mixed methods design</i>
Rutledge, Harris, and Ingle (2010)	Impacts of policies on teacher quality and test-based accountability on principals' hiring decisions	Across-stage <ol style="list-style-type: none"> 1. Qualitative: interviews with principals and district administrators; observations of hiring fairs 2. Quantitative: numerical coding of interview responses; correlational and multivariate analyses
Eckert (2013)	How teacher qualifications predict new teachers' confidence and retention in high-poverty urban schools	Sequential-explanatory <ol style="list-style-type: none"> 1. Quantitative: analysis of National Center for Education Statistics (NCES) schools and staffing survey data, using multinomial ordinary least squares (OLS) and logistic regression, multilevel modeling 2. Qualitative: interviews with selected teachers in high-poverty urban schools
Hall and Ryan (2011)	A middle school's internal accountability system and its perceived collective capacity with regard to external accountability mandates	Qualitatively-driven case study with sequential design <ol style="list-style-type: none"> 1. Qualitative: document reviews, interviews, participant observation 2. Quantitative: questionnaires for faculty
Luft et al. (2011)	Induction programs for beginning secondary science teachers	Embedded-concurrent <ol style="list-style-type: none"> 1. Quantitative and Qualitative data collection: Interviews and observations with structured protocols 2. Quantitative analysis of scores on interview and observation protocols 3. Qualitative coding of interview data
Sharp et al. (2012)	Design of a multisite case study of a statewide education policy mandate	Pragmatic parallel <ol style="list-style-type: none"> 1. Quantitative: student outcome data and student questionnaires 2. Qualitative: content analysis of course catalogs, interviews, and focus groups

In another study focused on teacher quality, Eckert (2013) examined how certain measures of teacher qualifications predicted new teachers' confidence and rates of retention in high-poverty urban schools. Distinct from Rutledge et al.'s (2010) study, Eckert's (2013) work involved the use of a sequential-explanatory mixed methods research design (see Ivankova

et al. 2006), comprising an initial quantitative phase followed by a qualitative phase. In the quantitative phase, data were used to identify relationships among several key variables (e.g., amount of teacher preparation, teacher efficacy, and teacher retention), and in the qualitative phase, these relationships were probed further, using interview protocols developed from the quantitative findings. To illustrate, in the quantitative phase, Eckert (2013) found a positive and significant relationship between teachers' personal efficacy and their levels of preparation, which led to interview questions delving deeper into this finding (e.g., what elements of teacher preparation impact feelings of teacher efficacy?). The qualitative phase provided nuance to the teacher efficacy variable, which would not have occurred via the quantitative analysis alone (Eckert 2013). For both Rutledge et al. (2010) and Eckert (2013), the mixed methods research designs involved clear qualitative and quantitative phases of research, using one to inform the other, albeit in different ways.

Additional examples from education policy research demonstrate how scholars have used sequential mixed methods research designs to explore policy problems. For instance, in a study focused on educational accountability, Hall and Ryan (2011) argued that federal education policy initiatives (e.g., No Child Left Behind) have encouraged an "overreliance" (p. 105) on quantitative data and methods. This is presented as a concern because, in part, a qualitatively-driven mixed methods approach could be particularly valuable for investigating the challenging issue of accountability in a middle school setting; this type of approach allowed for focus on contextual complexities and engagement with multiple stakeholders and perspectives (Hall and Ryan 2011). Utilizing a case study with a sequential design, the researchers collected and analyzed both qualitative and quantitative data through phases of research, using findings from one phase to inform development of the next. Specifically, a preliminary qualitative phase led to quantitative data collection and analysis, followed by a qualitative case study. By mixing methods in sequence and clearly delineating between phases of their research, Hall and Ryan (2011) demonstrated how using mixed methods to explore a complex education policy issue can lead to deeper understanding of the problem from various perspectives.

Education policy scholars have also utilized mixed methods research in ways that are distinct from the previously discussed sequential designs. For instance, in a study of new science teachers participating in induction programs, Luft et al. (2011) used an embedded-concurrent mixed methods research design, in which quantitative and qualitative data were collected

and analyzed throughout the course of the study period. Qualitative and quantitative findings ultimately were integrated in addressing all research questions. In another study, Sharp et al. (2012) exemplified how mixed methods may be used early in the research process to select cases for a multisite case study. Relying on a pragmatic approach in designing a longitudinal study examining the impacts of a statewide college and career readiness initiative, researchers utilized several stages of quantitative and qualitative sampling strategies to select case study sites. Akin to sequential mixed methods research designs, in which qualitative or quantitative findings precede and then inform the next stage of research, Sharp et al. (2012) employed quantitative sampling strategies followed by qualitative sampling strategies to select the cases to enhance generalizability of the research findings.

Education policy scholars employ mixed methods in a variety of ways and for a variety of purposes. In justifying the application of mixed methods approaches, researchers often refer to the complexity of the education contexts under scrutiny. For example, states, districts, and schools implement different policies with the same objective, and contextual factors influence the application of a policy in distinct ways across settings. Some researchers opt for sequential mixed methods designs to account for such nuance, following qualitative phases of inquiry with quantitative phases, or vice versa, and some opt for mixing methods in only one phase or throughout all phases. Review of the recent research applying mixed methods to education research reveals that mixed methods design options are as varied as the scholars and their research questions. In addition to fitting under the umbrella of mixed methods research, what these examples share in common is a pragmatic approach to tackling complicated policy problems via methodology driven by the research contexts and questions rather than simply preference for a certain type of data or analysis.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

For education policy scholars interested in conducting mixed methods research, the following section provides guidance on choosing among different approaches, illustrating how to conceptualize a mixed methods design via an example of examining a current policy issue. As exemplified by studies cited previously and affirmed in scholarship defining mixed methods research (e.g., Johnson et al. 2007), there are numerous options for mixing methods in a research design. Qualitative and/or quantitative methods may be applied in discrete phases, or both types of methods may

be integrated throughout all stages of inquiry. Some scholars describe their mixed methods research as either qualitatively-driven (Hesse-Biber et al. 2015) or quantitatively-driven (Mark 2015), depending on the methods that predominate in their designs, whereas others adopt equal-status mixed methods designs, with emphasis on full-integration of methodologies (Johnson 2017). No matter the methods chosen or the stages of research in which they are used, scholars who choose mixed methods research designs frequently do so because they have determined that employing mixed methods research approaches will allow for the most rigorous examination of their research questions. Among the qualities outlined for *strong* mixed methods research, Tashakkori and Creswell (2007) specified that the study should “demonstrate the need for mixed methods to answer research questions that include clearly interconnected qualitative and quantitative components” (p. 207). Other scholars (e.g., Johnson and Onwuegbuzie 2004; Onwuegbuzie and Leech 2006; Plano Clark and Badiee 2010) have made a similar assertion: that the research questions should drive the selection of methodologies, and that there are several options for framing research questions in mixed methods studies to distinguish when and how different methodologies will be used. As identified by Plano Clark and Badiee (2010), research questions for mixed methods inquiry may include the following:

- *separate quantitative and qualitative research questions;*
- *general overarching mixed methods research questions;*
- *hybrid mixed methods issue research questions;*
- *mixed methods procedural/mixing research questions;*
- *combination research questions;*
- *independent research questions;*
- *dependent research questions;*
- *predetermined research questions;* and
- *emergent research questions.*

There are numerous options for planning, structuring, and conducting mixed methods research, and scholars should be intentional in grounding their choices throughout each phase of the inquiry process in the original research problem and purpose. The following section outlines how the application of a mixed methods research design could unfold via an example of an education policy problem.

An Example Consider the following education policy question: should schools/districts provide teachers with bonuses if their students demonstrate above-average progress? Akin to most education policy problems, this is a complex issue involving a variety of stakeholders. In this section, we demonstrate a hypothetical mixed methods approach to examining this question, to provide scholars with an illustration of the process. However, we remind readers that the series of methodological choices outlined here will not necessarily apply fully to their areas of research interest or expertise. Rather, we offer this example to display the process of making methodological choices, identifying the key factors to consider in the development of a mixed methods research design.

In studying whether schools should pay bonuses to teachers for student progress, the first step in deciding upon a mixed methods research design is to unpack the question. As Tashakkori and Creswell (2007) recommended, researchers should consider how to frame their questions for mixed methods research. There are aspects of this issue open to interpretation, such as who may benefit from the findings and what is meant by “above-average progress”. One scholar might be interested in whether bonuses increase teachers’ rates of retention, whereas another scholar might be interested in whether schools providing bonuses see greater student growth across all classrooms. The question is not simply whether bonuses are good; one must consider the stakeholders involved, the intended audience(s) for the research, and what information will enhance their understandings of the issue. How might bonuses be beneficial, to what extent, and for whom? If we dissect the issue into questions that invite both qualitative and quantitative methodologies, we come up with an overarching question: To what extent does the policy of providing teachers with bonuses if their students demonstrate above-average academic progress in a school year increase teachers’ job satisfaction? This could be followed by sub-questions that are more specifically quantitative and qualitative:

Quantitative

- How do rates of retention for teachers who have received bonuses compare to those for teachers who have not received bonuses?
- What are the characteristics of teachers who have received a bonus?

Qualitative

- How do teachers who received bonuses perceive their levels of job-related stress before and after receiving a bonus?
- How do teachers perceive the bonuses, particularly in terms of their value and their impacts on levels of motivation?

Although the overarching research question identifies teachers' job satisfaction as the subject of inquiry, the sub-questions provide further specificity to the concept by implying that job satisfaction could be measured via retention rates and via teachers' self-reported levels of job-related stress at two points in time. Furthermore, there can be value in understanding how teachers perceive the bonus policy. This, in turn, yields a reason for mixing; if, for example, the bonus does not appear to impact teacher retention rates, researchers might understand why, assuming that the qualitative findings indicate that teachers do not particularly value the bonus. Furthermore, the qualitative findings might show that teachers are already working hard, and there is little reason to believe that adding a bonus structure will systematically impact their ability to yield more effort.

After crafting research questions that indicate a potential need for a mixed methods approach, the next choices for researchers are to determine the types of data needed and appropriate methods for data collection. As Wolf (2010) noted, these choices regarding methodologies and data should be built upon "a solid theoretical foundation" (p. 147); researchers engaged in mixed methods inquiry should articulate both how and why they choose to apply certain methods. To return to the research problem and questions outlined previously, perhaps the researcher takes a philosophical perspective of dialectical pluralism, due to the diversity of stakeholders and the goal of honoring their perspectives. For example, the administrators of the school district under scrutiny may take a postpositivist approach—that is, they believe that statistics on teacher retention will demonstrate the success or failure of the bonus policy. However, a school board member may approach the issue pragmatically, recognizing the variety of schools in the district and the need to explore the perspectives of numerous teachers working within them. Of course, there will be many more stakeholders and possibilities for diverse viewpoints; considering the research via a lens of dialectical pluralism might allow the researcher to elucidate and to respect these in the inquiry process.

Assuming that the hypothetical research study is grounded in dialectical pluralism, inviting paradigmatic differences into the process, the researcher could proceed with choosing appropriate methodologies for exploring the research questions. The first two questions on teacher retention indicate a quantitative (i.e., postpositivist; Phillips and Burbules 2000) approach, presuming that data are accessible regarding which teachers received bonuses and which were retained from one year to the next. Additional descriptive data on variables like teachers' years of experience, gender, and education levels would provide further areas for analysis. For the third question on teachers' perspectives regarding stress as related to the incentive policy, a qualitative approach might suffice: interviewing teachers regarding their job-related stressors could reveal how/whether they perceive that the policy contributes to their levels of stress. Additionally, the qualitative data might illuminate trends or irregularities from the quantitative findings; perhaps the teachers tell interviewers that they plan to seek jobs elsewhere in spite of earning bonuses because they do not believe that the extra money adequately compensates them for the extra stress. Taken together, the findings resulting from collecting and analyzing both the quantitative and qualitative data would inform the overarching question on teachers' job satisfaction as related to the incentive policy and provide some additional nuance to the issue. Do retention rates or self-reported stress levels more accurately represent teachers' job satisfaction? For this mixed methods research study, this need not be an either-or question. Considering quantitative data and findings alongside the qualitative data and findings paints a more complex picture of the issue; the policy may be "successful" from the quantitative perspective, in that teachers who earn the incentives are retained, but if some of the same teachers also report a strong desire to transfer to less stressful work environments without the incentive policy, is the policy's success sustainable?

Considering questions such as these opens another door to understanding mixed methods research, namely, the door of *legitimation*, which is parallel to the concept of validity. There are at least nine types of legitimation in mixed methods research studies (or ways of thinking about mixed methods validity; cf. Onwuegbuzie and Hitchcock 2017; Onwuegbuzie and Johnson 2006), such as inside-outside, weakness minimization, conversion, and political (because of space limitations, not all are listed; interested readers can reference source material). Inside-outside legitimation deals with the degree to which the researcher accurately presents both the research participants' and researchers' views. The example of teacher

perceptions of bonus structures could get into nuanced views about the nature of the teaching profession, and findings may interact with the researcher's own beliefs. To understand this interaction requires careful presentation of the views held by different actors in the research process to present legitimate findings (i.e., legitimation). Weakness minimization deals with the extent to which the strength of one approach (e.g., quantitative) is compensated for by the other (e.g., qualitative) approach. Applied to the earlier example, it can be difficult to use qualitative methods to predict circumstances under which teachers with different characteristics are likely to receive a bonus, especially when utilizing large datasets. Dealing with this issue fits well with regression modeling. At the same time, exploring complex and dynamic thought processes of teachers regarding their motivations when/whether a bonus is offered is an issue well-suited for interviewing, focus groups, and archival analyses. The balance of these different modes of inquiry demonstrates how the weakness of one method is compensated for by the other and vice versa. Conversion legitimation becomes possible when one quantifies qualitative data (e.g., counting themes, ordinal modeling of themes) and/or qualitzes quantitative data (e.g., interviewing stakeholders about their responses to Likert-format survey items). The issue of legitimation, or mixed methods validity, arises when one explores whether conversion promotes high-quality meta-inferences, which are inferences supported by both quantitative and qualitative information. In the teacher bonus example, one could, for instance, identify which types of teachers are prone to agree with specific themes around motivation, and the validity of a design will become apparent if this effort yields findings that are difficult to achieve via a monomethod (i.e., quantitative only OR qualitative only) approach. Political legitimation might be especially relevant in this example and policy research in general because it addresses the degree to which a researcher accounts for different stakeholder values and interests. One way to think about the legitimacy of research through this lens is whether the inquiry shapes policy in a way that meets the needs of different groups, especially stakeholders with minimal power.

Of course, there could be many more ways to outline a mixed methods approach to this policy issue, depending on the context, stakeholders, and desired outcomes. By fleshing out some options for research design with the previous example, we hope to have provided general guidance for scholars and practitioners new to mixed methods research. Walking through a hypothetical research planning process is meant to remind

education policy researchers of the many methodological and analytical choices that they must consider and provide some direction for thinking through those choices.

CHAPTER SUMMARY

By moving from broad definitions of policy, policy inquiry, and mixed methods research to the paradigmatic and analytical choices involved in conducting mixed methods research, this chapter provides guidance to education policy scholars interested in mixed methods designs. Recent peer-reviewed mixed methods studies on complex education policy problems, such as teacher quality and school accountability, illustrate how researchers have used mixed methods designs to interrogate these issues, while attending closely to various perspectives and contexts. Ultimately, by outlining how to conceptualize a mixed methods inquiry for a hypothetical, yet realistic, education policy problem, we present an example of the methodological decision-making processes essential to effective mixed methods research design. The information and guidance we offer serves as a useful starting point for novices considering whether and how mixed methods research can inform their work to alleviate the complicated policy problems facing today's education leaders.

Recommended Readings

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks: SAGE.

Creswell and Plano Clark provide a comprehensive and practical text on mixed methods research, covering all stages, from design through writing and evaluation of findings. They offer guidance on each phase of the process, and this updated edition includes recent examples of published mixed methods studies from a variety of disciplines, including education leadership.

Ivankova, N. V. (2015). *Mixed methods applications in action research: From methods to community action*. Thousand Oaks: SAGE.

Ivankova provides practical guidance for researchers interested in applying mixed methodology to action research, starting with foundational descriptions of both mixed methods and action research. The bulk of the text focuses on the specific steps of conducting mixed methods action research, from design and planning through data collection, analysis, and dissemination of findings. This text may be informative to both students and practitioners interested in applying mixed methodology to action research in a variety of fields.

Tashakkori, A. M., & Teddlie, C. B. (Eds.). (2010). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). Thousand Oaks: SAGE.

Tashakkori and Teddlie's edited volume presents perspectives from numerous scholars in mixed methods research. The text provides guidance from these scholars on conceptual issues of mixed research, issues related to research design and analysis, and practical application of mixed method research. This text may be useful to novice researchers seeking a variety of perspectives to inform their decisions regarding mixed methods designs.

REFERENCES

- Brannen, J., & Moss, G. (2012). Critical issues in designing mixed methods policy research. *American Behavioral Scientist*, 56, 789–801. <https://doi.org/10.1177/0002764211433796>.
- Creswell, J. W., & Garrett, A. L. (2008). The “movement” of mixed methods research and the role of educators. *South African Journal of Education*, 28, 321–333 Retrieved from http://www.scielo.org.za/scielo.php?pid=S0256-1002008000300003&script=sci_arttext&tlng=pt.
- DeLeon, P. (1994). Reinventing the policy sciences: Three steps back to the future. *Policy Sciences*, 27(1), 77–95. <https://doi.org/10.1007/BF00999600>.
- Eckert, S. A. (2013). What do teaching qualifications mean in urban schools? A mixed-methods study of teacher preparation and qualification. *Journal of Teacher Education*, 64(1), 75–89. <https://doi.org/10.1177/0022487112460279>.
- Guba, E. G. (1984). The effect of definitions of policy on the nature and outcomes of policy analysis. *Educational Leadership*, 42(2), 63–70.
- Hall, J. N., & Ryan, K. E. (2011). Educational accountability: A qualitatively-driven mixed-methods approach. *Qualitative Inquiry*, 17(1), 105–115. <https://doi.org/10.1177/1077800410389761>.
- Hesse-Biber, S., Rodriguez, D., & Frost, N. A. (2015). A qualitatively-driven approach to multimethod and mixed methods research. In S. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford handbook of multimethod and mixed methods research inquiry* (pp. 3–20). New York: Oxford University Press.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20. <https://doi.org/10.1177/1525822X05282260>.
- Johnson, R. B. (2017). Dialectical pluralism: A metaparadigm whose time has come. *Journal of Mixed Methods Research*, 11, 156–173. <https://doi.org/10.1177/1558689815607692>.

- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112–133. <https://doi.org/10.1177/1558689806298224>.
- Kerr, D. (1976). *Educational policy: Analysis, structure, and justification*. New York: David McKay.
- Lincoln, Y. S., & Guba, E. E. (1986). Research, evaluation and policy analysis: Heuristics for disciplined inquiry. *Review of Policy Research*, 5, 546–565. <https://doi.org/10.1111/j.1541-1338.1986.tb00429.x>.
- Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed-methods study. *Journal of Research in Science Teaching*, 48, 1199–1224. <https://doi.org/10.1002/tea.20444>.
- Mark, M. M. (2015). Mixed and multimethods in predominantly quantitative studies, especially experiments and quasi-experiments. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *The Oxford handbook of multimethod and mixed methods research inquiry* (pp. 21–41). New York: Oxford University Press.
- Mertens, D. M. (2010). Transformative mixed methods research. *Qualitative Inquiry*, 16, 469–474. <https://doi.org/10.1177/1077800410364612>.
- Mertens, D. M., Bazeley, P., Bowleg, L., Fielding, N., Maxwell, J., Molina-Azorin, J. F., & Niglas, K. (2016). Expanding thinking through a kaleidoscopic look into the future: Implications of the mixed methods international research association's task force report on the future of mixed methods. *Journal of Mixed Methods Research*, 10, 221–227. <https://doi.org/10.1177/1558689816649719>.
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1, 48–76. <https://doi.org/10.1177/2345678906292462>.
- Nagel, S. S. (1990). Bridging theory and practice in policy/program evaluation. *Evaluation and Program Planning*, 13, 275–283. [https://doi.org/10.1016/0149-7189\(90\)90058-5](https://doi.org/10.1016/0149-7189(90)90058-5).
- Onwuegbuzie, A. J., & Hitchcock, J. H. (2017). A meta-framework for conducting mixed-methods impact evaluations: Implications for altering practice and the teaching of evaluation. *Studies in Educational Evaluation*, 53, 55–68. <https://doi.org/10.1016/j.stueduc.2017.02.001>.
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools*, 13(1), 48–63.
- Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures. *The Qualitative Report*, 11, 474–498. Retrieved from <http://www.nova.edu/ssss/QR/QR113/Onwuegbuzie.pdf>

- Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham: Rowman and Littlefield Publishers.
- Plano Clark, V. L., & Badice, M. (2010). Research questions in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (2nd ed., pp. 275–304). Thousand Oaks: Sage.
- Rutledge, S. A., Harris, D. N., & Ingle, W. K. (2010). How principals “bridge and buffer” the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring. *American Journal of Education*, 116, 211–242. <https://doi.org/10.1086/649492>.
- Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy-oriented learning therein. *Policy Sciences*, 21(2–3), 129–168. <https://doi.org/10.1007/BF00136406>.
- Sharp, J. L., Mobley, C., Hammond, C., Withington, C., Drew, S., Stringfield, S., & Stipanovic, N. (2012). A mixed methods sampling methodology for a multisite case study. *Journal of Mixed Methods Research*, 6(1), 34–54. <https://doi.org/10.1177/1558689811417133>.
- Tashakkori, A., & Creswell, J. W. (2007). Exploring the nature of research questions in mixed methods research. *Journal of Mixed Methods Research*, 1, 207–211. <https://doi.org/10.1177/1558689807302814>.
- Wolf, F. (2010). Enlightened eclecticism or hazardous hopscotch? Mixed methods and triangulation strategies in comparative public policy research. *Journal of Mixed Methods Research*, 4, 144–167. <https://doi.org/10.1177/1558689810364987>.



Program Evaluation Methods for PK-12 Education

Liz Hollingworth

What is *evaluation*? Weiss (1998) defines evaluation as “the *systematic assessment* of the *operation* and/or the *outcomes* of a program or policy, compared to a set of *explicit* or *implicit standards*, as a means of contributing to the *improvement* of the program or policy” (p. 4). Within the context of their schools, school leaders are tasked with many kinds of evaluation. Program evaluation is specifically focused on programs, products, policies, and not the evaluation of personnel, such as teacher performance. *Program evaluation* is a way to assess program effectiveness, to improve the implementation of a program, to analyze the costs and benefits of a program, and to document program accomplishments.

Importantly, program evaluation has applications in many contexts, including the health sciences, the arts, business, disaster preparedness, the nonprofit sector, the military, and social work. In this chapter, the focus is on program evaluation in the service of improving the programs and policies in K-12 schools. Nevo (2006) conceptualizes educational evaluation as

L. Hollingworth (✉)
Educational Policy and Leadership Studies, The University of Iowa,
Iowa City, IA, USA
e-mail: liz-hollingworth@uiowa.edu

a complex concept serving five major functions: decision-making, improvement, accountability, professionalism, and certification. School leaders are uniquely positioned to use evaluation methods to understand student learning, to evaluate teacher performance, to make decisions about instructional materials, and to understand how programs are working in schools.

COMPARING EVALUATION AND RESEARCH

Imagine the research methods described thus far in this book are specialized tools to be used to conduct educational research. Evaluators use many of these tools to consider the effectiveness of programs. Evaluators often use surveys, focus groups, interviews, document analysis, large-scale data sets, observations, hierarchical linear modeling, and all of the other data collection techniques researchers use. While the tools themselves are the same, there are four distinct ways evaluation and research are different: (1) purpose, (2) audience, (3) use of results, and (4) standards that guide the work. To illustrate this, Table 16.1 provides a set of questions that differentiate the two.

First, evaluation and research serve a different purpose. For research, the purpose is to understand a phenomenon or theory and then generalize the results to as many domains as possible. Levin-Rozalis (2003) explains that, in contrast, “evaluation is intended to amass knowledge and understanding of a concrete activity—a project—and to give this understanding back to the project as feedback” (p. 6). In fact, in evaluation, the findings

Table 16.1 The differences between research and evaluation

	<i>Research</i>	<i>Evaluation</i>
1. Purpose: why is the project happening?	Researcher-driven, based on the needs of the field	Client-driven, based on the needs of the program
2. Audience: who wants the results?	The research community	Client and stakeholders
3. Results: how are the findings used?	To increase scientific knowledge; to apply the findings as widely as possible	For program improvement; to understand what is happening with a specific project during a given period of time
4. Standards: what ethical principles guide the process?	Institutional Review Board	Program evaluation standards and guiding principles

are typically only relevant to the project that is evaluated, and even then, findings can be applied only to the time of the evaluation.

Second, the audience for evaluation differs from the audience in a traditional research study (Berk and Rossi 1999). Typically, research questions are developed by the same team of people who will conduct the study and interpret the results. In contrast, with input from the evaluator and program stakeholders, a client generates the questions for the evaluation. The evaluator carries out the evaluation, including all data collection and analysis. The findings are then interpreted collectively by the client, stakeholders, and the evaluation team. The utility of the evaluation itself hinges on the relevance of the evaluation questions to the client and the program's stakeholders.

Third, research and evaluation differ with respect to the way that evaluation results are used. In research, results typically appear in a peer-reviewed journal and are used to contribute to the broader literature about an issue in a given field. Conversely, in evaluation, the results are used to inform the program or policy under review. Results are typically aimed at key program decision-makers. For example, formal reports might be written to provide funding agencies with feedback on the progress of an initiative. These reports typically are free from jargon and are focused specifically on the evaluation questions developed collaboratively between the evaluator and client. This is not to say that program evaluation reports cannot result in publications. In fact, the American Evaluation Association publishes several journals devoted to the advancement of the field of evaluation (see, e.g., *New Directions for Evaluation* and the *American Journal of Evaluation*).

Fourth, evaluation research has an established set of standards that guide an evaluator's practice. The program evaluation standards created by the Joint Committee on Standards for Educational Evaluation are guidelines that require reflective practice and judgment (Yarbrough et al. 2010). In other words, leaders need evidence and argument to show an evaluation has attended to the standards. There are five overarching evaluation standards: **Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability**.

In addition to the Program Evaluation Standards, the American Evaluation Association created Guiding Principles for Evaluators (Miller et al. 2005), which also impact practice. These include:

- (a) **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

- (b) **Competence:** Evaluators provide competent performance to stakeholders.
- (c) **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process.
- (d) **Respect for People:** Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
- (e) **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

The adherence to these program evaluation standards and guiding principles are another important way evaluation differs from research, which is typically guided by ethical rules set forth by Institutional Review Boards. Of these differences, the most important difference between program evaluation and research relates to the way the questions are developed.

IMPORTANCE OF QUESTION FRAMING

Program evaluation rests on the development of a coherent **theory of action**. Alkin (2011) defines a theory of action as a way of asking “what do those who run the program believe will result from the various actions (or activities) conducted?” (p. 71). In short, the theory of action describes how and why your program will work and reveals the strategic thinking behind the change expected because of the program. Typically, if-then statements are used to identify assumptions regarding how a program is expected to achieve the stated goals.

Relatedly, a **logic model** is a visual representation of the inputs, activities, outputs, and outcomes of the program to allow school leaders an opportunity to make decisions about how to change the program so that it will work better. Most logic models include resources (inputs), activities (outputs), and goals (outcomes). A logic model is most effective when an explicit theory of action is embedded in the program components (Knowlton and Phillips 2012). Logic models play a role in question framing by helping researchers to articulate a comprehensive plan for understanding the program, thus avoiding utilizing a random string of data collection ideas and activities.

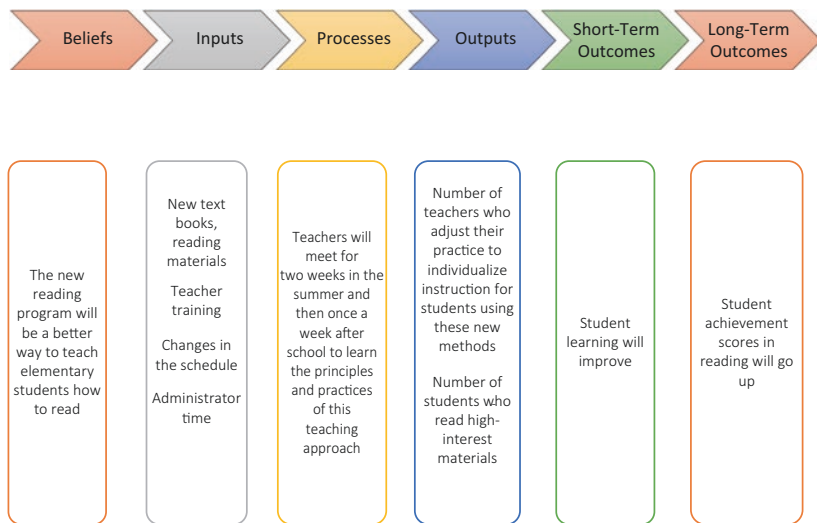


Fig. 16.1 Sample logic model for a new reading program

Consider the following example: a school district wants to try a new reading program with elementary school students. The program will require a significant financial investment in new curriculum, professional development for teachers, and additional time for reading instruction. Figure 16.1 provides one example of a logic model illustrating the theory of action for this reading program.

Creating a figure like this to describe the intended outcomes and outputs of a program can help to clarify the value of the intervention for stakeholders, like parents, teachers, and students. Inputs should include human resources like teacher time and administrative staff, financial support, supplies, in-kind donations, and any other assets that are necessary for the program to be successful.

Evaluation methods are the ideal approach for school leadership teams wanting to understand how a given initiative is working, whether all of the people responsible for the enactment of the program or policy are consistently and faithfully following the process, and if the initiative is working as intended. There are multiple types of program evaluations, and selecting the appropriate one depends on the questions the administration team has about the program itself. The next section defines a few of these using examples from schools.

TYPES OF PROGRAM EVALUATION

Internal Versus External Evaluation When a program administrator or someone from within the same school district conducts the evaluation, it is considered an *internal evaluation*. Internal evaluators have the advantage of understanding the context of the program and knowing the historical and political forces at work in the school district. What is more, internal evaluators know how the members of the team have struggled to make the program work, and will be sensitive to the personalities involved.

Sometimes it makes sense for an outside person to be invited into a school to conduct an evaluation to get a fresh perspective and to give teachers, students, and parents the chance to speak openly and anonymously about their concerns. This kind of *external evaluation* allows for an unbiased, objective analysis of program implementation and effectiveness. An external evaluator typically will not have a stake in the program and thus is thought to bring an unbiased perspective to the analysis. External evaluations are often conducted by professional evaluators. Professional evaluators have significant expertise and thus might be better positioned to identify strengths, weaknesses, and potential flaws in the logic of a program that internal evaluators might not be able to see.

Funding for a program evaluation often comes from specific line items in grant applications or as a gift from a philanthropic foundation. Many granting agencies will allow a percentage (i.e., 8–10%) of the total grant budget to be allocated for program evaluation services. This is a cost-effective solution for many schools, as it frees up administrators from doing evaluation work over and above their regular duties. In addition, a trained evaluator can assist with the writing of the evaluation portion of a grant proposal, to ensure that the plan will provide the school with the information needed to make the best use of scarce resources.

Informal Versus Formal Evaluation It is important to note that evaluation can be conducted formally and informally. *Informal program evaluations* tend to rely on general observations that are not systematically documented. Informal evaluations often rely upon impromptu, unsystematic procedures for observation. In contrast, *formal program evaluations* rely on more structured research designs. Evaluators thus approach a formal evaluation systematically with a particular aim being to support data-based decision-making or long-term program improvement. Formal evaluations

are thus intensive research activities designed to account for and control a variety of extraneous variables that could produce, influence, or skew evaluation results.

Formative Versus Summative Evaluation Evaluation can be both formative and summative in its orientation, with either approach using a variety of methodologies to meet the needs of project stakeholders. Scriven (1996) suggests viewing the formative/summative as a dichotomy to explain how the evaluation results will be used. *Formative evaluation* (McClintock 1984) is often conducted in the early stages of a program to provide stakeholders with feedback about improving the program. Formative evaluation is often used with a pilot program and can be used as a form of progress monitoring to ensure the program is meeting its stated objectives. Conversely, in *summative evaluations*, programs are evaluated at the end of a cycle, and findings are used to determine whether a program should be continued, ended, or modified for future use. A summative evaluation is particularly important for reporting requirements for funding agencies, as in the case of a granting agency wanting to know if the program is effective (Wholey 1996).

To illustrate, consider a high school that is in the process of implementing a new computer science curriculum designed to teach all students basic computer programming. In the first year of the program, the school implements a new curriculum. As such, an instructional leader in the building or another stakeholder might want to offer classroom teachers feedback on the implementation of the new curriculum and/or obtain insights about how the implementation process could be carried out with greater fidelity. As such, an administrator might conduct a formative evaluation guided by the question: How is the new curriculum working for students?

To answer this question, an internal evaluator, such as the principal, could survey the students in the school and/or conduct focus groups with classroom teachers. The survey and/or focus groups might pose questions about the new curriculum materials. These questions might inquire about their ease of use or the teaching methods themselves. Likewise, feedback from student surveys might be used to inform the selection of specific teaching approaches or call for adjustments related to the teacher's instructional practice. Alternatively, the evaluator might conduct focus groups with selected teachers to identify how they are using the curriculum, what challenges they are encountering, and what support(s) they need. Taken

together, this information could be used to make changes to the school's implementation of the new curriculum as part of a larger improvement effort.

Beyond asking formative questions, evaluators also ask summative questions which aim to address the effect or impact of programs. For instance, an evaluator might ask: Has the new program improved student learning (as measured by student reading achievement)? The temptation with a question like this is to find pre- and post- scores on a test of academic achievement to use as a barometer for gains in student learning. The challenge with this approach is that schools seldom initiate only one improvement at a time, making it difficult to know the true cause of any change in student performance on a given academic measure. Moreover, because districts rarely implement school improvement initiatives in only some schools, leaving other schools to serve as controls, leaders must be cautious in interpreting any gains as the result of the program alone. This challenge is one all educational researchers face and is not unique to program evaluation.

An experienced evaluator recognizes that simply subtracting gain scores from two points in time is not the best way to assess student learning. A more appropriate way to approach this question might be to use a hierarchical linear model to control for confounding variables like socioeconomic status by including a variable to represent school-level fixed effects, like Free and Reduced Lunch (FRL) populations. Then, Cohen's *d* effect sizes can be used to estimate the relative impact of each independent variable on the dependent outcome. A summative evaluation need not be framed in purely quantitative terms. For instance, an evaluator could conduct a summative student survey about the strengths and weaknesses of the new curriculum (i.e., student learning experiences) that would allow for a different perspective on the program's success. Regardless of the approach taken, the program evaluation standards on feasibility should drive the decisions on how to answer an evaluation question about student learning.

Implementation Fidelity Evaluation Sometimes, the reason a program is not working as intended is that it is not implemented correctly. *Implementation fidelity* is the degree to which a program's procedures are delivered as intended. Implementation evaluation has the potential to provide rapid feedback about operations and outcomes and can be thought of as part of a management cycle of "plan, implement, evaluate improve" (Love 2004, p. 63).

The first step in an implementation evaluation is to collect and synthesize the research base supporting the program; be sure the evaluator understands the theoretical basis for the program's efficacy. This research review is critical to understanding the mechanisms by which the program is expected to work. For example, consider a new district-wide program to provide math tutoring to elementary students after school. An evaluation team might notice that there is one school where the program is not working as well as it is everywhere else. An analysis of implementation fidelity would allow for an understanding of whether the after-school program is consistently conducted competently and according to the protocols determined by the district. First, the evaluation team will need to understand the research base that supports the math tutoring program itself. A good place to start is the What Works Clearinghouse, sponsored by the Institute of Education Sciences, which provides reviews of existing research on different programs, products, practices, and policies in education. By fully understanding why the math tutoring program is supposed to work, the team can then begin to document each factor or component that is hypothesized to influence outcomes. This can then create a framework that draws on relevant intellectual theory from established academic disciplines or on accumulated knowledge from past studies of similar programs or policies (Nightingale and Rossman 2004). An implementation fidelity evaluation might ask a question such as: Are teachers implementing the new program as it was designed or intended? The evaluator would then focus on verifying that the program is, in fact, being implemented as designed by the creator of the program.

Implementation fidelity evaluations require multiple observations of teacher practice and an in-depth analysis of documents like lesson plans, worksheets, student work, and other educational materials created in the service of the program. The person who conducts the observations will need expertise in the new teaching methodologies, which means the school leader conducting the study will be expected to go through the professional development training along with the teachers. What is more, the observer will need to be clear that the observation is in no way tied to a personnel evaluation, in order to build trust with the teacher under observation. Teachers who are taking a risk and trying something new in their classrooms will not be as eager to volunteer for future school improvement initiatives if the observation conducted by the evaluator becomes part of their human resources record. In fact, program evaluation standards on propriety provide guidance for how to navigate a situation like this and make clear that evaluation data should never be used to evaluate individual employees.

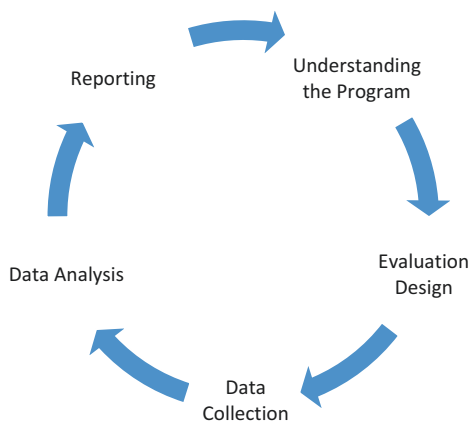
An evaluator could also frame an Implementation Fidelity Evaluation in more specific, practically related terms. For example, an evaluator could ask: Are the professional development activities changing teacher practice? To answer this question, an evaluator would need to document what classroom practice looked like both *before* and *after* the teacher went through the professional development, most likely through observations. Answering this question would require at least one entire school year of observations: several observations at the beginning of the year (before the professional development intervention) and then several at the end of the year (after the professional development is completed). In addition, interviews with teachers would allow the evaluator to understand the observed phenomena and to triangulate the data. The program evaluation standards on utility would guide the evaluation team to be sure the program stakeholders find evaluation processes and products valuable in meeting their needs.

Impact Evaluation *Impact evaluation* allows evaluators to assess how the program or intervention affects outcomes in both intended and unintended ways using a cause and effect analysis. An impact evaluation can also provide a framework to understand whether it is the program providing the benefit to its participants or whether changes might be attributed to some other factor (Khandker et al. 2009). For school leaders who are interested in whether or not to keep an existing program, an impact evaluation can be used to serve an accountability purpose to determine if and how well the program worked. Impact evaluation is often used with pilot programs and is particularly useful when a pilot intervention is going to be scaled up or tried in a different school setting. An example of an impact evaluation question might be: How would outcomes have changed had the program not been introduced to the participants? Such a question would allow the evaluator to determine the extent to which the outcome can be attributed to the program or intervention. An evaluator might follow-up with summative evaluation to assess the program's impact or a formative evaluation to provide targeted feedback to the program to sharpen its impact on the desired outcome.

FIVE PHASES OF PROGRAM EVALUATION

As illustrated in Fig. 16.2, the program evaluation cycle can be considered in five phases: understanding the program, evaluation design, data collection, data analysis, and reporting.

Fig. 16.2 Five phases of program evaluation



In Phase 1, an evaluator should review the program goals and objectives to ensure a common understanding of the program itself. Evaluators must work with stakeholders to develop evaluation questions and to come to a mutual understanding of a theory of action that will guide the selection of data collection methods. Generally, the development of a visual logic model is the best way to clarify the intended goals and the theory of action for the program itself. This is also the phase where the type of evaluation will be determined, including whether to hire an external evaluator. In addition, knowledge of the theories and scholarly literature that support the program will be vital to the success of Phase 1. Understanding the specific mechanisms that will be observed is the foundation of quality evaluation design.

Phase 2 is designing the evaluation. In this phase, the evaluation team will work together to determine the scope of the evaluation, the cost, and the plan for data collection, including selecting participants. The critical part of this phase is the decision about how to use the data that will be collected and analyzed. Russ-Eft and Preskill (Russ-Eft and Preskill 2009) note that without this, the evaluation will fail to meet its full potential. Comparison group designs and case studies can allow the evaluator to see how the program is functioning across multiple contexts.

Phase 3 of the evaluation cycle involves data collection. Data collection might involve interviews, surveys, observations, documents, and retrieval of archival data. In addition, the evaluator will have to determine who is going to be included in the evaluation: a sample of participants or everyone.

Patton (2005) suggests using a purposive sampling strategy to gather rich data from a sample. This is a particularly useful strategy in program evaluation, where it is often expensive to try to interview every person who may be impacted by a particular program (Patton 2014).

In Phase 4 the data are analyzed according to the evaluation questions. Qualitative data that have been collected will be analyzed according to the themes and evaluation questions that were determined by the evaluation team and stakeholders in Phase 1. Quantitative data will be similarly interpreted, and pooling data across multiple sites is preferred if the difference among the sites is not extreme (Rog 2010).

In Phase 5, the results are reported. One of the Program Evaluation standards is Utility, and Patton has written extensively about the importance of ensuring stakeholders are able to use the results of the evaluation by engaging them in every phase (see for example, Patton 2008). Evaluation reporting is different from research reports, as the tone is not usually scholarly and the results are focused tightly on the program itself, without an attempt at generalization or toward advancing a specific theory. In this iterative model, the evaluation cycle repeats as the program evaluation reports are used to make changes to the program itself.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

It is important that novice scholars embrace the practical aspect of program evaluation as an approach to gathering data. Evaluation can be useful in supporting both policy development and school-based decision-making. By bringing together groups of teachers, leaders, parents, community members, and other stakeholders to understand the goals of a given school improvement initiative, for example, evaluators are able to harness enthusiasm for a program while simultaneously helping the stakeholders understand how the program is supposed to work and could work better. Weiss (1998) suggests the following question to surface a program's theory of change: "What ideas and assumptions link the program's inputs to the attainment of the desired ends?" (p. 55). Once this question has been answered, evaluation questions can be developed to probe the program itself.

What follows are three recommendations for scholars interested in learning more about program evaluation. First, attend the annual meeting of the American Evaluation Association. This academic conference usually occurs in late fall and consists of a group of scholars who are interested in sharing best practices in program evaluation across multiple

contexts. The conference also includes training sessions around specific methodologies, so even seasoned evaluators find something new to learn each year. Second, volunteer with an organization with expertise in program evaluation so you can learn by doing. Local nonprofits will often have programs that would benefit from evaluation support. Such support might include conducting phone interviews, focus groups, or offering support for data analysis. In addition, many universities have evaluation centers that provide program evaluation services in the fields of public health, medicine, or education. Some have funding to support graduate student assistants who are interested in learning more about how to conduct high quality program evaluations. Third, there are excellent journals devoted to the study and use of program evaluation: e.g., the two mentioned earlier in this chapter, *New Directions for Evaluation* and the *American Journal of Evaluation (AJE)*. These journals provide the most up-to-date research in this area and can be accessed online and in hard copy. Novice and emerging scholars can benefit from reading past issues to see how program evaluation is used across multiple contexts. These journals are also ideal outlets for scholars who have conducted an evaluation and are interested in contributing to the scholarship around evaluation methods.

CHAPTER SUMMARY

This chapter provided an introduction to program evaluation, beginning with a discussion of the differences between evaluation and research. In short, the purposes of evaluation and the audience for the results are important differences between research and evaluation. What is more, the design of a program evaluation study is dependent upon the questions stakeholders have. The chapter also provided the program evaluation standards and guiding principles that drive the work of evaluators, another key difference between evaluation and research.

Question framing is of paramount importance in program evaluation. A logic model is a visual representation of the program's theory of action, which includes inputs, activities, outputs, and outcomes of the program. As a follow-up exercise for this chapter, the logic model provided in Fig. 16.1 can be used as a template to outline the features of another program either under consideration or already in the implementation phase. Understanding the theory of action for a program can then lead to the development of evaluation questions and a study design. Program evaluation can be considered as

a cyclical process, where results from an evaluation study are primarily used to inform the program itself, whether it is for future improvement or to determine the viability of the program itself.

In general, programs are evaluated to gain information and to make decisions. The type of evaluation is determined by the kinds of questions being asked. Types of program evaluation that are often used in K-12 schools include formative evaluation, summative evaluation, implementation fidelity evaluation, and impact evaluation.

Recommended Readings

Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. Jossey-Bass.

For readers interested in learning more about logic models, Frechtling's handbook is an accessible and easy to follow explanation of logic modeling and its importance in connecting theory with implementation and outcomes in program evaluation in the social sciences. In addition to detailing the components of the logic model in multiple contexts, it also explores common mistakes, making it an excellent resource for novice evaluators.

Stake, R. (2003). *Standards-Based and Responsive Evaluation*. Sage Publications.

This text is an invaluable resource that provides an overview of the conceptual decisions evaluators make by setting up the evaluation design process as a series of choices. Stake details how to attend to stakeholders, how to attend to the ethical challenges of program evaluation, and how to write an evaluation report.

REFERENCES

- Alkin, M. C. (2011). *Evaluation essentials: From A to Z*. New York: Guilford Press.
- Berk, R. A., & Rossi, P. H. (1999). *Thinking about program evaluation* (2nd ed.). Thousand Oaks: Sage.
- Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on impact evaluation: Quantitative methods and practices*. Washington, DC: World Bank Publications.
- Knowlton, L. W., & Phillips, C. C. (2012). *The logic model guidebook: Better strategies for great results*. Thousand Oaks: Sage.
- Levin-Rozalis, M. (2003). Evaluation and research: Differences and similarities. *The Canadian Journal of Program Evaluation*, 18(2), 1.

- Love, A. (2004). Implementation evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed., pp. 63–97). San Francisco: Jossey-Bass.
- McClintock, C. (1984). Toward a theory of formative program evaluation. *New Directions for Adult and Continuing Education*, 1984(24), 77–95.
- Miller, R. L., Chelimsky, E., House, E. R., Smits, P. A., Champagne, F., Preskill, H., ... Mason, M. (2005). American evaluation association: Guiding principles for evaluators. *American Journal of Evaluation*, 26, 106–123.
- Nevo, D. (2006). Evaluation in education. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *Handbook of evaluation: Policies, programs and practices*. Thousand Oaks: Sage.
- Nightingale, D. S., & Rossman, S. B. (2004). Collecting data in the field. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed., pp. 321–346). San Francisco: Jossey-Bass.
- Patton, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks: Sage.
- Patton, M. Q. (2014). *Qualitative research & evaluation methods* (4th ed.). Thousand Oaks: Sage.
- Rog, D. J. (2010). Designing, managing, and analyzing multisite evaluations. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed., pp. 208–236). San Francisco: Jossey-Bass.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change* (2nd ed.). New York: Basic Books.
- Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17(2), 151–161.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River: Prentice Hall.
- Wholey, J. S. (1996). Formative and summative evaluation: Related issues in performance measurement. *Evaluation Practice*, 17(2), 145–149.
- Yarbrough, D. B., Shulha, L. M., Caruthers, F. A., & Hopson, R. K. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks: Sage.



International Network as Sites for Research on Successful School Leadership

Christopher Day and David Gurr

The experience of research on school leadership for most readers of this chapter is likely to have been small scale, designed and conducted alone or with one or two others, as full or part-time doctoral students, or as university tutors with limited resource support. In education particularly, many will have engaged in single paradigm research: qualitative studies or those based on survey design and analysis. Relatively few will have used mixed methods or had opportunities to work as part of a collaborative team of researchers over two or more years. Even fewer will have worked with colleagues from different countries and cultures. Yet, the press of academia, and the increasingly inter-connected policy and practice world in which we live, demands that international collaborations increase. Therefore, it is important to discuss how understandings between researchers—who are

C. Day (✉)

School of Education, The University of Nottingham, Nottingham, UK
e-mail: Christopher.day@nottingham.ac.uk

D. Gurr

Graduate School of Education, The University of Melbourne,
Carlton, VIC, Australia
e-mail: d.gurr@unimelb.edu.au

© The Author(s) 2018

C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,

https://doi.org/10.1007/978-3-319-93539-3_17

meeting and managing different challenges to traditional definitions of intellectual rigor and successful processes of collaboration, in research worlds which are increasingly interacting with policy-led excursions into territory which was once regarded as belonging to academia—may be achieved. There are as yet few empirically informed reports on these collaborations which address the processes of engagement in longitudinal research of scholars from different jurisdictions that are at different points in their development of educational systems and who work in policy environments that are culturally diverse.

This chapter provides an exploration of these issues through the lens of one longstanding multi-national research network, the International Successful School Principalship Project (ISSPP). It describes how this network grew from research on successful school principalship in eight countries into more than twenty over a decade, with three strands of research. Although the chapter will discuss methodological and other technical issues, the focus is not only on these. A focus on principles, processes, and practices of researcher engagement in multi-national research is important for those who seek to ‘cross borders’ in working with others over time. This is because sustained engagement in network research is only one of several parts of the work of university professors, and so there are likely to be fluctuations in individual commitment, resource availability, and occasional discontinuities in active participation. Discussions on researcher engagement are under-represented in the literature, yet are likely to be relevant to the work of all cross-country research collaborations.

THE ISSPP

The origins of the ISSPP and its use of largely qualitative methodologies can be found in an earlier multi-perspective study of schools in England (Day et al. 2000). The primary aim of that research had been to find out what caused schools to be effective under the leadership of the same principal over time. The research involved multiple perspective case studies of primary, secondary, and special schools in which head teachers were widely acknowledged as being effective over time and which were judged to be successful in terms of student learning outcomes. Head teachers, deputy head teachers, governors, parents, students, classroom and school support staff, and teachers were interviewed. The study found that effective head teachers were constantly managing competing tensions and dilemmas, and that they demonstrated leadership approaches that were

values- and visions-based. These leaders acted with integrity, adapted to multiple contexts, promoted continuous professional learning among staff, and used self-reflection to improve their own practice.

Largely as a result of this work, the ISSPP was established in 2001 by school leadership academics from universities in eight countries. At the time, research about successful principals had rarely been the focus of research on school leadership internationally. The founding group agreed that rather than use the term ‘effective’ leadership, a more accurate term would be ‘successful’ leadership. The latter was distinguished from the former because it did not only focus on school outcomes as a measurable metric (e.g. tests/examinations), but also the broader educational teaching and learning vision and practices in which students were given opportunities to participate.

The network now has members from 24 countries. Further information can be found on the project website (www.uv.uio.no/ils/english/research/projects/isspp). The members are a mix of less and more experienced researchers who have joined at different times in the network’s history. The only ‘entry’ requirements are that they agree to use the ISSPP agreed research protocols, produce at least one case study in the first two years of membership, acknowledge this in any subsequent publications, and attend at least one ISSPP meeting each year. New members always begin with research on strand one. The use of agreed research methods and protocols has meant that the ISSPP has built an extensive international archive of case studies of school principalship in successful primary and secondary schools in different cultures and socio-economic contexts. Systematic analysis of multi-perspective case study data within and across the participating ISSPP countries has provided a wealth of rich qualitative data internationally which point to a number of key qualities, skills, and behaviors central to successful principalship in all contexts; a selection of the findings are reported below.

There are currently three research strands of ISSPP research:

- *Strand 1.* This explores the work of successful principals through a focus on three areas: principals in primary and secondary schools; those working in areas of high social and economic disadvantage; and, those principals who sustain success.
- *Strand 2.* To expand the research on successful school leadership, this strand considers the work of principals in visible and invisible under-performing schools.

- *Strand 3*. The most recent strand explores the professional identities of principals leading successful schools.

Each of these employs different methods. For the purposes of this chapter, only the first strand is discussed.

REVIEW OF RELEVANT LITERATURE: RESEARCH CONTEXTS

Here we consider ideas about networks that provide insights into how the project has operated over more than 15 years and the knowledge base that both surrounds the project and to which it contributes.

Researchers need to demonstrate, especially in the early stages of a research network, an awareness of the often complex and competing existing published literature in the field that they are entering. When working collaboratively, there must be collective agreements on key meaning from the outset. In the ISSPP case, in the late twentieth century, there was already, on the one hand, a significant international literature on effective schools. Much of this was based primarily on quantitative measures of school students (e.g. student test/examination results) and then sought to find key indicators for their effectiveness rather than how principals contributed to this. On the other hand, researchers on school leadership itself had for many years constructed different, often competing, models of leadership which were often based upon self-report by principals, critical theory stances, or adaptations of existing research into business leadership (Mulford 2007). These were not always empirically founded or able to be directly or indirectly associated with students' measurable performance. Indeed, at that time most research had been about leadership rather than effective leadership. Notable exceptions to this in the late 1990s and early in this century were in Canada, with the conceptually sound empirical research of Leithwood and colleagues (see Leithwood and Riehl 2005), and in Australia, with the work of Silins and Mulford (2002) into the effects of secondary school leadership on student learning.

As the work of the ISSPP progressed, it was important to continue to keep up to date with emerging research. Alongside this, the awareness by governments of the key roles played by school principals in school improvement increased. In America, the Task Force on Developing Research in Educational Leadership of Division A of the American Educational Research Association presented a summary of well-documented understandings (Leithwood and Riehl 2005). These were later developed into

seven and then ten strong research-based claims for successful school leadership, as part of a government-funded project in England (Leithwood et al. 2006; Day et al. 2011). In their wide-ranging review of secondary schools functioning in the context of educational reform, Silins, Mulford, and Zarins (2002) went so far as to identify three major, aligned, and sequential factors in high school reforms, that is, how people are treated, the presence of a professional learning community, and the presence of a capacity for learning. A later meta-analysis of the quantitative research on effective leadership (Robinson et al. 2009) identified the importance of 'instructional' leadership.

CONTRIBUTIONS TO KNOWLEDGE OF THE FIELD

As it progressed, the ISSPP began contributing to this knowledge base. By 2005, the project began its major contributions, beginning with a special issue of the *Journal of Educational Administration* (43(6), 2005), six further journals (*International Studies in Educational Administration*, 35(3), 2007; *Journal of Educational Administration*, 47(6), 2009; *Journal of Educational Leadership, Policy and Practice*, 24(1), 2009; *Leadership and Policy in Schools*, 10(4) 2011; *International Journal of Educational Management*, 26(5), 2012; *Journal of Cases in Educational Leadership*, 1(3), 2012), books (e.g. Day and Gurr 2014; Day and Leithwood 2007; Moos et al. 2011; Ylimaki and Jacobson 2011), and many individual journal papers.

The research has revealed that, regardless of cultural, policy, and school system contexts, there are more similarities than differences between the values, qualities, strategies, skills, and actions of successful school leaders. It has challenged the efficacy and pervasiveness of single models that have dominated much of the academic writing and research that investigates school principals as against successful school principals. The ISSPP has shown that successful principals do not all lead and manage in the same way. Relatively few are only heroic, charismatic, transformational, or instructional, though all demonstrate features of these and other models. Some may begin as autocrats, but gradually build more democratic, participative structures and cultures. All display qualities of discernment, intellectual and emotional wisdom, commitment, courage, and resilience. Principals in schools in challenging circumstances especially seem to be acutely aware of the need to nurture staff, students, parents, and others in order to build successful learning communities. The findings both

confirm, complement, and go beyond findings from others' previous and subsequent research. They provide new insights and powerful images, particularly about the kinds of reflexivity, inner values, philosophies, entrepreneurial, and intra- as well as inter-professional qualities, skills, and strategies that the leadership and management of successful principals demand, regardless of culture, country, and school context.

INTERNATIONAL COMMUNITIES OF PRACTICE: MESO-SYSTEMS AND DEMOCRATIC PARTICIPATION

In many ways, the ISSPP network has come to illustrate a community of practice (Wenger and Snyder 2000). It provides an example of how collaboration over time may not only stimulate individual growth but also contribute to raised expectations, an increased collective sense of trust, sense of individual and collective identity and wellbeing, and high standards of research. As a network of equals, with a leadership steering group that set the initial vision and direction of travel, the ISSPP operates with a minimum level of bureaucracy. It seeks no subscriptions from its members and encourages new collectively agreed initiatives, providing peer support and mentoring when requested. This may be one reason for its continuing growth and sustainability. There are few 'rules'. Any researcher may apply to join. Collaboration was the founding principle in the ISSPP. All research instruments, analyses, and reporting frameworks were designed to a set of agreed common protocols across English and non-English speaking countries in order to provide rigor in the research itself and a 'baseline' for providing an on-going dialogue about emerging understandings of and insights into the work and lives of successful principals and school improvement in the participating countries. These principles were translated into practice through agreed, generic detailed survey and interview questions and protocols, data analyses, and reporting procedures. It is a condition of entry into the network that each researcher uses these, though they are also able to add to them.

The ISSPP is essentially a 'meso-system' (Bronfenbrenner 1979; Flaspohler et al. 2012) and may be distinguished from an organization by its deliberate rejection of bureaucratic hierarchies. Rather, it seeks to create possibilities for democratic relations. In this sense, it parallels Fielding's (2012) notion of community as both people- and task-oriented. Unlike many research collaborations by academics, it aims to be not only instrumentally but also morally successful, through, for example, working collaboratively with experienced and less experienced researchers who are in different phases

of their career journeys, and from educational systems in different jurisdictions which are themselves in different phases of their own development and where the development and use of research methods varies (Ärlestig et al. 2016). This moral dimension distinguishes the ISSPP from the more conventional researcher groupings normally brought together temporarily to conduct time-related research project work.

APPLICATIONS TO THE STUDY OF LEADERSHIP AND POLICY: METHODOLOGICAL ISSUES

This section highlights important methodological aspects from the work of the ISSPP that have ensured the rigor and standing of the project.

Guiding Questions It is important to focus upon formulating research questions that either replicate those in other research (which allows case accumulation possibilities), or seek to advance existing understanding by building further on these. In the ISSPP, we formulated five questions that did not, at that time, seem to have been answered by others:

- (i) What similarities and differences can be identified in the beliefs and behaviors of successful school principals across national cultures and policy contexts?
- (ii) Do different countries have different ways of defining success?
- (iii) How do high-stake assessments and accountability measures influence the practices of successful principals?
- (iv) Do different socio-economic contexts in which schools operate affect the ways in which successful principals work? Are different qualities and skills needed?
- (v) How do successful principals come to be successful? How do they learn about their work and acquire the skills needed to create and sustain school improvement?

Principal Selection In the ISSPP research, the schools of successful principals were selected against one or more of the following criteria:

- Schools that had received a positive external and independent inspection/review report with regard to student progress and attainment, and the quality of leadership and management of head principals;

- Schools which, on the basis of national tests and examination results (or the nearest equivalent) could be shown to be improving in performance over time;
- Schools in which the principals were widely acknowledged by their professional peers (through the local, regional, and/or national networks of professional associations) as being successful leaders;
- Schools that had been successful, or grown to become successful, during at least a five-year consecutive period of the principal's leadership. It was agreed that this was a sufficient minimum term during which the principal could reasonably be expected to have exerted a sustained influence throughout the school.

Principals were, therefore, selected in each research site in each country 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. Selection of schools and principals has to cover a wide range of contexts and leadership challenges. This is achieved by means of a matrix which is constructed around four dimensions which will allow for theory-generating case studies to be chosen on the basis of representing extreme cases or maximum variation sampling. These dimensions are: school size and education phase; socio-economic and socio-cultural settings; time spent as principal; and sex of the principal.

Multi-perspective Methods The use of multi-site case study methods is based on the assumption that the concept of success when applied to the leadership provided by principals is a contextualized relationship construct, as well as an attribution on the part of those who experience such leadership. The research is guided by the agreement between the founding group that: (i) multi-perspective data about successful principals was likely to provide richer, authentic, more nuanced data than had hitherto been available; (ii) such data were best provided by those with close knowledge of the principal, such as teachers, students, parents, non-teaching members of the school, students, and other community members. The exact nature of who is interviewed can vary from context to context.

In Australian cases from Strand one, the following data were collected (Gurr et al. 2007, 2010):

- individual interviews with the principal (two interviews), assistant principal, curriculum coordinator (or equivalent), up to six other teachers, school council/board chairperson, and a school council/board parent member
- group interviews with parents (two groups of 8–10) and students (two groups of 8–10)
- documents illustrating school achievements and student attainment

The following were questions asked of the principals in the first interview of the Australian cases (taken from Gurr et al. 2003).

1. Describe the ethos or philosophy of the school. (What direction is it heading?)
2. What is your personal vision for the school?
3. What are your priorities as a principal for this school?
4. What are you planning over the next few years?
5. Identify the challenges facing the school and your principalship. (How do you lead the teachers—school community—in dealing with these challenges?)
6. To what extent has the school become more or less challenging?
7. Can you describe what makes this school successful? In what specific ways is it successful? What criteria do you use to measure its success?
8. How do you know you are doing a good job?
9. How do you see your role as leader in this school?
10. What leadership strategies have worked well and which ones do you think are less effective?
11. Can you describe a situation, a complex issue, or challenge you handled well? Can you describe a difficult challenge or issue you would like to have handled differently?
12. As a principal you will have relationships with numerous groups—community, colleagues, staff, council, parents, and students. What are the differences, if any, in relating with these groups?
13. What non-professional sources of support and encouragement do you use in doing your job?
14. On average, how many hours per week do you work during term time? How do you relax or ‘get away from it all’? How often? Does this help? In what ways?

Conclusion: In the follow-up interview we would like to hear about a narrative of examples of professional success and failures.

Generally, two researchers conducted each interview. Interviews were recorded and transcribed, and notes were taken by the second researcher. In addition to the collection of primary data, evidence was also obtained from such secondary sources as school development plans, school prospectuses, inspection reports, newsletters, and examples of media coverage. These sources were used to contextualize the empirical data and were a means of confirming their validity and reliability.

The choice to collect multiple-perspective data was based in part upon recognition that single-source reports of principal work was of limited value. It was based, also, on the observation made at the beginning of this chapter: that most researchers are only able, because of circumstance, to conduct relatively small-scale research. In methodological terms, it was clear that the triangulation of multiple sources of data would provide trustworthy, credible, and authentic cases. Entry, access, and ethics were made easier to attain in this case because we were investigating why and how, rather than whether, school success had been achieved.

Content Analysis: An Inductive Approach The ISSPP guidelines provide members with an overall structure on how the data are analyzed systematically among the project research teams using a common framework. In this way, we have been able to contrast like with like across the teams.

The analysis of the field evidence should not be deferred until the end of the project; it should begin early and proceed concurrently with the collection of data in schools so that the two become closely integrated. Indeed, in subtle ways the collection and analysis of data begin to inform each other (Miles and Huberman 1984), and it is a reflective activity, both individually and for the team as a whole. This process of reflection takes a number of forms including periodic meetings of the project team at which progress is reviewed and ideas and opinions discussed, usually as a result of analytical notes (or memos) based on the reflections of individuals. These reflective documents and the discussions prompt further the process of analysis by helping the team to move between concrete field data and conceptualization. They also provide an accountable record, or audit trail, which demonstrates that the process of reflection and inductive data analysis is not only occurring but also proceeding in a manner which, while being flexible and eclectic, is nevertheless orderly and systematic.

The sheer volume and diversity of the field data collected by the team make it necessary to organize it into smaller homogenous units of

information in order to begin to make sense of it. Consequently, the project, by means of the reflective process described above, is likely to arrange the data into segments of material based on an organizing system derived from the data themselves. The process, therefore, is essentially one of inductive cross-case analysis (Miles and Huberman 1984), the main outcome of which is a two-dimensional matrix with issues and themes related to school leadership on one axis and the different sources of evidence (e.g. principals, teachers, students, parents, etc.) on the other. In the early stages of this process the organizing categories, especially the issues and themes, is regarded as being tentative and provisional and are likely to be subsequently modified in the search for a more satisfactory system. Manipulating the fieldwork data in this way is a time-consuming and intellectually-exacting activity.

Having arranged the data for purposes of analysis in such a way, it is possible to compare what the different sources of evidence have revealed about leadership in their schools or what a particular category of stakeholder has to say about different aspects of leadership. The richness of the insights provided by the analysis of the different perspectives of the various stakeholders interviewed is reflected in the way in which the research is reported.

While the process of inductive analysis consists of deconstructing and processing or reconfiguring the data in intellectually rigorous and trustworthy ways, the ultimate objective of the research is a synthesis of the project's empirical evidence with the theoretical constructs on leadership in schools that are already available in the literature. In other words, the aim is to go beyond what Hycner (1985) called a 'composite summary', and what Patton (1990) has referred to as a 'descriptive account of patterns and themes'. Instead, the final goal is to offer a critique of existing theories of effective leadership in schools in the light of the project's empirical findings and to suggest some provisional hypotheses, new concepts, or theoretical categories and some tentative thoughts on substantive theory.

Sustaining the Network: Micro-politics and Professional Identities An issue rarely raised in writings about international research networks is the challenge of generating and sustaining momentum. Motivations and abilities to meet deadlines and write well for publication are often determined by individuals' capacities to manage context-specific factors. Thus, there are inevitable fluctuations in the research work of individuals. Not everyone in

a network will be able to draw upon financial or other human resources, and at any given time, other priorities may delay or interrupt their leadership research. Without an acceptance of this, and a willingness to adjust and re-adjust deadlines, it is unlikely that research networks of university colleagues will survive. Yet this juggling act is framed by an individual and collective desire for progress. There are no easy answers, and it needs to be regarded as a dilemma to be managed rather than a problem to be solved.

It is likely that all experienced colleagues, consciously or unconsciously, will wish to reinforce the findings of their previous work. Within the ISSPP network, for example, issues of ideologies and past work that colleagues have worked so hard and long to establish and cannot abandon need to be managed. Almost by definition, these forms of diversity (e.g. a love of inventing and re-inventing leadership models, or investigating success through Foucauldian lenses, relying on the use of empirically outmoded terms) must be respected and managed. Indeed, paradoxically, they add to the richness of the work.

Because language, motivations, relationships, and agendas will also vary among researchers, there are likely to be several impediments to building and sustaining network trust such as:

- (i) Power (Who decides on the focus, process, and intended outcomes? How can parity be assured?)
- (ii) Roles, responsibilities, relationships (Who is responsible for what, and how are resources gained and distributed?)
- (iii) Time (How may work, once begun, be sustained? How is time financed? By whom?)
- (iv) Reciprocity (How can trust and trustworthiness be built?)
- (v) Credibility (How may university staff show that their knowledge and skills may add value?)
- (vi) Micro-politics (How may contrived collaboration move to authentic collaboration?)
- (vii) Outcomes/reward (How can both/all parties benefit in ways which will enhance their own moral purposes and fulfill organizational requirements?)
- (viii) Boundary management and boundary crossing as languages, norms of thinking/habits of mind implicit in the different academic and practitioner communities need to be understood and learned (between roles and responsibilities within the partnership)

- (ix) Sustainability (How can parties plan for the ending of the partnership?)
- (x) Knowledge (How can academic and practical knowledge be equally respected and used?)

There are likely, then, to be several predictable and unpredictable problems that need to be managed and, over time, resolved. All researchers will not necessarily always be able to maintain a consistently high level of motivation throughout their membership of a collaborative network. They are likely to experience fluctuations, associated with the quality of relationships and leadership in their work environments, as well as the nature of new policy demands, changes in societal expectations, and unanticipated changes in personal circumstances (e.g. retirement, change of job). Energy levels may also erode over time.

RECOMMENDATIONS FOR NOVICE AND EMERGING SCHOLARS

In addition to the discussions in the previous section, two areas are explored for those less experienced scholars wishing to engage in international research networks.

Understanding and Accepting Differences It is difficult to overestimate the time and immense leap that researchers from different countries need to make in order to understand the cultures and language that informs colleagues. We are all products of particular national, cultural, and personal histories, and members of ISSPP found that they needed to regularly reflect on the ways that these influenced them. It is easy to underestimate the amount of time needed in order to come to a deep understanding of the ways schools work in our different countries. For example, the *Bildung/Dannelse* of liberal education in Scandinavian schools was and is very different from the Democratic principles that underpin the work of those in America. Success itself is difficult to quantify when students are not tested until the age of 16, as is the case in Swedish schools.

Leadership: The Need for Researcher Resilience For those in leadership roles within networks, resilience can be fostered or diminished through the environment (e.g. the network leadership interventions in establishing nurturing structures and cultures). As with any group, factors which have the strongest impact upon research network members' satisfaction are

trust and fairness, followed by sense of community, meaning of work, availability of resources, and work-privacy conflict (the compatibilities or incompatibilities of working and private lives). Colleagues may respond positively or negatively in the presence of challenging circumstances, and this will depend on the quality of organizational or colleague leadership, as well as the strength of their own commitment. Extended collaborations in networks need to be managed in order to avoid their potential for collaborative inertia (Huxham and Vangen 2005: 13). It follows that academic optimism (Hoy et al. 2006) is a necessary constituent not only for researchers but also for leaders of research projects. Indeed, Beard et al. (2010) also associate academic optimism with enabling cultures, defined by Hoy and Miskel (2005) as hierarchies that help rather than hinder, and systems of rules and regulations which guide problem solving rather than punish failure. Successful networks that are sustained are, by definition, journeys of hope based upon a set of ideals. Arguably, it is our ideals that sustain us through difficult times and changing personnel and professional environments. They are an essential part of researcher resilience.

CHAPTER SUMMARY

This chapter has described and discussed issues in the design and conduct of a large, sustained multi-national, collaborative research network. There has been insufficient space to elaborate further upon the data collection, analyses, and results (See website for further information). The issues raised here are, however, relevant to all researchers as they begin to lay down the foundations for their work and also to those who seek to collaborate with colleagues nationally and internationally. In order to advance knowledge, all research must demonstrate its robustness, fitness for purpose, trustworthiness (whether this is framed as reliability, generalizability, or validity) and relevance (whether this is framed in terms of influence, utilization, or impact). All research that is co-designed needs to demonstrate features associated with successful collaboration such as trust-building, sustained interactivity, emotional literacy, commitment, resilience, and acceptance of diversity.

The claim of this chapter is that growing and sustaining multi-national collaborative research networks goes beyond attending only to the technical merits of particular research methodologies and methods, though it includes this. Successful research networks, like successful schools, need many leaders. Successful research network leaders also need to understand

and be able to work with the intellectual, social, and emotional capital embedded in relationships between all members of the research network community and build the organizational capital embedded in the network's structure and cultures if its members are to achieve success in producing rigorous research that is also persuasive in today's international policy contexts.

Recommended Readings

Arlestig, H., Day, C., & Johansson, O. (Eds.). (2016). *A decade of research on school principals: Cases from 24 countries*. Cham: Springer.

This edited volume presents research covering the past 10–15 years from International Successful School Principalship Project. The book is organized into five parts, with each part illustrating leadership studies in nations with similar contexts.

Day, C., & Leithwood, K. (Eds.). (2007). *Successful school leadership in times of change: An international perspective*. Toronto: Springer.

This edited volume presents research from several countries which describes the similarities and differences in the work of principals. In particular, it describes what successful principals do and the ways in which the international contexts shape their work.

REFERENCES

- Arlestig, H., Day, C., & Johansson, O. (2016). *A decade of research on school principals*. Heidelberg: Springer.
- Beard, K. S., Hoy, W. K., & Hoy, A. W. (2010). Academic optimism of individual teachers: Confirming a new construct. *Teaching and Teacher Education*, 26(5), 1136–1144.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Day, C., & Gurr, D. (Eds.). (2014). *Leading schools successfully: Stories from the field*. London: Routledge.
- Day, C., & Leithwood, K. (Eds.). (2007). *Successful school leadership in times of change*. Dordrecht: Springer.
- Day, C., Harris, A., Hadfield, M., Tolley, H., & Beresford, J. (2000). *Leading schools in times of change*. Buckingham: Open University Press.
- Day, C., Sammons, P., Leithwood, K., Hopkins, D., Gu, Q., Brown, E., & Ahtaridou, E. (2011). *Successful school leadership: Linking with learning*. Maidenhead: Open University Press.

- Fielding, M. (2012). Education as if people matter: John Macmurray, community and the struggle for democracy. *Oxford Review of Education*, 38(6), 675–692.
- Flaspohler, P. D., Meehan, C., Maras, M. A., & Keller, K. E. (2012). Ready, willing, and able: Developing a support system to promote implementation of school-based prevention programs. *American Journal of Community Psychology*, 50(3–4), 428–444.
- 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., & Mulford, B. (2007). Instructional leadership in three Australian schools. *International Studies in Educational Administration*, 35(3), 20–29.
- Gurr, D., Drysdale, L., & Mulford, B. (2010). Australian principal instructional leadership: Direct and indirect influences. *Magis*, 2(4), 299–314.
- Hoy, W., & Miskel, C. (2005). *Education administration: Theory, research, and practice* (7th ed.). New York: McGraw-Hill.
- Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, 43(3), 425–446.
- Huxham, C., & Vangen, S. (2005). *Managing to collaborate: The theory and practice of collaborative advantage*. London: Routledge.
- Hycner, R. H. (1985). Some guidelines for the phenomenological analysis of interview data. *Human Studies*, 8, 279–303.
- Leithwood, K., & Riehl, C. (2005). What we know about successful school leadership. In W. Firestone & C. Riehl (Eds.), *A new agenda: Directions for research on educational leadership* (pp. 22–47). New York: Teachers College Press.
- Leithwood, K., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006). *Seven strong claims about successful school leadership*. Nottingham: National College of School Leadership.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills: Sage Publications.
- Moos, L., Johansson, O., & Day, C. (Eds.). (2011). *How school principals sustain success over time: International perspectives*. Dordrecht: Springer.
- Mulford, B. (2007). *Overview of research on Australian educational leadership 2001–2005*. Monograph No. 40. Melbourne: Australian Council for Educational Leaders.
- Patton, M. (1990). *Qualitative evaluation and research methods*. Newbury Park: Sage Publications.
- Robinson, V., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why. Best evidence syntheses iteration (BES)*. Wellington: Ministry of Education.

- Silins, H., & Mulford, B. (2002). Schools as learning organisations: The case for system, teacher and student learning. *The Journal of Educational Administration*, 40(5), 425–446.
- Silins, H., Mulford, B., & Zarins, S. (2002). Organisational learning and school change. *Educational Administration Quarterly*, 38(5), 613–642.
- Wenger, E. C., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, 78(1), 139–146.
- Ylimaki, R., & Jacobson, S. (Eds.). (2011). *US and cross-national policies, practices and preparation: Implications for successful instructional leadership, organizational learning, and culturally responsive practices*. Dordrecht: Springer.



Taking Stock of Complementary Methods: The Perpetual Quest for Good Research Methods for Educational Leadership and Policy

Carolyn J. Riehl

The search for appropriate methods for documentation and research in educational leadership and policy has been going on for a very long time. Under the encouragement of educators like Henry Barnard and Horace Mann, governments and academics began gathering systematic information about students and schools in the mid-nineteenth century (Beadie 2016). Data scrutiny was an integral component of the scientific management approach that transformed educational administration over 100 years ago, and district leaders and state superintendents of public instruction became avid users of surveys about students, teachers, and schools (Callahan 1962; Strayer 1915; Timar 1997). Something akin to policy research began as early as 1915, when observers explored how multiple

C. J. Riehl (✉)
Department of Education Policy and Social Analysis, Teachers College,
Columbia University, New York, NY, USA
e-mail: riehl@tc.columbia.edu

© The Author(s) 2018
C. R. Lochmiller (ed.), *Complementary Research Methods for
Educational Leadership and Policy Studies*,
https://doi.org/10.1007/978-3-319-93539-3_18

school districts measured “teacher efficiency” and pondered how it could be managed more effectively (Boyce 1915). The lure of statistics, as Labaree (2011) calls it, helped lend legitimacy and prestige to both educators and educational researchers in eras when their knowledge base and applied practice were variable and uncertain.

It is important to have robust methods at one’s disposal for building theoretical and actionable knowledge in a field. As this volume demonstrates, many interesting and useful options exist for leadership and policy researchers. The chapters describe a rich range of methods, some in considerable detail and others through more general overviews, and they also illustrate the utility of methods through specific examples of research studies in which they are employed. The methods discussed run the gamut from what are traditionally thought of as quantitative research, to a broad family of methods that draw on qualitative data, and then mixed methods for combining the two. All of this should be useful for readers.

Of course, productive research does not always begin with the selection of a research method. Instead, scholars suggest it can be important to identify methods after posing research questions. Nonetheless, knowing that diverse methodological options are available can be very helpful for stimulating interesting research questions. Several chapters in this volume begin this way, using a particular method as an opportunity to frame unusual questions that might be asked in leadership or policy research.

Just as methods can inspire questions, rich and unique data sources can also encourage researchers to formulate different kinds of questions. The refinement of analytic methods with which to use them has often followed. Large cross-sectional data sets, for example, have made it possible to ask questions about the nested effects of policies, districts, schools, and teachers on student learning, and analytic methods for using those data have become more robust. Similarly, as district fiscal data, longitudinal administrative records, or log data from student use of interactive electronic learning platforms have become available, different analytic methods such as lagged reciprocal-effects models, latent class analysis, and educational data mining have received more attention (Baker and Corbett 2014; Bowers 2010; Boyce and Bowers 2016; Figlio et al. 2017; Loeb and Strunk 2003; Hallinger and Heck 2011). Several chapters in this volume describe data sources that may inspire new research projects.

In short, there are many points of entry in this text for helping novice researchers understand and expand their options. Some time ago, William Firestone and I co-edited a volume of essays laying out promis-

ing themes for new research in educational leadership (Firestone and Riehl 2005). In a chapter for that book titled “What Methods Do We Need?” (Riehl and Firestone 2005), we described three questions to be asked about the research methods available for addressing those new themes. First, could the methods answer the question of “what works, and why?” by using experimental approaches or otherwise accounting for causal relationships, while also taking seriously the imperatives of local adaptation and recognizing important covariates and the mechanisms underlying causal effects? Second, could research methods accommodate the epistemological foundations of diverse knowledge paradigms, expanding the basis of research beyond realist, post-positivist approaches to incorporate other ways of knowing? And third, could research methods be used to generate knowledge that might address underlying patterns of privilege and power in education, rather than implicitly or explicitly supporting the status quo? In this concluding chapter, I will try to synthesize what the authors in this volume have presented by discussing how their work helps to answer these questions anew, and what might come next.

DETERMINING WHAT WORKS AND WHY

Researchers have long-established interests in understanding what leads to particular educational outcomes and effects. This is as true for those who research educational leadership and policy as for those who study curriculum or pedagogy. We want to know what matters, what works, and why it works given particular conditions. Some researchers eschew any idea that explanations for outcomes are fixed, determinate, or generalizable, and they search for contextualized, socially constructed reasons for and interpretations of human action. Others look for more regularized patterns of causes and effects. Causal research with large numbers of cases generally involves statistical computations. But such analyses depend first of all upon the logical foundations of causal reasoning that have been developed by philosophers ranging from Aristotle to David Hume to John Stuart Mill to J. L. Mackie. These logical arguments often entail explanations of how to determine the necessary and sufficient conditions leading to particular outcomes, and they form the basis for various experimental and quasi-experimental research designs such as those described by Campbell and Stanley (1963), Cook and Campbell (1979), Ragin (2013), and King et al. (1994).

An important consideration in the logic of causality is that of the “counterfactual,” how to account for what might have happened, but was not observed—for example what the learning of a student receiving an educational intervention would be had that same student not received the intervention. Counterfactuals ideally make the treatment effect completely transparent, but they pose dilemmas for researchers because the same subject typically cannot both receive an intervention or treatment and not receive it. A related consideration is selection bias, the problem that arises if persons receiving an intervention under consideration differ from those not receiving it in ways that might interfere with assessing the true effect of the intervention.

Experimental designs with randomized assignment to treatment and control groups have been used to approximate the counterfactual and minimize selection bias (Borman 2009; Fisher 1935). Experiments figure prominently in the tradition of educational psychology research and, to some degree, classroom pedagogical research. For a long time, they were considered inappropriate or not feasible in leadership and policy research, largely because of perceived political, logistical, and ethical challenges of random assignment to teachers, classrooms, leadership approaches, or policy alternatives. The very small number of randomized studies that are frequently cited in the policy literature—especially the Perry Preschool study and the Tennessee STAR experiments with class size—attest to this, although the use of randomized controlled trials (RCTs) is slowly growing (e.g., Spillane et al. 2010). Instead, researchers studying leadership and policy tended to adopt the approaches of many social science researchers, building causal propositions from conceptual theorizing and exploratory empirical evidence, and then testing those propositions with statistical analyses.

Analytic methods in this “structuralist tradition” (Kaplan 2009) included path analysis, structural equation modeling, ordinary least squares regression, and hierarchical linear models. These quantitative techniques had the extra benefit of permitting the examination of covariates, or other social conditions or attributes (race, poverty status, etc.) that social theorists often believed to be important in producing or mediating educational outcomes. They measured associations, but could not completely deal with selection bias or account for causes; however, paired with clear logic models, they were often considered satisfactory for making causal arguments.

Many of these mathematical strategies were closely linked to the economists’ production function model. The federal *Equality of Educational*

Opportunity study, known as the Coleman Report (Coleman et al. 1966) is often considered the first consequential use of the production function and correlational analyses to study the effects of policy-relevant factors connecting school and non-school resources to student outcomes. While these analyses were rudimentary by today's standards, the report spawned an industry of school effects research that is 50 years old and going strong.

In the early 2000s, the federal government began to emphasize what it considered to be more rigorous research for education. In addition to calling for scientifically-based evidence in the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), known as the No Child Left Behind (NCLB) legislation, Congress passed the Education Sciences Reform Act of 2002, which established the Institute of Education Sciences (Orland 2009). Like NCLB, this law presented research standards stipulating that causal relationships could be demonstrated “only in random assignment experiments or other designs (to the extent such designs substantially eliminate plausible competing explanations for the obtained results).” The first director of Institute for Education Sciences (IES), Grover (Russ) Whitehurst, was a heavy advocate for RCTs. In its grant-making activity, IES began to differentiate among the types of research that were appropriate for exploratory, developmental, and scale-up research efforts, much as the Food and Drug Administration had been doing for years with regard to medical treatments and pharmacology (Riehl 2006a). The federal What Works Clearinghouse (WWC) was also established with the purpose of assessing the quality of research evidence on educational treatments and programs and providing research reviews and practice guides for decision makers and practitioners. IES and WWC promulgated the use of a framework of four tiers of research evidence, ranging from Tier 1, based on experiments using random assignment, to Tier 4, research that provides a rationale for an intervention based on a logic model and some empirical evidence of effectiveness. Only studies labeled as Tier 1 or Tier 2 met the standards set by WWC for being adequately “evidence-based.” In the 2015 reauthorization of ESEA, known as the Every Student Succeeds Act, once again states and localities were required to use evidence-based practices based on research evaluated on a four-point scale, and in late 2017, the WWC published its *Standards Handbook, Version 4.0*, over 120 pages in length, with extensive guidance for handling statistical analyses, missing cases, and other aspects of experimental and quasi-experimental research.

These developments were a boon for the experimentalist tradition (Kaplan 2009). But experiments are not always feasible in education, and experiments can be compromised by many factors (Gobo 2008), such as treatment/control group crossover, non-response, non-compliance, non-implementation, or measurement error. As Kaplan (2009) argued, random trials do make it possible to make causal and counterfactual claims, but that does not mean that there are no other ways to draw valid causal inferences from non-experiments or other forms of research. So, researchers also continued to use correlational statistical analyses, which fortunately had been evolving. As Morgan and Winship (2007) note, prior to the 1990s, researchers had become so enamored of their preferred statistical models that they neglected to attend to the logic behind causality, particularly the matter of the counterfactual:

For causal analysis, the rise of regression led to a focus on equations for outcomes, rather than careful thinking about how the data in hand differ from what would have been generated by the ideal experiments one might wish to have conducted. (p. 13)

Gradually, educational researchers who recognized these problems began to adopt analytic strategies for causal inference that were being developed and used in economics and statistics (Murnane and Willet 2011). These techniques deal more explicitly with selection bias and counterfactual unknowns; they include fixed effects models, instrumental variables, regression discontinuity analyses, difference-in-difference designs, and propensity score matching (Angrist and Pischke 2009; Gamoran 2009; Morgan and Winship 2007).

These statistical analyses are not a silver bullet. Although they seem to take counterfactual reasoning more seriously, it is not always clear whether or how they improve upon estimates of effects derived from earlier analytic strategies. Also, they continue to rest on some aspects of production functions that do not seem to map well onto the realities of schooling, including assumptions that effects are the “marginal productivity of a given unit of input” and the lack of accounting for mechanisms connecting inputs and outputs, especially those that have something to do with the actual processes of teaching and learning (Carnoy 2009, pp. 30–31).

In observing these developments in both the use of randomized experiments and the refinement of statistical methods for non-experiments,

Kaplan (2009) finds common ground that reflects positively on the developments and the variations of research they have engendered:

In the final analysis, neither the experimental nor structural approach to causal inference can legitimately lay claim to being a “gold standard” for methodological rigor. Both approaches have many well-known strengths and weaknesses... In terms of basic principles, both approaches (a) rightly reject the nihilistic post-modern relativism that seems to have infected education research of late, (b) urge rigorous standards of empirical data collection and measurement, and (c) subscribe to some variant of the counterfactual theory of causation. (p. 151)

Orland (2009) adds another positive interpretation of the methodological debates that have often seemed polarizing:

What makes the current emphasis on RCTs so interesting is not that it is viewed as uniquely scientific, but rather that it represents a deliberate policy strategy for more closely linking the fruits of educational research with the needs of decision-maker audiences. That view is based on the belief that, above all, educational decision makers are looking to the research community for clear guidance on very particular questions. (pp. 122–123)

Orland’s comment resonates with the impulses of others who are not completely absorbed by the challenge of finding determinate, discrete causal patterns. For example, when John Q. Easton became the second director of the federal IES, he explained that although he affirmed the good work IES had done to date in supporting more rigor in research methodologies, including the use of better forms of causal analysis, still ...

Going forward as we shape our new grant programs and priorities, we will expect our funded researchers and our evaluation contractors to better understand educational and learning processes and the mechanisms through which schooling policies and practices affect students. This means looking beyond what works and what doesn’t, but “how?” and “why?” and “for whom?” and “under what conditions?” This will require supporting research on the effects of practices and programs on different subgroups of students, testing hypotheses regarding mediating processes and mechanisms, studying the roles of classroom, school, and social contexts in moderating the effects of policies and practices. (Easton 2010)

All of these tensions and developments are reflected in patterns of research in educational leadership and policy over the past few decades, and they are also evident in chapters of this volume. Several chapters address topics related to the new class of quantitative analysis strategies for causal analysis. For example, Yongnam Kim, Stanley Lubanski, and Peter Steiner discuss the use of matching strategies for estimating treatment effects. In this method, pairs of treatment and control cases that appear to be similar based on other potential confounding variables are created, and non-matched cases are removed from the analytic sample. This technique compensates for the non-observation of the counterfactual for every case in the data set by assuming that each treatment or control partner in a pair serves as the counterfactual for each other. An estimate of the treatment effect is derived from averaging the difference in outcome scores for each matched set. One benefit of this approach is that various matching strategies are available to use with data sets that are structured differently, including single-level, multilevel, and clustered data, which is important given the complex, nested nature of educational inputs and outcomes. However, with these matching strategies as with older correlational models, an accurate specification of important covariates is essential. Matching strategies cannot compensate for the lack of a strong logic model explaining possible confounding or mediating conditions, and they will produce faulty estimates if the analytic model does not properly “identify” the likely treatment effect by restricting the contributions of other covariates.

Cassandra Guarino’s chapter also addresses the specification of models when using quantitative methods to assess the effectiveness of various educational treatments or conditions or programs (e.g., class size or teacher quality), and especially when using such models for high-stakes accountability or evaluation. Conceptually, the value-added and growth models she discusses are not particularly unusual or new; like others, they account for covariates and measure the effects of particular inputs and outcomes. But as Guarino emphasizes, it is important to specify models carefully to avoid bias and imprecision. Even something as basic as whether to include prior learning as a covariate or part of the computation of the outcome learning variable can have significant consequences. She argues that value-added models are preferable to growth models when evaluating teachers, because they account for more of the confounds that emerge when students are not randomly assigned to teachers. Even so, not all value-added models are equally appropriate, and some may actually add bias when estimating teacher effects.

Another way to study the impact of educational programs or treatments is through program evaluation, a multifaceted approach that incorporates qualitative and quantitative methodologies to inform action and decisions at multiple points in program life cycles. As Liz Hollingworth notes in her chapter, the objectives and uses of program evaluation differ from those of more basic research. This affects not only the questions that are pursued in an evaluation, but also how the research activities can be conducted. Hollingworth presents a basic but thorough framework that incorporates important elements in program evaluation that are sometimes passed over, such as evaluation of the logic model or theory of action for a program, and assessment of benefits relative to costs.

Even though James Coleman himself was far more interested in the effects of social factors on student performance in school (Riehl and Lyon 2017), *Equality of Educational Opportunity* was framed around Congress's request for a study of school resources that could be monetized, to inform and perhaps compel decision makers who set federal education budgets. Research on resources is one of the oldest forms of educational research (Hedges et al. 2016), but cost-benefit and cost-effectiveness studies and analyses of educational budgets and expenditures seem to occupy a somewhat separate, specialized space within the leadership and policy research domain, and they don't always answer bigger questions such as how educational costs relate not only to learning gains, but also to other social benefits (Weimer 2009). Nonetheless, studies of fiscal resources and processes can be connected to more general questions in education, and two chapters in this volume translate the "what works and why?" question into matters of educational economics and finance. Joshua Zender, Kenneth Smith, and John Kurpierz explain the value of analyzing fiscal data from an accounting perspective instead of the education production function, which is more commonly used by economists. These authors note that these forms of research can help improve fiscal planning and allocation decisions and can lead to more effective compliance, equity, and effectiveness for schools and school systems. This is especially important when courts are holding states and localities to account for many things, including educational adequacy, a relatively new standard in equity-related legislation (Rebell 2007, 2008). Zender and his colleagues describe necessary steps in preparing and using fiscal data for research purposes, including how to access fiscal data from a multitude of local and state sources, how to convert accounting data into formats more amenable to analysis, and how to communicate financial analyses to interested audiences.

While Zender and colleagues focus on the analysis of resources and expenditures (which are often obliquely reflected in accounting data), Tammy Kolbe and Rachel Feldman address the matter of evaluating education costs, another research approach that is somewhat underutilized by leadership and policy researchers. True program costs usually reflect far more than simple expenditures for a given educational service or treatment. The costs of educational interventions vary with degree of implementation and other factors, and these variations would be difficult to determine from expenditure data alone. Program costs are perhaps more familiar to educational economists than accountants, but much of the relevant data must originate with accountants who keep fiscal records. Methods for determining costs range from an “ingredients analysis” (Levin 1983) to calculating opportunity costs. As Kolbe and Feldman explain, program costs can assist leaders and policy makers in accurately assessing resources needed to replicate programs and determining how to distribute cost burdens across stakeholders. They illustrate their points with an interesting case study of a college readiness program that linked up with a new partner, creating a novel context for determining program costs.

The Zender, Smith, and Kurpierz chapter and the Kolbe and Feldman chapter discuss methods for using particular forms of data—financial data. Angela Urick pivots to another data source and examines how large-scale data sets can be used in secondary data analyses. She provides a convincing rationale and many examples of this approach to leadership and policy research and reminds readers that the particular strategies used in secondary data analysis should be well matched to the research questions posed and the nature of the data. She also helpfully notes that while large-scale data sets are usually derived from prefigured frameworks and themes, researchers who use them have the opportunity to explore using the data as indicators or operationalizations of new concepts. It is worth adding that in some cases, sponsoring agencies invite input into the construction of data collection instruments, giving researchers even more opportunity to use large data sets for theoretical as well as empirical development.

Urick argues that “established national and international datasets can be used to analyze trends across the country, or countries, which better informs state and federal policy, extends research on commonalities or differences across systems of education, and provides a means for advancing quantitative analysis techniques.” On the other hand, however, Hutt (2016) cautions that, in the highly decentralized United States at least, national data sets since the Coleman Report have simultaneously presented

both a more precise and a more distorted view of educational patterns, by creating an account of an imaginary national system that obscures local details and variations that are the real source of educational outcomes and inequities.

Two chapters focus less on educational inputs, mechanisms, and outputs and more on the processes of policy work, by employing quantitative methods to study the social networks within which this work happens. Kara Finnigan, Daniela Luengo-Aravena, and Kim Garrison describe social network analysis as both “a theoretical lens and a unique methodology,” a characterization that has also been applied to production functions, suggesting the close link between a way of thinking about a problem and the methods used to examine it. Network analysis, in these authors’ view, can illuminate the processes of social cognition, policy advocacy and diffusion, and the development and deployment of social capital. These processes take place in teacher networks and administrator networks, among others, and they are relevant to understanding the exchanges that take place among persons and groups at all stages of policy cycles, from problem definition to program evaluation. The example described in this chapter, a study of advocacy networks around charter schooling, illustrates how network analysis can support descriptive analyses, but it is easy to see how this lens and methodology can be used for more critical analyses (discussed below).

Christopher Day and David Gurr take network research in a different direction and describe an international network of leadership researchers who are pursuing programs of research on specific leadership-related topics. The network represents an unusual effort to systematize and align the work of multiple researchers around research questions, methods, and sites. In addition to introducing the important theme of cross-national, comparative work in leadership and policy, Day and Gurr, perhaps unwittingly, allude to the possibilities of studying networks of researchers who are using deliberate strategies to create intellectual capital through their work.

The chapters discussed thus far draw mostly on various methods of quantitative research, but this volume of complementary research methods also, appropriately, places strong and useful emphasis on qualitative research for understanding what works and why. The chapters on qualitative methods ought to be very useful in concrete and more general ways to novice researchers. For example, in their overview of qualitative research for leadership and policy, Jeffrey Brooks and Anthony Normore note that while these fields began by being grounded in other disciplines such as anthropology, political science, and sociology, over time the disciplinary

linkages have become weaker and researchers have drawn on multiple perspectives to advance “understanding of how dynamics such as influence, power, communication, collaboration, administration, abuse, equity, management and organizations work in educational organizations and contexts.” However, they rue research designs that do not include in their analytic models factors, qualities, or considerations that are specific to education. This is especially problematic in the case of leadership research, which often focuses more generically on leadership skills or processes and pays less attention to the contexts, practices, or objectives that are particular to education. Don’t be afraid to wander beyond disciplinary boundaries, but don’t forget what’s distinctive about education, they suggest; this is heady advice for beginning researchers searching for ways to put a distinctive stamp on their chosen fields of inquiry.

Getting a bit deeper into the details of qualitative methods, Susan Bush-Mecenas and Julie Marsh describe the utility of the “case-ordered meta-matrix display” for developing theoretical patterns, themes, and analyses. First introduced by Miles and Huberman (1994), case matrices are a data reduction strategy that can make the visual examination of data easier and lead to more complex pattern discovery. Matrices don’t automatically ensure that, as countless qualitative researchers have written, “themes emerge from the data,” but they increase the probability that analysts will notice interesting and valid patterns. This is even more likely if researchers are consistent and explicit in laying out the criteria by which they describe, reduce, and classify phenomena. Bush-Mecenas and Marsh refer to several interesting resources for analyzing matrices from other disciplines and provide several clear examples from their own research. They note that this approach to multicase qualitative research can help illuminate local causality as well as generate working hypotheses for theoretical generalization (Lincoln and Guba 1985). Matrix analysis, they acknowledge, sometimes gives findings a “quantitative feel,” but “the underlying logics are firmly rooted in qualitative inquiry.”

Just as the Finnegan, Luengo-Aravena, and Garrison chapter illustrated how leadership and policy processes could be studied with the quantitative tools of social network analysis, two chapters in this volume explore how they can be studied with particular qualitative methods centered on text and discourse. It has long been acknowledged that educational leaders do things with words (Gronn 1983; Robinson 1995). Policy also functions discursively and cognitively. So it makes sense to employ language-based methodologies for studying leadership and policy. In this vein, Jessica

Lester and Justin Paulsen provide a detailed overview of three broad strategies: critical discourse analysis, discursive psychology, and conversation analysis. Each method has distinctive core assumptions, techniques, and applications, but all three methodologies share some foundational ideas about the performative nature of language, the important role of language in the dynamics of power, and language use as a site where macro-level forces and ideologies meet micro-interactions among situated individuals. Similarly, in their chapter, Michelle Young and Sarah Diem discuss how critical policy analysis can be used to illuminate the rhetorics, texts, and sub-texts of policy.

The research world is not neatly divided into two camps, and Colleen Chesnut, John Hitchcock, and Anthony Onwuegbuzie offer an overview of mixed methods research for studying leadership and policy. As they explain, not only can mixed methods research be useful for different kinds of policy inquiry, including policy research (understanding and addressing a policy problem), policy evaluation (assessing the value of a given policy), and policy analysis (providing specific information to inform policy choices), but it can also incorporate different paradigmatic points of view, including transformative and pragmatic epistemologies along with realist, post-positivist ones. The authors include helpful references to a set of strategies for ensuring the “legitimation” of mixed methods studies. Related to the concept of validity, these legitimation strategies allow researchers to address problems that are unique to mixed methods and document the quality and validity of their approaches.

The chapter by Morgan Polikoff, Shauna Campbell, and Shira Korn picks up the theme from Brooks and Normore about the need for more focus on education-specific aspects of leadership and policy, and puts it down at the feet of curriculum research. Polikoff and his colleagues situate curriculum materials as an educational input with significant potential consequences for student learning and posit that tracing the adoption, implementation, and effects of curriculum materials will put school-specific flesh on the bones of education leadership and policy. They suggest that both quantitative and qualitative methods can be used to explore how curriculum materials affect students’ opportunity to learn, how they are used by teachers, and how they affect learning. These approaches could include the analysis of state or local data on textbook purchases, ethnographic research with teachers to observe textbook use in the enacted curriculum, content analysis of particular textbooks, perhaps even random trials of textbook use on student outcomes. Expanding outward, studies

of textbooks could expand into research on teachers' use of online sources. This discussion of the processes by which curricula and textbooks are adopted suggests numerous opportunities for qualitative research on decision makers' sense-making and even points to network analysis as an option for tracking the interactions and influences among key players in the adoption and implementation process.

As this discussion of just some of the chapters in this volume illustrates, studying "what works and why?" continues to draw the attention and efforts of researchers working with diverse questions, data sets, and methodologies. The chapters show that researchers have taken the call for more rigor in causal analysis seriously, but have not been dissuaded from using other methods that can shine a light on many different corners of leadership and policy. Orland (2009), drawing on the work of Nancy Cartwright (2007), contrasts research methods that "clinch conclusions" but are narrowly applicable with those that "vouch for conclusions," including "well-designed and scientifically rigorous descriptive and correlational studies, formal meta-analyses and evidence-based research syntheses, as well as qualitative investigations of significant educational programs and phenomena" (p. 124). The chapters in this text suggest that the next generation of leadership and policy researchers will be well poised to deliver these multiple forms of research.

INCORPORATING EPISTEMIC DIVERSITY AND UNMASKING POWER AND PRIVILEGE

Research in educational leadership has deep roots planted in the realist, post-positivist knowledge paradigm (Donmoyer 1999; Willower and Forsyth 1999). Even when research involves qualitative methods, it frequently leans more toward "soft positivism" (Miles and Huberman 1994), serving the needs of objective, variable-oriented research, than toward interpretivism. Still, research on leadership emanating from the interpretive and critical traditions has been robust and continues to help define the field. Critical and interpretive studies of leadership extend from the work of Thomas Greenfield (Greenfield and Ribbins 1993) and William Foster (1989), to studies of women educational leaders, social justice leadership, and critical race theory, to exploration of multiple dimensions of identity and leadership. Interpretive and critical studies of leadership often focus on local interpretations and adaptations, which to many is the

most meaningful context or level at which leadership can be understood. The same can be said for research in education policy. While realist and rationalist epistemologies dominate policy research, especially evaluation research and research conducted for government agencies and some foundations, interpretive and critical studies frequently dot the landscape of policy research.

In this volume, the critical paradigm is very well represented in a clutch of chapters that point in promising, exciting, and important directions for leadership and policy research. Irene Yoon explains why critical perspectives are needed in leadership research. These perspectives recognize that “schooling and society are inherently biased and normativized to exclude and even dehumanize students of color, students perceived as disabled, gender identity, students who live in poverty and students who are undocumented.” Moreover, she argues, “research has played an important role in colonization, slavery, and legalized oppression.” Leadership research can challenge the structures and logics of schooling and pursue social change, particularly by staying attuned to the dynamics of power and oppression in local contexts. Yoon goes on to discuss the epistemologies and methodologies that can support the critical project in leadership research. They include critical ethnography and participatory research, two methods that require a careful examination of one’s own positionality and transform the distance between researchers and those they seek to understand. While this may be unsettling for researchers who themselves benefit from the status quo, Yoon maintains that “dismantling institutions and critiquing the actions of people in power does not mean treating them as simplistic, monolithic, or uncomplex and evil.”

In their overview of qualitative methods, Brooks and Normore note that researchers have options for the epistemologies and purposes that will define their approaches to research. But they also note that qualitative researchers should consider “the relational, power, and gatekeeper dynamics” that may influence their research, including those they create themselves if they study their own students as well as those imposed by others who offer or deny access to different settings and informants.

Meagan Call-Cummings and Melissa Hauber-Özer write, in their chapter on action research (AR) and participatory action research (PAR), that these methodologies can democratize research and even go farther to create “an arena of resistance and struggle” by securing the authentic participation of subjects who have often been marginalized in educational contexts. In particular, they argue, forms of participatory research, including those

focused on youth, can contribute by “challenging and interrupting systemic oppression and fomenting institutional change.” These authors offer a number of concrete and inspiring examples of AR/PAR research, and they caution novice researchers to take the commitment to full participant involvement seriously, not just symbolically.

Lester and Paulsen’s chapter on language-based methodologies also has a distinctively critical tilt. Not surprisingly, critical discourse analysis is often used to interrogate the presence of power and privilege, ideology, and exploitation in policy texts and discourses. As noted above, language-based analyses offer opportunities for researchers to explore the penetration of macro-social ideologies and structures into micro-level meanings and interactions. These methods have great potential for unmasking the most subtle and surreptitious forms of oppression in leadership and policy.

Finally, the chapter by Young and Diem is a provocative comparison between traditional and critical policy analysis in education. The authors maintain that while traditional policy analysis (TPA) often seeks to determine the “best” policy outcomes through rational investigation, including outcomes that serve values such as equity, critical policy analysis (CPA) typically incorporates a different starting and ending point. CPA begins with a critical epistemology that assumes the presence of enduring structures of inequality in education and defines the knowledge project as uncovering those structures with the express purpose of altering them. CPA connects this paradigmatic stance to interpretive theories such as feminist theory, post-structuralism, or critical race theory, which establish the analytic parameters through which findings are made meaningful. Critical policy analysis seeks to get under the surface of policy processes and products, even of seemingly equitable policies, to understand the dynamics of power that drive them. CPA takes very little for granted and even less at face value, save for the assumption that oppression exists. This approach can seem a rebuke to traditional policy analysts who believe the best route to truth is through an objective examination of the facts, uncluttered by a priori assumptions about inequity. But critical policy analysts are unapologetic about the need to reveal the structural contradictions and systematic biases in policy activity.

These forays into critical research unfortunately do not say much about critical race theory or other methodological or theoretical approaches that are specifically relevant to issues of race, class, gender, or other forms of identity and difference (Blaisdell 2016; Brown and De Lissovoy 2011; Dixson and Rousseau 2005). That is a weakness of the critical emphasis in

this volume, and an unfortunate reflection of the limited diversity among leadership and policy scholars. Indeed, this limitation can and must be corrected by ensuring that many more researchers who are persons of color or represent marginalized populations and perspectives are allowed to have important places at the leadership and policy research table.

EVEN MORE OPTIONS

Novice researchers reading this volume might easily be overwhelmed by the extensive menu of methods. They might become even more overwhelmed by the consequences of selection, for when researchers choose the methods they will focus on, they are also usually choosing the discourse community within which they will work. Still, it is important to remember that this volume reflects just the beginning of an even longer list of possibilities. In these final pages, I offer several comments about additional options.

First, research in education policy and leadership can draw more not only on the theoretical frameworks but also the methodological tools of the academic disciplines. Discipline-based frameworks and methods relevant for leadership and policy might include historiography, visual sociology, comparative politics, and behavioral economics. Also, additional causal inference methods from econometrics are not described in detail in this volume but certainly are relevant. Even more, because leadership and policy researchers frame questions that cross disciplinary boundaries, creative combinations of these methods offer additional options.

Second, researchers can explore options for more structured causal analysis using qualitative data. Not all qualitative methods lend themselves to causal study and some analysts argue that qualitative research should remain focused on interpretations, reasons, and justifications for actions and their consequences as befits the “qualitative template” (Brady and Collier 2010). But I have already noted how often qualitative research is used in post-positivist research. Qualitative comparative analysis is one technique that leadership and policy researchers have begun exploring as a way to develop more systematic causal explanations from case study research using “fuzzy set” Boolean logic (Caves et al. 2016; Ragin 2014; 2013).

Third, researchers can employ quantitative research in the service of critical questions. Some higher education researchers have been working for years to develop critical quantitative methods (e.g., Carter and Hurtado 2007; Stage 2007; Stage and Wells 2014). Two special issues of the journal

New Directions for Institutional Research have been devoted to this approach and its applicability is probably much wider.

Fourth, researchers can use methods that help them explore the future. Research and evaluation methods, including those described in this volume, often use as their objects of study existing policies, practices, or conditions. But many new research methods train their attention on emerging conditions and innovations; indeed, these methods are meant to deliberately help nurture innovations along. These methods include design research (Penuel et al. 2011), design-based implementation research, developmental evaluation (Patton 2011), and improvement science (Bryk 2009, 2015; Bryk et al. 2015). Though these methods differ, they all share an orientation toward adaptive change, continuous improvement, and system learning. One of the hallmarks of improvement science, for example, is the idea of networked improvement communities that pursue learning and innovation together. This relates to yet another idea for novice researchers—how to utilize research methodologies and build programs of inquiry that honor the knowledge needs of practitioners and decision makers. Research-practice partnerships are not a methodology per se, but can be an important venue for pursuing improvement-oriented research (Coburn and Stein 2010).

Finally, it is important to note that the methods of research described in this volume, and those to come that will no doubt improve upon current methods, are of necessity “complementary” methods. Desimone (2009) writes that since educational policy is itself a complex endeavor, multiple methods are required to capture that complexity. Gamoran (2009) agrees, positing that “policy analysts need information of at least the following sorts: estimates from a production function, which specifies the impact of introducing a program or practice into the existing system; a cost-benefit analysis, including the full costs of implementation; an interpretive understanding of the meaning and significance of the new program or practice to participants, including possible unintended consequences; and an understanding of the political context of the potential reform” (p. 109). Not only can different methods generate very particular knowledge about the contexts, implementation, and outcomes of educational leadership or policy, they can also serve a conceptual or enlightenment function to help people think differently and pay attention differently (Weiss 1980). In this regard, I end with an inspiring comment from the chapter on mixed methods by Chesnut, Hitchcock, and Onwuegbuzie,

who strike a resonant chord when they explain the value of mixed methods research for addressing “wicked problems,” which are “problems involving multiple interacting systems, replete with social and institutional uncertainties, for which there is no certainty about their nature and solutions, and for which time is running out to find solutions” (Mertens et al. 2016, p. 225). Wicked problems need researchers to question their inherent power inequities and obstructions of social justice in order to frame strategies for action. This work can and must be done with methods that are as multifaceted as the wicked problems themselves.

REFERENCES

- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.
- Baker, R. S., & Corbett, A. T. (2014). Assessment of robust learning with educational data mining. *Research & Practice in Assessment, 9*, 38–50.
- Beadie, N. (2016). The federal role in education and the rise of social science research: Historical and comparative perspectives. *Review of Research in Education, 40*(1), 1–37.
- Blaisdell, B. (2016). Exorcising the racism phantasm: Racial realism in educational research. *Urban Review, 48*, 285–310.
- Borman, G. D. (2009). The use of randomized trials to inform education policy. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds). *Handbook of education policy research* (pp. 129–138). New York: Routledge.
- Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research, 103*(3), 191–207.
- Boyce, A. (1915). Present methods of measuring teacher efficiency. *Teachers College Record, 16*(7), 11–28.
- Boyce, J., & Bowers, A. J. (2016). Principal turnover: Are there different types of principals who move from or leave their schools? A latent class analysis of the 2007–2008 schools and staffing survey and the 2008–2009 principal follow-up survey. *Leadership and Policy in Schools, 15*(3), 237–272.
- Brady, H. E., & Collier, D. (2010). *Rethinking social inquiry: Diverse tools, shared standards* (2nd ed.). Lanham: Rowman & Littlefield Publishers.
- Brown, A. L., & De Lissovoy, N. (2011). Economies of racism: Grounding education policy research in the complex dialectic of race, class, and capital. *Journal of Education Policy, 26*(5), 595–619.
- Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan, 90*(8), 597–600.

- Bryk, A. S. (2015). 2014 AERA distinguished lecture: Accelerating how we learn to improve. *Educational Researcher*, 44(9), 467–477.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Callahan, R. E. (1962). *Education and the cult of efficiency*. Chicago: The University of Chicago Press.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company
- Carnoy, M. (2009). Policy research in education: The economic view. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds.). *Handbook of education policy research* (pgs. 27–38). New York: Routledge.
- Carter, D. F., & Hurtado, S. (2007). Bridging key research dilemmas: Quantitative research using a critical eye. *New Directions for Institutional Research*, 133, 25–35.
- Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge, UK: Cambridge University Press.
- Caves, K. M., Meuer, J., & Rupiatta, C. (2016). Advancing educational leadership research using qualitative comparative analysis (QCA). In B. G. Barnett, A. R. Shoho, & A. J. Bowers (Eds.), *Challenges and opportunities of educational leadership research and practice: The state of the field and its multiple futures* (pp. 147–170). Charlotte, NC: Information Age Publishing Inc.
- Coburn, C. E., Stein, M. K., & ProQuest (Firm). (2010). *Research and practice in education: Building alliances, bridging the divide*. Lanham: Rowman & Littlefield.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Co.
- Desimone, L. M. (2009). Complementary methods for policy research. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds.). *Handbook of education policy research* (pp. 163–175). New York: Routledge.
- Dixon, A. D., & Rousseau, C. K. (2005). And we are still not saved: Critical race theory in education ten years later. *Race, Ethnicity, and Education*, 8(1), 7–27.
- Donmoyer, R. (1999). The continuing quest for a knowledge base: 1976–1998. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (2nd ed., pp. 23–44). San Francisco: Jossey-Bass.
- Easton, John Q. (2010). *Out of the tower, into the schools: How new IES goals will reshape researcher roles*. Presidential Talk, Annual Meeting of the American Educational Research Association, Denver, May 2.
- Figlio, D., Karbownik, K., & Salvanes, K. (2017). The promise of administrative data in education research. *Education Finance and Policy*, 12(2), 129–136.
- Firestone, W., & Riehl, C. (Eds.) (2005). *A new agenda for research in educational leadership*. New York: Teachers College Press

- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyde.
- Foster, W. (1989). The administrator as a transformative intellectual. *Peabody Journal of Education*, 66(3), 5–18.
- Gamoran, A. (2009). Commentary: The disciplinary foundations of education policy research. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds.). *Handbook of education policy research* (pp. 106–110). New York: Routledge.
- Gobo, G. (2008). Re-conceptualizing generalization: Old issues in a new frame. In P. Alasuutari, L. Bickman, & J. Brannen. (Eds.), *The SAGE handbook of social research methods* (pp. 193–213). Los Angeles: Sage.
- Greenfield, T. B., & Ribbins, P. (1993). *Greenfield on educational administration: Towards a humane science*. New York: Routledge.
- Gronn, P. C. (1983). Talk as the work: The accomplishment of school administration. *Administrative Science Quarterly*, 28(1), 1–21.
- Hallinger, P., & Heck, R. H. (2011). Conceptual and methodological issues in studying school leadership effects as a reciprocal process. *School Effectiveness and School Improvement*, 22(2), 149–173.
- Hedges, L. V., Pigott, T. D., Polanin, J. R., Ryan, A. M., Tocci, C., & Williams, R. T. (2016). The question of school resources and student achievement: A history and reconsideration. *Review of Research in Education*, 40(1), 143–168.
- Hutt, E. L. (2016). Surveying the nation: Longitudinal surveys and the construction of national solutions to educational inequity. *Ethics and Education*, 11(2), 240–258.
- Kaplan, D. (2009). Causal inference in non-experimental educational policy research. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds.). *Handbook of education policy research* (pgs. 139–153). New York: Routledge.
- King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton: Princeton University Press.
- Labaree, D. F. (2011). The lure of statistics for educational researchers. *Educational Theory*, 61(6), 621–632.
- Levin, H.M. (1983). *Cost-effectiveness: A primer*. Beverly Hills, CA: SAGE Publications.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills: Sage Publications.
- Loeb, S., & Strunk, K. (2003). The contribution of administrative and experimental data to education policy research. *National Tax Journal*, 56(2), 415–438.
- Mertens, D. M., Bazeley, P., Bowleg, L., Fielding, N., Maxwell, J., Molina-Azorin, J. F., & Niglas, K. (2016). Expanding thinking through a kaleidoscopic look into the future: Implications of the mixed methods international research association's task force report on the future of mixed methods. *Journal of mixed methods research*, 10(3), 221–227.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications.

- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York: Cambridge University Press.
- Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford: Oxford University Press.
- Orland, M. (2009). Separate orbits: The distinctive worlds of educational research and policymaking. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds). *Handbook of education policy research* (pp. 113–128). New York: Routledge.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford Press.
- Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40(7), 331–337.
- Ragin, C. C. (2013). New directions in the logic of social inquiry. *Political Research Quarterly*, 66(1), 171–174.
- Ragin, C. C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. Oakland: University of California Press.
- Rebell, M. (2007). Professional rigor, public engagement and judicial review: A proposal for enhancing the validity of education adequacy studies. *Teachers College Record*, 109(6), 1303–1373.
- Rebell, M. A. (2008). Equal opportunity and the courts. *Phi Delta Kappan*, 89(6), 432.
- Riehl, C. (2006a). Feeling better: A comparison of medical research and education research. *Educational Researcher*, 35(5), 24–29.
- Riehl, C. (2006b). Research on educational leadership: Knowledge we need for the world we live in. In F. W. English & G. C. Furman (Eds.), *Research and educational leadership: Navigating the new National Research Council guidelines*. Lanham: Rowman and Littlefield Publishers.
- Riehl, C., & Firestone, W. (2005). What research methods should be used to study educational leadership? In W. Firestone and C. Riehl (Eds.), *A new agenda for research in educational leadership* (pgs. 156–170). New York: Teachers College Press.
- Riehl, C., & Lyon, M. A. (2017). Counting on context: Cross-sector collaborations for education and the legacy of James Coleman’s sociological vision. *The Annals of the American Academy of Political and Social Science*, 674(1), 262–280.
- Robinson, V. M. J. (1995). The identification and evaluation of power in discourse. In D. Corson (Ed.), *Discourse and power in educational organizations* (pp. 111–130). Cresskill: Hampton Press, Inc.
- Spillane, J. P., Pareja, A. S., Dorner, L., Barnes, C., May, H., Huff, J., & Camburn, E. (2010). Mixing methods in randomized controlled trials (RCTs): Validation,

- contextualization, triangulation, and control. *Educational Assessment, Evaluation and Accountability*, 22(1), 5–28.
- Stage, F. K. (2007). Answering critical questions using quantitative data. *New Directions for Institutional Research*, 133, 5–16.
- Stage, F. K., & Wells, R. S. (2014). Critical quantitative inquiry in context. *New Directions for Institutional Research*, 158, 1–6.
- Strayer, G. D. (1915). The methods of a school survey. *Teachers College Record*, 16(1), 40–44.
- Timar, T. B. (1997). The institutional role of state education departments: A historical perspective. *American Journal of Education*, 105(3), 231–260.
- Weimer, D. L. (2009). Making education research more policy-analytic. In G. Sykes, B. Schneider, D. N. Plank, with T. G. Ford (Eds.), *Handbook of education policy research* (pp. 93–100). New York: Routledge.
- Weiss, C. H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 1(3), 381–404.
- Willower, D. J., & Forsyth, P. B. (1999). A brief history of scholarship on educational administration. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (2nd ed., pp. 1–23). San Francisco: Jossey-Bass.

INDEX¹

A

Across-stage mixed methods design, 312
Action research (AR), 9, 108,
121–137, 373
Adjacency matrix, 234–237, 240
Adoption of curriculum, 194, 196
Advocacy, 69, 100, 104, 232,
239–240, 243, 244, 247, 369
Advocacy Coalition Framework
(ACF), 243–245
Agodini, Roberto, 193, 197, 205, 206
American Evaluation Association
(AEA), 327, 336
Apple, Michael W., 81, 102
Average treatment effect (ATE), 175,
176, 180–183, 187

B

Bacchi, Carol, 79
Ball, Stephen J., 80–82, 91
Bellibas, Mehmet S., 158

Bhatt, Rachana, 193, 198, 205, 206
Bias, 25, 51, 100, 104, 131, 136, 159,
176–178, 180, 181, 184, 186,
187, 214, 217, 220, 223, 225,
362, 364, 366, 373, 374
Big data, 160
Borgatti, Stephen P., 231, 232, 237,
245, 246
Bowen, Glenn A., 25, 29
Bowers, Alex J., 148, 360
Brooks, Jeffrey S., 8, 19, 87, 107, 369,
371, 373
Butler, Judith, 91, 102

C

Capper, Colleen A., 20, 107,
108, 114
Carnoy, Martin, 364
Carpenter, Bradley W., 88, 105, 107
Case-ordered meta-matrix, 8, 35,
38, 370

¹Note: Page numbers followed by 'n' refer to notes.

- Case study, vii, 8, 21, 26, 33–54,
88, 200, 201, 207, 209,
267–275, 314, 315, 343, 348,
368, 375
- Categorizational analysis, 70
- Causal analysis, 364–366, 372, 375
- Causal inference, vii, 9, 158, 159,
173–188, 364, 365, 375
- Causality, 33, 53, 362, 364, 370
- Centrality, 104, 106, 240, 242, 246
- Chudgar, Amita, 151
- Clark, M. H., 187
- Cochran, William G., 174, 180
- Coding, 34–37, 40–43, 45, 47, 51,
54, 84, 202
- Cognitive social structures (CSS),
243–244
- Coleman Report, 363, 368
- Collins, Patricia, 91, 92, 101–103,
112–114
- Communities of practice, 346–347
- Complementary methods, vi, 6, 7, 376
- Confounding factors, 174
- Conversation analysis (CA), 8, 57–59,
63, 66, 70, 371
- Core-periphery network, 240
- Counterfactual, 362, 364–366
- Covariance, 181, 187
- Creswell, John W., 12, 311, 312,
316, 317
- Critical discourse analysis (CDA),
57–62, 62n1, 67, 70–73, 83, 86,
88, 108, 109, 371, 374
- Critical policy analysis (CPA),
8, 79–93, 371, 374
- Critical quantitative methods, 375
- Critical race theory, 102, 106,
372, 374
- Critical research, 9, 71, 99–115, 374
- Cross-national comparisons, 144, 156
- Curriculum materials, 9, 193–209,
331, 371
- D**
- Daly, Alan, 232n2, 237n3, 238,
238n4, 246
- Data displays, 33–54
- Day, Christopher, 10, 369
- Deconstruction, 85–87, 102–105, 351
- Density, 246
- Denzin, Norman K., 84, 85, 92
- Design research, 376
- Dialectical pluralism, 310, 311, 318, 319
- Diffusion theories, 243, 244
- Directed graph, 234, 240
- Discourse analysis (DA), vii, 8, 57, 58,
60, 71, 85, 86
- Discursive psychology (DP), 57, 60,
63, 64, 371
- Documents, 25, 26, 29, 34–36, 38,
39, 61, 62, 67, 68, 70, 86, 88,
233, 261, 262, 266, 269, 290,
300, 325, 326, 333–335, 349,
350, 371
- Drysdale, Lawrie, 348
- E**
- Econometrics, 375
- Education policy, 3, 57–61, 64, 69,
70, 79, 84, 90, 93, 157, 193,
237, 242, 284, 294, 300,
307–321, 373, 375
- Education policy analysis, 90, 93
- Effective school leadership, 344
- Elite interviews, 24
- Embedded concurrent mixed methods
design, 314
- Enacted curriculum, 195, 196, 198,
207, 208, 371
- Epistemologies, 101, 102, 107,
371, 373
- Equal-status mixed methods
design, 316
- Estimands, 175–177, 183–184, 188

- Evaluation, vii, 10, 38–43, 47–52, 62, 65, 67, 80, 81, 83, 88, 194, 196, 200, 213–215, 218, 221, 226, 239, 254, 282, 295, 299, 300, 308–310, 312, 325–338, 365–367, 369, 371, 373, 376
- External evaluation, 330
- F**
- Fairclough, Norman, 60–62, 68, 83, 100, 108
- Feminist critical policy analysis, 89, 90
- Fernández, Erica, 82
- Formative evaluation, 331, 334
- Foucault, Michel, 60, 71, 91, 102, 125
- Framework analysis, 40, 43, 80, 81, 291
- Frameworks, 37, 41–43, 47, 50, 51, 53, 54, 68, 80–82, 84, 85, 88, 89, 91, 101, 104, 107, 109, 114, 115, 124, 144, 147, 152, 157, 158, 160, 174, 175, 195, 198, 201, 202, 244, 245, 261, 269, 284, 291, 295, 308, 310, 311, 333, 334, 346, 350, 363, 367, 368, 375
- G**
- Gooden, Mark A., 104, 106, 107
- Grissom, Jason A., 148, 156
- Guarino, Cassandra M., 9, 215, 218, 222–225, 366
- Guba, Egon G., 84, 309, 370
- Gurr, David, 10, 369
- H**
- Harris, Alma, 82, 290
- Hesse-Biber, Sharlene Nagy, 316
- High School Longitudinal Study (HSLs), 144
- Homophily, 246
- Huberman, A. Michael, 34–37, 350, 351, 370, 372
- I**
- Impact analysis, 203
- Impact evaluation, 334
- Implementation, vi, 1, 3–6, 8–10, 13, 27, 28, 34, 35, 38–40, 42, 43, 47–49, 51, 52, 80–83, 87–89, 93, 156, 159, 173, 174, 177, 179, 185, 194–203, 205–209, 232, 236, 239, 240, 242, 243, 246, 247, 256, 281–284, 286, 290–296, 300, 301, 315, 325, 330–332, 337, 368, 371, 372, 376
- Implementation fidelity evaluation, 332
- Improvement science, 376
- Inductive analysis, 351
- Institute of Education Sciences (IES), 333, 363, 365
- Institutionalized power, 104
- Intended curriculum, 194, 195, 198, 207
- Intensive explication, 86
- Internal evaluation, 330
- International Successful School Principalship Project (ISSPP), 342–347, 350, 352, 353
- Interpretive research, 127, 373, 374
- Inverse propensity weighting, 180–181
- Ivankova, Nataliya V., 314
- J**
- Jacobson, Stephen L., 345
- Johnson, R. Burke, 308, 310–312, 315, 316, 319
- Justice, 8, 82, 89, 100, 104–107, 109, 125, 137, 255, 308, 310, 372, 377

K

Kerr, David, 309
 Koedel, Cory, 193, 198, 203–206, 226

L

Language use, 59, 104, 110, 371
 Large scale datasets, 156
 Large-scale survey research, 200, 201
 Leadership, v–vii, 1–13, 19–29, 33, 34, 36, 43, 45, 47, 48, 53, 57–74, 88–90, 99–115, 132–134, 143–161, 183–185, 198, 206, 208, 238, 256, 290–299, 310, 311, 329, 341–355, 359–377
 Leadership networks, 238, 246
 Leadership preparation programs, 107
 LeCompte, Margaret D., 35
 Lee, Jaekyung, 148, 154, 158, 175
 Legitimation, 308, 319, 320, 371
 Leithwood, Kenneth, v, 1, 344, 345
 Lester, Jessica N., 8, 11, 371, 374
 Levin-Rozalis, Miri, 326
 Lincoln, Yvonna S., 84, 85, 309, 370
 Lochmiller, Chad R., 20, 59, 60
 Logic model, 37, 42, 300, 328, 329, 335, 337, 362, 363, 366, 367
 López, Gerardo R., 82, 104, 108, 114
 Luft, Julie A., 314

M

Marginalized communities, 104, 105
 Marsh, Herbert W., 158
 Matrix analysis, 42, 43, 48, 51, 370
 Membership categorization analysis (MCA), 57, 60
 Merriam, Sharan B., 21, 24, 26, 33, 34, 200
 Mertens, Donna M., 308, 310, 311, 377
 Milani, Michelle, 82, 89

Miles, Matthew. B., 34–37, 350, 351, 370, 372
 Mixed methods, v–vii, 6, 10, 307–321, 341, 360, 371, 376, 377
 Moos, Lejf, 345
 Morgan, David L., 311, 364
 Mulford, Bill, 344, 345, 348
 Multiple case study, 33–54
 Multiple-perspective methods, 350
 Multi-site case study, vii, 315, 348
 Multivariate matching, 177, 178, 187

N

National Assessment of Educational Progress (NAEP), 144, 157
 Naturally occurring data, 63, 65
 Networks, 10, 37, 82, 107, 111, 178, 231–247, 341–355, 369, 370, 372, 376
 NVivo, 39, 41–43, 45, 47, 49, 51, 54

O

Objectivity, 38, 71, 101, 103, 104
 Observations, 23–26, 29, 34–36, 38–40, 52, 86, 87, 173–188, 232n2, 326, 330, 333–335, 350
 One-mode network, 234, 236
 Open online educational resources, 194, 197
 Opportunity to learn (OTL), 195, 196, 198, 200–202, 208, 371

P

Peer networks, 237, 238
 Policy analysis, 4, 8, 10, 70, 79–93, 308–311, 371, 374
 Policy lever, 195, 197
 Porter, Andrew C., 154, 195, 201, 202
 Post-positivism, 91

Potter, Jonathan, 58, 62, 63, 73
 Power dynamics, 22, 24, 82, 89, 90, 128
 Pragmatism, 310
 Problemizing, 81, 85, 108
 Program evaluation, vii, 10, 213,
 214, 221, 226, 254, 325–338,
 367, 369
 Program evaluation standards, 327,
 328, 332–334, 336, 337
 Programme for International Student
 Assessment (PISA), 144, 147,
 157, 160
 Propensity score matching, 159, 364

Q

Qualitatively-driven, 316
 Quantitative analysis, 143, 209, 314,
 366, 368
 Quantitatively-driven, 316
 Queer theory, 85, 86

R

Race, 51, 60, 81, 84, 91, 102, 104,
 105, 107, 112, 114, 123, 129,
 148, 152, 154, 187, 362,
 372, 374
 Racism, 61, 91, 102, 105, 106, 123,
 127, 129, 130, 132, 136
 Random experiments, 173, 362–364
 Randomized controlled trials (RCTs),
 362, 363, 365
 Reciprocity, 47, 246, 352
 Reliability, 26, 186, 350, 354
 Research design, 7, 10, 12, 21–23, 26,
 29, 71, 82, 84, 85, 92, 99, 126,
 127, 260, 307, 308, 311–317,
 320, 321, 330, 361, 370
 Research framework, 144, 147, 157
 Research paradigm, 79–93, 311,
 341, 361

Riehl, Carolyn J., 5, 11, 344, 361,
 363, 367
 Rigor, 21, 26–27, 34, 37, 38, 50, 52,
 54, 101, 131, 201, 342, 346,
 347, 365, 372
 Robinson, Viviane, 345
 Rodriguez, Deborah, 255
 Rosenbaum, Paul R., 174, 176,
 178, 185
 Rubin, Donald B., 174, 176, 178,
 184, 185

S

Saldana, Johnny, 35, 37
 Sampling, 52, 53, 157, 159, 160, 180,
 315, 336, 348
 Schensul, Jean J, 35
 Schmidt, William, 154, 196
 School and classroom contexts,
 150–152, 154
 School leadership, 9, 62, 99–115, 147,
 150, 157, 329, 341–355
 Schools and Staffing Survey (SASS),
 144, 150, 157
 Secondary data analysis, vii, 9,
 143–161, 203, 368
 Selection bias, 362, 364
 Sequential analysis, 70
 Sequential-explanatory mixed methods
 design, 313
 Shadish, William R., 159, 187
 Silins, Halia, 344, 345
 Smith, Linda T., 100–104, 112, 114
 Smithson, John L., 195
 Social capital, 238, 239, 243, 369
 Social networks, vii, 10, 231–247,
 369, 370
 Spencer, Liz, 37
 Stable Unit Treatment Value
 Assumption (SUTVA),
 176, 177, 183

Stake, Robert, 21, 34, 200
 Stokoe, Elizabeth, 63, 66
 Strategic transformation, 104
 Stratification, 82, 85, 187
 Stuart, Elizabeth A., 179, 180
 Successful school leadership, 341–355
 Summative evaluation, 331, 332, 334
 Supplemental materials, 194, 207
 Support, vi, 3, 5, 12, 13, 36, 40, 47,
 82, 89, 90, 107, 109–111, 144,
 147, 154, 196, 206–208, 237,
 245, 269, 281, 291, 294,
 296–299, 329–333, 335–337,
 341, 342, 346, 349, 361, 365,
 369, 373
 Survey of Enacted Curriculum (SEC)
 Framework, 201, 202, 262
 Systems of power, 100, 112

T

Teaching and Learning International
 Survey (TALIS), 144, 147, 157
 Thematic analysis, 36, 37
 Theory of action, 42, 43, 328, 329,
 335, 337, 367
 Thoemmes, Felix J., 180, 182
 Traditional policy analysis (TPA), 4,
 80, 81, 84, 374
 Transcription, 36, 40, 62, 66, 73
 Transferability, 27
 Transformative paradigm, 310, 311
 Transparency, 25, 27

Trends assessment, 154
 Trends in International Mathematics
 and Science Study (TIMSS),
 144, 157, 160
 Trustworthiness, 26, 352, 354
 Two-mode network, 236

U

Undirected graph, 234

V

Validity, 34–37, 52, 91, 103, 108,
 131, 135, 206, 276, 308, 319,
 320, 350, 354, 371
 Value-added models, 5, 9, 194, 205,
 213–220, 222–226, 366
 van Dijk, Teun A., 60, 61, 74

W

Weiss, Carole, 325, 336, 376
 Welton, Anjale D., 82
 What Works Clearing House
 (WWC), 205, 333, 363
 Winton, Sue, 82, 89
 Wolcott, Harry F., 20, 21, 25, 26

Y

Yin, Robert K., 21, 33, 34, 52, 53
 Ylimaki, Rose, 345