

PISA Under Examination

Changing Knowledge, Changing Tests, and Changing Schools

Miguel A. Pereyra, Hans-Georg Kotthoff
and Robert Cowen (Eds.)

SensePublishers



PISA Under Examination



CESE Comparative Education Society in Europe
Association d'Éducation Comparée en Europe
Gesellschaft für Vergleichende Erziehungswissenschaft in Europa

The Comparative Education Society in Europe (CESE) is an international non-profit making association of scientific and educational character. CESE was founded in 1961 in London and is a founding society of the World Council of Comparative Education Societies (WCCES).

CESE has traditionally promoted a space for dialogue amongst scholars, specialists and young researchers from the field of education and other disciplines. More specifically, its purpose is to encourage and promote comparative and international studies in education by:

- promoting and improving the teaching of comparative education in institutions of higher learning;
- stimulating research;
- facilitating the publication and distribution of comparative studies in education;
- interesting professors and teachers of other disciplines in the comparative and international dimension of their work;
- co-operating with those who in other disciplines attempt to interpret educational developments in a broad context;
- organising conferences and meetings;
- collaborating with other Comparative Education Societies across the world in order to further international action in this field.

Every two years CESE organises an international conference of high scholarly standards which attracts academics, scholars, practitioners and students from all parts of Europe and around the world. Throughout its history, CESE has organised twenty-four such conferences, a special conference for the 25th anniversary of the Society, a symposium, and two 'CESE In-Betweens'. In-Betweens are international symposia organised between the biennial conferences. A web site of CESE is maintained at <http://www.cese-europe.org/>

The CESE Series

Series Editor: Miguel A. Pereyra, CESE President

Editorial Board: The CESE Executive Committee

Robert Cowen (Institute of Education, University of London)

Elisabeth Buk-Berge (Norwegian Ministry of Education & Research)

Hans-Georg Kotthoff (Pädagogische Hochschule Freiburg)

Vlatka Domović (University of Zagreb)

Lennart Wikander (University of Uppsala)

Eleftherios Klerides (European University of Cyprus)

PISA UNDER EXAMINATION

Changing Knowledge, Changing Tests, and Changing Schools

Edited by

Miguel A. Pereyra

Hans-Georg Kotthoff

Robert Cowen



SENSE PUBLISHERS
ROTTERDAM/BOSTON/TAIPEI

A C.I.P. record for this book is available from the Library of Congress.

ISBN: 978-94-6091-738-7 (paperback)

ISBN: 978-94-6091-739-4 (hardback)

ISBN: 978-94-6091-740-0 (e-book)

Published by: Sense Publishers,
P.O. Box 21858,
3001 AW Rotterdam,
The Netherlands
www.sensepublishers.com

Printed on acid-free paper

Photo cover: The *Gran Telescopio CANARIAS*, one of the largest telescopes in the world located in one of the top astronomical sites in the Northern Hemisphere: the *Observatorio del Roque de los Muchachos* (Observatory of the Boys' Rock) in the island of La Palma in the Canaries.

All Rights Reserved © 2011 Sense Publishers

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

This book is dedicated to Germán González (1940-2011), a great schoolteacher who deeply cared for the cultural improvement of his island of La Palma and its people and could not see this book printed unfortunately

CONTENTS

Preface and acknowledgments	xi
1. PISA under examination: Changing knowledge, changing tests, and changing schools <i>Miguel A. Pereyra, Hans-Georg Kotthoff and Robert Cowen</i>	1
Section I: The comparative challenges of the OECD PISA programme	15
2. PISA as a political instrument: One history behind the formulating of the PISA programme <i>Ulf P. Lundgren</i>	17
3. PISA: Numbers, standardizing conduct, and the alchemy of school subjects <i>Thomas Popkewitz</i>	31
4. Constructing the OECD programme for international student assessment <i>Clara Morgan</i>	47
5. The dissatisfaction of the losers: PISA public discourse in Ibero-American countries <i>Antonio Bolívar</i>	61
Section II: PISA and school knowledge	75
6. The context for interpreting PISA results in the USA: Negativism, chauvinism, misunderstanding, and the potential to distort the educational systems of nations <i>David C. Berliner</i>	77
7. PISA, international comparisons, epistemic paradoxes <i>David Scott</i>	97
8. Competencies vs. interculturality. Student exchanges in the age of PISA <i>Donatella Palomba and Anselmo R. Paolone</i>	109

TABLE OF CONTENTS

Section III: The assessment of PISA, school effectiveness and the socio-cultural dimension	123
9. The introduction of state-wide exit examinations: Empirical effects on math and english teaching in german academically oriented secondary schools <i>Katharina Maag Merki</i>	125
10. The PISA girls and ticking the boxes an examination of students' perspectives on PISA testing <i>Gerry Mac Ruairc</i>	143
11. From the appealing power of PISA data to the delusions of benchmarking: Does that challenge any evaluation of educational systems? <i>Marie Duru-Bellat</i>	157
12. Are you on the educational production frontier? Some economic insights on efficiency from PISA <i>Javier Salinas Jiménez and Daniel Santín González</i>	169
Section IV: PISA and the immigrant student question	183
13. PISA's potential for analyses of immigrant students' educational success: The German case <i>Aileen Edele and Petra Stanat</i>	185
14. Why do the results of immigrant students depend so much on their country of origin and so little on their country of destination? <i>Julio Carabaña</i>	207
Section V: Extreme visions of PISA: Germany and Finland	223
15. Education politics and contingency: Belief, status and trust behind the Finnish PISA miracle <i>Hannu Simola and Risto Rinne</i>	225
16. Concepts, cultures and comparisons. PISA and the double German discontentment <i>Daniel Tröhler</i>	245
17. CODA <i>Robert Cowen</i>	259

TABLE OF CONTENTS

Annexes	265
Annex I	267
Visualizing PISA scientific literature versus PISA public usage <i>Antonio Luzón and Mónica Torres</i>	269
Annex II The Poster Exhibition	303
Annex III	319
PISA a examen, cambiar el conocimiento, cambiar las pruebas y cambiar las escuelas <i>Jesús Romero, Antonio Luzón and Mónica Torres</i>	321
Notes on contributors	327

PREFACE AND ACKNOWLEDGMENTS

The *Comparative Education Society in Europe* (CESE) today is the oldest European scientific society in the field of academic education in our continent. Since its creation in London in 1961, CESE has traditionally promoted a space for scientific dialogue amongst scholars, specialists and young researchers from the field of education and other disciplines.

Throughout its history, CESE has organized twenty-four conferences and two specialized symposia. At present CESE is continuing this tradition and from time to time seeks to organize between the biannual conferences an international symposium with the title “CESE in-between”. Our main goal is to invite leading scholars and experts both within and outside Europe to engage in independent and intellectually balanced conversations about urgent and contemporary educational *problématiques*.

From the 23rd to 26th of November 2009, with the important sponsorship of the *Cabildo Insular of La Palma* (Board of Towns of the Island of La Palma), the *Consejería de Educación, Universidades, Cultura y Deportes* (Regional Ministry of Education, Universities, Culture & Sports) of the Government of the Canaries, the Spanish *Ministerio de Ciencia e Innovación* (Ministry of Science & Innovation), and *Rayas* (Museum and Archive of History of Education of La Palma), CESE organized an international symposium entitled *PISA under Examination: Changing Knowledge, Changing Tests, and Changing Schools*. The subject of PISA was chosen because of its widespread interest to academics and policy-makers and working educationists as well as parents and local communities. Across Europe, there is exceptionally strong interest in this topic on both practical and theoretical levels.

For four days, seventeen leading scholars presented their contributions in the symposium, and 105 delegates from Spain, Europe and America (mainly from Latin America) met in the *Teatro Chico* (Small Theatre), a historical place built, on the remains of a 16th century church, by the liberal freemason bourgeoisie of La Palma in the 19th century. It was indeed a memorable event, as memorable as this island of the Canaries, designated by UNESCO as one of the “biosphere reserves” of the world – a place where the hybridising of European and Latin American cultures is unique.

I would like sincerely to thank those whose generous participation made it possible for the international symposium on *PISA under Examination* to be organized by CESE in La Palma, as well as the economic support granted by the *Cabildo de La Palma*, the *Consejería de Educación, Universidades, Cultura y Deportes* of the Government of the Canary Islands, and the Spanish Ministry of Science & Innovation. My colleagues and friends on the Executive Committee of CESE and in particular the Immediate Past President of CESE, Bob Cowen, were very stimulating and supportive all through the process of organising this international symposium. Without the participation of the distinguished keynote speakers we invited, this event would not exist; and, in fact, in most cases all our

PREFACE AND ANCKNOWLEDGEMENTS

invitations to come to La Palma for presenting and discussing ideas about PISA and its most relevant *problématiques* were accepted. The support of the President of the *Cabildo de La Palma*, Ms Guadalupe González Taño, was from the beginning essential, and my dear old friend Germán González, director of *Rayas* (Archive and Museum of History of Education of La Palma), also contributed decisively to the full accomplishment of the event. Unfortunately Germán has been unable to see this book printed since he very regrettably passed away last March.

Eliás Bienes and Javier Jerónimo from *Nuevo Rumbo-Historia Viva* were in charge of the organization of the symposium in La Palma, and the inspiration and qualities of their organization helped us to be both creative and efficient. My colleagues from the University of Granada Antonio Luzón and Mónica Torres, who were the secretaries of this international symposium, performed their work wonderfully well, and wrote a solid report on PISA which was included in the booklet of the symposium, printed by Gustavo Gómez and creatively designed by María Torres (retrievable at <http://www.cese-europe.org/conferences/45-i-cese-in-between-las-palmas-2009/324-pisa-booklet>). Antonio's help was very important in the heavy and time-consuming process of preparing the 'camera-ready' manuscript of this book. Rocio Lorente prepared efficiently the Index Name of this book.

To all of them I want to express my most sincere acknowledgment.

Miguel A. Pereyra
President of CESE and coordinator of the international symposium
PISA under Examination

MIGUEL A. PEREYRA, HANS-GEORG KOTTHOFF
AND ROBERT COWEN

PISA UNDER EXAMINATION

Changing Knowledge, Changing Tests, and Changing Schools

AN INTRODUCTION TO THE COMPARATIVE PUZZLE

PISA or the Programme for International Student Assessment of OECD (Organisation for Economic Co-operation and Development) is one of the most famous educational events of the last decades. Thousands of students from sixty-two different countries (the OECD countries plus country partners which signed a contract with this institution) have been recently involved in its tests for the 2009 PISA (the fourth report of this kind was presented in December 2010) (see Fig. 1).

A map of PISA countries and economies



Source: OECD—The PISA website: <http://www.pisa-oecd.org>

Overall, PISA has been a remarkable phenomenon. Rarely has educational information translated so fast into the word 'disaster' – and domestic political crisis. Rarely has educational information translated so fast into the word 'stardom' –, and sudden international attention being given to countries which hitherto were

un-noted and uncelebrated. PISA was not merely been an educational event. It was also a media circus. It involved the public rehearsal of reasons for failure or success; and even, in some cases, public and political and academic explanations about why 'failure' was not really that, and why 'success' was not really that either.

At the centre of all these indications, we find the growing influence of international agencies on education and schooling which has decisively contributed to a marketisation of the field of education in the context of an increasingly multilevel and fragmented arena for educational governance (Jones, 2007a and 2007b; Henry et al., 2001; Martens, Rusconi & Leuze, 2007; Mundy, 2007, and Moutsios, 2009). In recent years, their influence has not been limited to a particular geographical area or specific area of education and schooling, but has become a generalized phenomenon giving rise to an increasing internationalization of education. In fact, the “cognitive horizon” of these international agencies, such as the OECD, reaches beyond traditional borders and national and regional identities of its member countries, as shown by the universally applicable models to inform ‘best practices’ to achieve more efficient education and schooling. In this context, the “cognitive horizon” assumes a linear administrative chain of steering of our educational systems, which runs from the political level via the political body of school owner without considering any model able to explain the complexity of the relation of the different levels of the educational system; on the contrary, what is mainly considered is the instructional setting organised within each school to individual learning (see Landgeldt, 2007, p. 236). An additional distinctive feature of this “cognitive horizon” is its goal of generating policy-based *regulatory competition* on objective criteria, scientifically researched with more or less sophistication and presented in an easily accessible manner (through the use of tools useful for trying to solve various problems and issues, as PISA seems to do precisely up to the point of becoming at the present a kind of ‘soft power’ in education, as recently stated by Bieber & Martens, 2011). (See also on PISA, and the OECD, Rizvi & Lingard, 2009; Knodel et al., 2010, and Knill & Tosun, 2011, on the dynamics of these policy mechanisms.)

International agencies are becoming, therefore, independent agents in the field of education, rather than simply providing advice for their Member States which had originally been their responsibility, and their influence is today very notorious across the different fields of policy by generating a standardization by harmonization of educational systems, increasingly clear in Europeanisation educational processes (Lawn, 2011). Actually they are generating “soft mechanisms” for the formulation, the regulation and the transnational coordination and convergence of policies, buttressed by the diffusion of persuasive discursive practices which promote isomorphic policy emulation processes subject to rapid institutional imitation in today’s globalizing world (Meyer & Rowan, 1983 and DiMaggio, 1983).

Following these patterns, the OECD as one of the leading international organizations has been ending to reach greater recognition lately. Since the last past decades of last century, and in particular since the 90s, has consolidated a steady ‘comparative turn’ in its education policy by introducing a framework of *governance by comparison* which emphasizing the interplay between the interplay between the

actors (the OECD governing body and its member states) and resulting policy (Martens, 2007, p. 54).

In this context the first comparative puzzle which attaches to PISA is: why all the fuss? What are the politics and sociology and anthropology of the international testing movement as if 'educational results' were a sporting event?

The second comparative puzzle which attaches to PISA is: in what sense is it 'comparative education'? At what point do numbers become or represent or stand for cultures, and what needs to be explained about the cultures/numbers symbiosis? What kind of comparative education does PISA signify? A comparative education of measured outcomes? Outcomes of what and from what, in the broader social and historical context?

The third comparative puzzle which attaches to PISA is: in what sense is it good 'big sociology'? What is – sociologically, in the workings of schooling systems – being tested?

The fourth comparative puzzle which attaches to PISA is: in what ways is this good empirical work? Which technical criteria does this kind of 'comparative work on an international scale have to satisfy and in what senses may we (technically) believe in the numbers?

The fifth comparative puzzle, noting the style and extent to which we 'believe' in those numbers, is whether, by whom, and with what consequences may we deduce policy action from such research? Is this the 'robust and relevant research' of which politicians dream? Can we move from these research results to policy action quickly, cautiously, or not at all?

All these questions and others were approached during the debates of our international symposium and accordingly they are addressed in the contributions of this book. We have organised the content of the book on the pattern we followed in the symposium.

In the first part of this book entitled "The comparative challenges of the OECD PISA programme, the authors contextualize and situate the OECD PISA programme within the broader social and historical context of the development of international comparative student assessment. PISA is viewed and analysed from a variety of angles and disciplines, including historical, political, administrative, economic, educational, cultural, governance and comparative perspectives. However, while the authors in this first part analyse the same phenomenon from a wide range of very different analytical and theoretical perspectives, they all share one common assumption: they regard PISA as a form of international and transnational governance and as a disciplinary technology, which aims to govern education in the 21st century.

In his contribution **Ulf Lundgren**, who was himself professionally involved in the development of the PISA programme, provides the reader with a detailed and intimate history of the formulation of the PISA programme. The main focus of his historical analysis is on the development of international assessment as a device for political governing. Thus, Lundgren analyses the economic and political context which formed the background for international assessment in general and PISA in particular. Lundgren traces the genesis of educational assessment back to the 19th century and identifies the early decades of the 20th century as a first milestone in the

development of educational assessment, when the idea of educational evaluation as a base for educational reforms was established and educational assessment was increasingly linked to social positions and salaries. According to Lundgren, the 1950s and 1960s mark a further milestone in the development of international assessment because it became comparative. International comparative assessment seemed to be particularly useful in a political Cold War climate that had an interest for the comparison of competitive education systems. The establishment of important agencies for comparative educational assessment such as the IEA, which followed quickly after the Sputnik shock in 1957, had a major impact on the further development of international assessment in that it drew the politicians' attention to the possibility of governing education by goals and results, i.e. measured outcomes. According to Lundgren, it became obvious in the 1980s that earlier planning models in education had failed and that new ways of political governing of the education system had to be developed, which required new and more specific goals: "To govern education by expressing goals to be achieved and evaluating the achievements demanded new conditions for governing. To be a steering device, goals have to be clear" (p. 23). Against this background the PISA programme was launched in the 1990s. It became particularly successful because it coincided with global changes in the 1990s which led to a global knowledge society in which education has become an international commodity. According to Lundgren's analysis, it is this particular *Zeitgeist*, which is characterised by the competition between new emerging knowledge societies that is not only restricted to natural but also to intellectual resources, which explains the PISA effect to a large degree.

In the second paper **Thomas Popkewitz** aims to analyse and to understand the system of reason through which OECD's PISA technologies and classifications are made intelligible. In order to do this Popkewitz examines firstly historically how "the numbers of PISA can be seen as 'facts' and as a way of 'telling the truth' about society, schooling, and children" (p. 33). Following Popkewitz's argument, PISA's narratives are built on the premise that numbers tell the 'truth' about national schooling systems and children. However, numbers as categories of equivalence are not merely numbers. Measurements provide constant performance indicators in a continual process of locating one's self in the world that are analogous to global positioning systems: "PISA globally positions the child and nation through a style of thought that differentiates and divides through creating categories of equivalence among countries" (p. 36). In the second part of his paper Popkewitz turns to the principles of school subjects and investigates how disciplinary knowledge is translated into school subjects. He argues that the "practical knowledge" measured by PISA has very little to do with the disciplinary knowledge. The translation of disciplinary knowledge into school subjects is rather an attempt to govern conduct through the insertion of particular rules and standards or even moral qualities about modes of living. PISA's assessment of students' knowledge and skills, then, can not only be seen as measurements about what "practical" knowledge children know. PISA also has strong normative function in that it tells us, who the child is and who or what it should be in future, i.e. a 'self-motivated lifelong learner' who is to live in the 'knowledge society'.

Clara Morgan analyses the construction of the PISA programme from a multi-disciplinary perspective which draws on political economy and international relations as well as sociology. Seen from a political economic perspective, Morgan situates the construction of PISA in the broader political rationality of neo-liberalism. As the role of education in the 1980s and 1990s was increasingly viewed in neo-liberal, i.e. instrumental terms (e.g. to reduce unemployment rates etc.), the OECD educational activities became increasingly concerned with the development of a competitive and highly skilled labour force: “Under neoliberalism, OECD education policy focused on implementing accountability and performance measures, improving educational quality and monitoring of educational systems” (p. 49). The formulation of PISA fitted into this context, in that it defined measurable outcomes which are required for competitive accountability. Analysing PISA from the international relations perspective, Morgan comes to the conclusion that there has been a strong American influence on the formulation of the PISA programme and, more generally, on the governance of international organisations, including the OECD. Finally, Morgan draws on theories from sociology and from Michel Foucault’s conceptualization of the ‘power bloc formation’ to understand how PISA ‘works’ and how it is used to exercise power. From this analytical perspective PISA reflects a ‘power bloc formation’ that works because it “serves the needs”, as Morgan puts it, “of politicians, policymakers and international and regional organisations as an accountability engine for governing education in the 21st century” (p. 56).

In the final paper of the first part **Antonio Bolívar** takes a very different perspective on the PISA public discourse by analysing the PISA results from the perspective of the “losers”. These are, according to Bolívar, the Ibero-American countries, who feel discontented and dissatisfied with their PISA results, which do not correspond to the desires and expectations of their societies. On the basis of several empirical studies on the PISA media discourse in Spain and Latin America, Bolívar argues that the PISA reports have been presented with a certain degree of sensationalism, with a lack of rational analysis and simplifications or even manipulations of the data. In summary, Bolívar argues, “each of the PISA reports has been received from a political and ideological duality, serving the educational policy that interested each ideological group, and producing ideological manipulations of the results” (p. 62). Bolívar’s own analysis of the performance of the Ibero-American countries in PISA 2006 in the second part of the paper reveals that the Ibero-American average score in science (426) is far from the OECD average (500) and that this performance is even overestimated because those youths who do not take PISA tests often do not attend schools and would thus lower the scores even further. On the whole, Latin America obtains lower results than the countries in Europe and Asia and the Latin American countries present more unequal distribution. Obviously, the educational reforms that were taken in the last few decades have not had a decisive effect on the quality of teaching in the classroom. According to Bolívar, one reason for this is that rather than extracting lessons from the results and ‘rationalizing’ educational policies, the data have been

instrumentalized and used to justify the changes already made or to provide support for educational policies already in place.

The second part of this volume, which is devoted to the theme of ‘PISA and School Knowledge’, takes the PISA discourse closer to schools and schooling. The contributors analyse and discuss the impact of PISA on school knowledge and the school curriculum in particular. Thus, their analyses focus on questions like what kind of knowledge is tested through PISA, how the achievement in PISA is related to knowledge acquired at schools and in which respect PISA challenges and shapes definitions of school knowledge and definitions of competencies.

David Berliner focuses in his paper on PISA’s potential to distort national educational systems in general and school curricula in particular. He starts off by analysing the interpretive context for the publication of the PISA 2006 scores, which were greeted in the USA with negativism, exaggerated fears about the allegedly poor performance of the US American schools and chauvinism. Berliner’s own analysis of the PISA 2006 scores reveals first of all that the US American schools and pupils are far better than depicted in the media. However, the PISA results also reveal that there is a huge problem with inequality and inequity in the US American school system. According to Berliner’s analyses, this problem is not primarily caused by the school system, but rather by a number of out-of-school factors such as gross domestic product per capita and the huge inequality in wealth within the nation. In this respect PISA scores do not merely represent schools and schooling, but, according to Berliner, “schools and society *in interaction*” (p. 83). Thus, the PISA scores are a powerful indicator of the USA’s uneven income distribution and housing segregation and of the effects of social class on school achievement. In the final part of his paper Berliner looks at the consequences that could arise, if PISA tests became high-stakes tests. On the basis of numerous empirical studies in the USA and the UK, which analysed the effects of high stakes testing, Berliner shows that high stakes assessment systems can corrupt teachers as well as the indicator, has a narrowing influence on the school curriculum, (e.g. more time for maths and reading in the curriculum, marginalisation of art and music etc.) and has a standardizing influence on the teaching methods. If PISA became a high-stakes assessment system, it would probably result in an international standardization of school curricula and a narrowing of the skill set that pupils and students possess, which is, according to Berliner, exactly the opposite of what is needed in the 21st century.

In the second contribution of this part **David Scott** takes a closer look at the forms of knowledge that are tested in PISA from a critical realist perspective. After his initial and fundamental differentiation between two forms of knowledge, i.e. knowledge (a), that represents knowledge sets, skills and dispositional states of a person, and knowledge (b), which represents knowledge sets, skills and dispositional states which allow this person to do well in tests, Scott unmasks false beliefs or assumptions about the characteristic features of these two forms of knowledge and about the problematic relationship between knowledge and its assessment (i.e. marker error, cultural bias effect, epistemic differences etc.). According to Scott’s analysis, the relationship between knowledge and its assessment is further

complicated or even aggravated by various ‘examination technologies’ such as whether an incentive is attached to the taking of the test, the students’ motivation to take the test and the test format (i.e. multiple choice or free-ranging essay formats), which might favour some groups in comparison with others. International comparative student assessments (like PISA) face the additional difficulty of trying to construct curriculum-free tests underpinning the idea of a universal form of knowledge. PISA tests are therefore, according to Scott’s analysis, not related to national school curricula and they are consequently not a measure of what the student have been taught or what they have learnt in any formal sense, which means that the test are likely to favour some countries at the expense of others. According to Scott, the notion of a universal form of knowledge makes a number of reductionist assumptions and does not account properly for cultural differences which might affect test performance in several ways. By doing this PISA also operates as a standardizing device (i.e. it creates a norm) by stressing certain forms of performative knowledge which are becoming the norm. The final criticism is directed at the way PISA results are published in comparative national tables thereby putting emphasis on position rather than score. According to Scott, such league tables do not provide countries with very useful information for the improvement of their education system, but rather contribute to the nation’s (negative) self image.

In the final paper by **Donatella Palomba** and **Anselmo R. Paolone** the theme of PISA’s relation to school knowledge is analysed from a very specific angle. In their case study, Palomba & Paolone focus on the question of teachers’ attitudes towards long-term students’ exchanges at secondary schools. The case studies were conducted in several Italian secondary schools which are involved in year-long individual student exchange programmes. The research consisted of qualitative fieldwork based on participant observation, semi-structured interview, recorded ‘open discussions’ and the study of available school documents (e.g. school profile, etc.). In two schools the teachers drew in some ways on PISA tests in order to assess the acquired competencies of their returning students. These two schools, which are reported in this paper, approached PISA in two completely different ways and integrated aspects of PISA in their own culture, translating and transforming these elements, according to their local tradition, previous experiences and actual needs. The results of the two case studies show that the “familiarity with PISA makes the teachers less mistrustful towards what returnees have studied and learned abroad”. In their discussion of these findings the authors stress that PISA’s concentration on competences (rather than knowledge) which are spread internationally, school experiences in Italy and elsewhere are getting more similar. As result Italian teachers tend to think what returnees have learned abroad is not inconsistent with what is being taught at home. So while the intercultural experience of the students, i.e. their feeling of ‘otherness’ is probably reduced, the acceptance of the Italian teachers of their pupils’ competencies acquired abroad is stronger, because the competencies are deemed to be universal. Within the “intercultural exchanges”, these effects can be seen as facilitating an international dialogue, but also as a cultural homologation.

The third part of this volume entitled “The assessment of PISA, School Effectiveness and the Socio-cultural Dimension” focuses on the assessment of PISA and the question if and in which ways we can deduce policy action from this kind of research for educational policy, school improvement and school efficiency. The authors analyse the assessment of PISA on different levels and from different disciplines, including, in the last paper, the economics of education perspective. While most papers discuss the possible consequences of the PISA results on the systems level, one paper focuses on the student’s perspective by asking how individual socio-economically disadvantaged students react to PISA tests and engage in the process of testing.

In the first paper **Katharina Maag Merki** examines the effects of external achievement tests on teaching quality. Since changes in the teaching quality which are the result of the participation in international comparative achievement studies cannot be investigated in the framework of the PISA studies, Maag Merki analyses to what extent external state-wide exit examinations have an effect on the teaching quality in maths and English in the final year of upper secondary education in the German *Gymnasium*. Following the below average performance of the German education system in earlier PISA studies (2000, 2003, 2006), all 16 states introduced state-wide *Abitur* exit examinations unless they had not already instituted them earlier (e.g. Bavaria, Baden-Württemberg). In her longitudinal empirical study Maag Merki focuses on two German states: Bremen which introduced state-wide *Abitur* exit examinations in some advanced-level courses (e.g. English and Maths) in 2008 and the German state of Hesse, in which state-wide exit examinations have been introduced in all subjects in 2007. Comparing the teaching quality before and after the introduction of state-wide exit examinations led, according to Maag Merki, to the following results: “the introduction of state-wide *Abitur* exit examination in advanced English and maths courses in Bremen was accompanied by an improvement in instructional quality in those courses” (p. 131). These positive effects on instructional quality remain stable over time and can be found again in 2009. In contrast to Anglo-American empirical findings on the question of the impact of external achievements tests on the teaching quality, negative consequences could not be observed at this early stage. The main reason for this difference could be, according to Maag Merki, that the German *Abitur* exit examinations must be characterized – in international comparison – as low-stakes assessment, which “allow teachers more room to employ functional approaches that can be tailored to students’ needs” (p. 132).

The second paper by **Gerry MacRuairc** moves from the macro-level analysis to the students’ perspective on PISA testing. While the high level of correlation between educational attainment and the socio-economic background of the students is empirically well established, Mac Ruairc wants to analyse in his study how individual socio-economically disadvantaged students react to tests and engage in the process of testing. In order to do that, “it is”, according to Mac Ruairc, “important to take on board the perspectives of individual students themselves” (p. 135). By examining the views of students on PISA testing in one case study the author provides an insight into how the PISA assessment (2009) was experienced

by a group of working-class girls in a disadvantaged inner city school in a large urban area in the Republic of Ireland. The study comprised a visit to the school on the day following the administration of the 2009 PISA test and included focus group interviews with three groups of students and the principal. The thematic analysis of the interviews and the focus groups transcripts revealed three themes: (1) the intensity of the testing process was too high and most students, especially those with special educational needs, felt overstretched by the amount and the content and difficulty of the reading test items. (2) children who simply ticked the boxes to complete the test in time have implications for the validity of some of the responses to test items (3) students complained about too many personal questions and a lack of anonymity in the student questionnaire, which was to collect data in relation to a number of background variables including family and home circumstances. In his conclusion Mac Ruairc highlights the need for a more proactive approach to student support and a more nuanced model of assessment in future PISA tests to take account of social class difference.

Marie Duru-Bellat analyses in her contribution the ability of PISA data in assessing the quality of education systems. The author starts off by discussing the question why PISA data are so appealing for policy-makers despite their limitations. In her analysis Duru-Bellat points out that PISA data are so attractive because, rather than assessing conformity to academic knowledge, PISA gives a concrete picture of 15-year-old students' performance in subjects or exercises that are supposed to be relevant for daily life ("life skills"). In addition to this, PISA data, even if they are imperfect and questionable, are very helpful in highlighting differences in educational outcome across countries. According to Duru-Bellat, the misuses and limitations of PISA become obvious, when PISA data are used for benchmarking and when countries are ranked as result of cross-comparative comparisons: "The core problem with benchmarking is that benchmarks are set using the most readily available data" (p. 154). Since PISA data are readily available, they are used as if there were no other relevant indicators of educational quality of an education system (e.g. equity), which is of course highly questionable. However, indicators are isolated pieces of information, which according to Duru-Bellat, are not sufficient for assessing a whole 'system'. For the comprehensive assessment of a whole education system, evaluation is far more useful than indicators, because evaluation requires "the combination of indicators and most of all, the more qualitative interpretation of their meaning" (p. 155). In her conclusion Duru-Bellat points out that her criticism, which is focused on the misuse of PISA data for benchmarking processes, should not lead us "to renounce processes that evaluate education systems based on their output" (p. 157). The student output is and remains an important factor in assessing the quality of education systems. However, according to Duru-Bellat, it needs to be supplemented by additional data: "it is important not to limit oneself to measurement of student achievement but rather to include measurements of system characteristics such as coverage, financing (public/private) and tracking (early/comprehensive tracking, types of student groups etc.)" (p. 156).

Javier Salinas and **Daniel Santín** analyse the PISA reports and results from the economics of education perspective. In their paper the authors present an overview of the problems related to the assessment of efficiency in education and describe how the PISA data have been used for carrying out these studies. The possibility of obtaining educational data every three years for many countries allows economists of education to keep studying the technological relationship between educational inputs and outputs. The aim of a major part of the research done with PISA is to measure the productivity of educational resources and to establish the efficiency level of the schools responsible for producing education. The paper discusses the main educational concepts that have been used in empirical studies to measure productivity using the data coming from PISA and summarizes the main results obtained thus far: e.g. that a greater decision-making autonomy at the school-level tends to be associated with higher levels of efficiency or that, holding resources constant, PISA scores could be boosted by an average of 5% for OECD countries etc. In their conclusion, the authors stress that the PISA reports constitute a very valuable source of information for the analyses of educational efficiency and that they provide very useful information for evaluating educational policy. Finally, the authors provide some concrete advice on what additional information should be included in future PISA reports in order to improve the quality of the empirical analyses that could be conducted using PISA data (establishing a longitudinal database etc.).

The fourth part of this volume entitled “PISA and the Immigrant Student Question” focuses on the potential of PISA for the analysis and understanding of one specific aspect, which is of major importance for most education systems: in many countries immigrant students lag behind their peers from native families in terms of achievement and school success. The relatively poor performance of immigrant students in PISA tests has been one of the most controversial issues in the intense debate about the PISA results. In this part of the volume two papers are presented which both draw on PISA data, but arrive at very different explanations with regard to the reasons for this performance gap between native and immigrant students.

Aileen Edele and **Petra Stanat** assess PISA’s potential for analyses of immigrant students’ educational success by referring to the German case. The authors start by claiming that large-scale assessment studies, such as PISA, “have advanced our understanding of immigrant students’ educational disadvantage considerably” (p. 175) and they prove their point by contrasting what was known about the immigrants students’ educational disadvantage in the German school system before and after PISA. According to the authors, the PISA study established a more comprehensive indicator of immigration background by recording students’ and parents’ countries of birth, which proved that immigration into Germany was much higher than earlier German studies (e.g. *Microcensus*), which had defined immigrants strictly on grounds of their citizenship rather than their migration history, had shown. On the basis of the PISA data, Edele and Stanat are able to identify determinants of immigrant students’ disadvantages in German schools on different levels. On the national/societal level immigration and integration policies as well as differences in the approaches to support second language acquisition seem to play a crucial role. On the school level and with regard to the composition

of the student body, “there is”, according to the authors, “little evidence for the assumption that high proportions of immigrant students, [students not speaking the language of instruction at home, or immigrant students speaking a particular language at home] affect student achievement above and beyond the effects of social composition and average prior achievement of the student body” (p. 185). With regard to the individual level, the language spoken at home is the strongest single predictor of immigrants’ students reading achievement. In addition, immigrant students showed higher levels of instrumental motivation than native students and their achievement disadvantages do not seem to be due to a lack of motivation or aspirations. In conclusion, Edele and Stanat indicate that studies like PISA are powerful tools for identifying strength and weaknesses of school systems and possible targets for intervention. However, they do not suggest concrete measures of how to remedy the identified problem. Especially for measures at the teaching and learning level PISA does not tell us how to improve the achievement of immigrant students. This requires different types of studies like randomized field trials for which Edele and Stanat also provide an illustrative example (e.g. the *Jacobs Summer Camp Project*) in their paper.

In the second paper of this part **Julio Carabaña** discusses why the results of immigrant students depend so much on their country of origin and so little on their country of destination. According to Carabaña, the PISA study opens up new possibilities of carrying out research about immigrant students using a design of the type ‘one origin-various destinations’. When the country of emigration has participated in PISA, a comparison of emigrants with non-emigrants becomes feasible for several countries. On the basis of his analysis of the scores extracted from PISA 2003 and 2006, Carabaña maintains that “with some exceptions, emigrants reproduce the PISA scores of their aboriginal counterparts wherever they go” (p. 202). According to Carabaña, the striking similarities between aboriginal and immigrant students become still stronger, if we account for the special composition of emigrants, which are usually not a random sample from their country population. This leads, in Carabaña’s words, to the following indication: “emigration hardly affects students’ PISA scores, which remain at the level of the country of origin and do not come closer to the level of the destination country” (p. 203). To explain this phenomenon, the author tests various explanations from macro level characteristics of the countries of destination and of origin to personal characteristics and cultural factors. In conclusion, the author arrives at the cognitive ability hypothesis as being the strongest determinant of scholastic achievement. According to Carabaña, the hypothesis of national differences in cognitive or learning ability greatly alleviates the schools in the host countries, because they are “free of the suspicion of depressing the results of immigrant students, or of being unable to help them to develop their full potential” (p. 207).

We have titled the last part of this book “Extreme Visions of PISA: Germany and Finland”, and it provides two papers which look at the PISA debate and results from two very different angles. The papers provide two ‘extreme visions’ of PISA: the first one is written from the Finnish perspective, which is ‘extreme’ because it is written from the perspective of the ‘PISA winners’. The second paper presents

another 'extreme' case because it focuses on the German PISA debate, which is characterized by terms like 'PISA shock' and by feelings of self-doubt. Both papers try to provide explanations for the specific performance of the two education systems: Simola and Rinne try to explain the 'Finnish miracle', while Tröhler tries to explain the 'German double discontentment' with PISA.

Hannu Simola and **Risto Rinne** start off by suggesting three concepts which they consider to be promising theoretical concepts for comparative education. These are: (1) bringing the theoretical concepts of path dependency, convergence and contingency together, (2) tracing the history of the *problématique* and (3) analysing national and local interpretations and translations as hybrids. In their following analysis of the 'Finnish PISA miracle' the authors focus exclusively on the concept of contingency to see whether the concept can facilitate a broader understanding on the national phenomenon of 'Finnish PISA success'. As a first step Simola & Rinne identify three national 'truths' that are widely accepted in Finland even though there is, according to the authors, not too much empirical research evidence behind them: the Finns share a high belief in schooling, teaching is a very highly regarded profession in Finland and the Finnish comprehensive school enjoys rather high trust on the part of both parents, authorities and politicians. In their analysis the authors illustrate that the genesis of these three national 'beliefs' is rather the result of coincidence and conjunction, than the result of rational and purposeful educational planning by educational politicians. In their conclusion, Simola and Rinne claim on the basis of their presented case that conceptualisations such as contingency must be taken seriously when pursuing an understanding of national education policies and politics. The alternative approach, i.e. operating only through functionalist and system models, emphasising mainly the transnational or national trends or focusing solely on rational decisions and choices "does not give theoretically adequate instruments for comparative research" (p. 227).

In the second paper **Daniel Tröhler** analysis the emergence of the lively or even fierce public and academic discussion on PISA in Germany, which he explains as a clash of two very different cultural self-understandings. To begin with, Tröhler clarifies the relationship between three fundamental concepts which lie at the heart of the debate in Germany: competence, *Bildung* and knowledge. According to Tröhler's analysis, the attempt by some German PISA experts to mate competence and *Bildung* has caused major irritation and raised scepticism in Germany. At the background of this conflict lies, according to the author, a 'clash of cultures' between American pragmatism on the one hand and the German concept of *Bildung* on the other. *Bildung* resists being operationalized, is meta-useful and is, finally, unmeasurable. In his historical analysis Tröhler points out that the roots of the present PISA ideology lie in late 1950s, when the Cold War was 'educationalized' in the USA. The 1950s and 1960s was also the time when the human capital theory was developed and increased emphasis was put on maths, science and foreign languages, when cognitive psychology became the main reference discipline for education and when the technical systems perspective became the dominant perspective in education. Comparing these ideological roots of the PISA experts with the German ideology of *Bildung* explains, according to

Tröhler, to a high degree “the harsh rejection in Germany of the merging of the concepts of competence and *Bildung*” (p. 238). This conflict between competence and *Bildung* is even made worse, because PISA’s focus is not directed at what students learn at school on the basis of their (national) curricula and textbooks. Instead PISA aims to test “young people’s ability to use their knowledge and skills in order to meet real-life challenges” (p. 233), which brings it even closer to the non-empirical German ideology of *Bildung*. Against this background Tröhler interprets the German PISA dispute as a double discontent. On the one hand PISA is calling into question the traditional German concept of *Bildung* by focusing on the outer world (‘to meet real-life challenges’) rather than focusing on the development of the inner world (*Persönlichkeit*). On the other hand the PISA results also irritate the PISA experts who had to realize how little their educational project of the harmonious “One World” of free, globally interacting and economically secure citizens had been realized. This is particularly true for Germany, where poor national unity and coherence was greatest, indicated by the vast differences between the PISA results of the immigrant and native students.

We conclude our volume by including three texts as Annexes. Annex I is a research report by Antonio Luzón and Mónica Torres which reviews and analyzes the scientific literature about PISA as well as the public use of it as a important subject which was given widespread coverage by newspapers.

The analysis of the scientific literature on PISA was verified through the publications found in the so-called *Web of Science* (WoS) of Thomson Scientific (better known by its former name of ISI or Institute for Scientific Information), and the database *Scopus* from Elsevier. In addition they included a search in *Google Scholar*, a fourth generation search engine increasingly used in scientific research. Following Luzón and Torres’ study, it appears clear that the coverage of PISA issue was within a very wide subject area within the field of social sciences mainly referred to as ‘Education’; although PISA is addressed by other areas such as economics, sociology, psychology, mathematics education, history and even philosophy, which offers a multidimensional aspect of its reception by the scientific literature. The German sources and the German reality of PISA had a very visible impact. However other publications on PISA tests are also very visible, such as those associated with the “g” factor of intelligence, or with learning techniques in the classroom, and the implications and consequences of PISA in specific learning contexts or for specific social divides, such as immigration.

Annex II gathers the abstracts of the posters exhibited during the holding of the international symposium *PISA under Examination* in La Palma. Before the exhibition, there was – in the symposium – an exposition of the content of each poster by each author. Finally, Annex III reproduces the text of a summary in Spanish of the symposium by Jesús Romero, Antonio Luzón y Mónica Torres. This appeared in the very important educational newspaper of Spain: *Escuela*.

REFERENCES

- Bieber, T. & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education* 46(1), 101-116.
- DiMaggio, P.J. & Powell, W. (1983). "The iron cage revisited" institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147-60.
- Henry, M., Lingard, B., Rizvi, F. & Taylor, S. (2001). *The OECD, globalisation, and education policy*. New York: Pergamon.
- Jones, P.W. (2007a). *World Bank financing of education: lending, learning and development* (2nd edn.). New York: RoutledgeFalmer.
- Jones, P.W. (2007b). Education and world order. *Comparative Education*, 43(3), 325-337.
- Knill, C. & Tosun, J. (2011). Policy making. In D. Caramani (Ed.) *Comparative Politics* (pp. 495-519) (2nd edn.). Oxford: Oxford University Press.
- Knodel, P., Martens, K., de Olano, D. & Popp, M. (2010). *Das PISA-Echo. Internationhale Reaktionen auf die Bildungsstudie*. Frankfurt am Main: Campus Verlag.
- Landfeldt, G. (2007). PISA – Undressing the truth or dressing up a will to govern? In S.T. Hopmann, G. Brinck & M. Retzl (Eds.), *PISA zufolge PISA – PISA according to PISA. Hält PISA, was es verspricht?/Does PISA keep what is promises?* (pp. 225-240). Vienna & Berlin: LIT (retrievable at <http://www.univie.ac.at/pisaaccordingtopisa/pisazufolgepisa.pdf>).
- Lawn, M. (2011). Standardizing the European education policy space. *European Educational Research Journal*, 10(2), 259-272.
- Martens, K. (2007). How to become an influential actor – The “comparative turn” in OECD education policy. In K. Martens, A. Rusconi & Kathrin Leuze (Eds.), *New arenas of education governance. The impact of international organizations and markets on educational policy making* (pp. 40-56). New York: Palgrave.
- Martens, K., Rusconi, A. & Leuze, K. (Eds.), *New arenas of education governance. The impact of international organizations and markets on educational policy making*. New York: Palgrave.
- Meyer, J.W. & Rowan B. (1983). Institutionalized organizations: Formal structure as myth and ceremony. In J. W. Meyer & W.R. Scott, *Organizational environments* (pp. 21-44). Beverly Hills: Sage.
- Moutsios, S. (2009). International organisations and translational educational policy. *Compare*, 39(4), 467-478.
- Mundy, K. (2007) Global governance, educational change. *Comparative Education*, 43(3), 339-357.
- Rizvi, F. & Lingard, B. (2009). The OECD and global shifts in education policy. In A. Kazamias & R. Cowen (Eds.), *Internatrional Handbook of Comparitive Educaiton* (vol. 1 pp. 437-453). Dordrecht: Springer.

Miguel A. Pereyra
Faculty of Educational Sciences
University of Granada

Hans-Georg Kotthoff
Pädagogische Hochschule Freiburg

Robert Cowen
Institute of Education
University of London

SECTION I

THE COMPARATIVE CHALLENGES OF THE OECD PISA PROGRAMME

ULF P. LUNDGREN

PISA AS A POLITICAL INSTRUMENT

One History Behind the Formulating of the PISA Programme

In the early nineties Ivan Illich reminded us that it was time to celebrate the 500 year anniversary of the creation of the educational sector and hence schooling as a system of ideas for power and control over knowledge (Illich, 1981). What he referred to was the first idea to establish a state control system over written texts and thereby mastering the degree of literacy. This distinguish idea was presented at the Spanish court the 18th of August 1492. The month of August that very year is often remembered as the time when Queen Isabel of Spain gave up after all the nagging of Columbus and allowed him to sail to India. But Illich tells another story. The 18th of August the Queen was courted by what we today would call a linguist. His name was Elio Antonio de Nebrija. De Nebrija had published a grammar for the Castilian language. At that time in Europe a grammar was a regulation of how a language should be used, not a description of how a language was used. De Nebrija had discovered that the spoken Latin had changed to some gibberish and no longer a well formed and common language. In twenty year he had tried to reconstruct the classical Latin in Spain but all in vain. Instead, it struck him, that it would be better to write a grammar for the popular language; for Castilian. It was this grammar he presented his queen. But, his idea was more sophisticated than just a set of language rules. The very rational behind introducing a grammar was a new danger. A risk that was discernable as a consequence of the new technical innovation, namely the printing techniques. Due to this invention people learnt to read and that in its turn resulted in all kind of leaflets and pamphlets that were spread around. And many of these texts presented ideas that were threatening to the power and the queen. Ideas were published that questioned what should not be questioned. And furthermore, people were reading in silence. This was also a new invention. Earlier, when there were few texts to read, reading were done loudly (cf. Saenger, 1997). Silent reading is of course more difficult to control and interfere with. The reading had to be controlled, was de Nebrija's clear message. The Queen and thus the state should organise education and teach people to read. If such an education was to be effective and the outcome to be controlled, it was necessary to construct an artificial language. This artificial language had to be constructed on central decided rules and organised on levels following the hierarchical structure of the state apparatus. In that way the reading could be controlled and the empire saved from the contamination of subversive ideas.

U.P. LUNDGREN

This idea of a radical turn from a people's everyday language to one by a grammar dictated language, taught in special institutions is a dramatic shift. It was according to Illich the invention of the public educational sphere. It is also the forming of the politics of education and the forming of devices for the control of the outcome of learning. Queen Isabel rejected the proposal of de Nebrija. She did not see any grand idea behind this proposal of a marriage between the Empire and the Language. She hold to the idea that the language belonged to the private sphere of her subjects. Such doubts and inhibitions have been exceptional among coming rulers.

The idea of this article is to present some reflections on the development of international assessment as a device for political governing. I will do that by pointing on the economic and political context in which international assessments have existed. This is the background for presenting how the PISA programme was formed. I will not go into details, mainly focus on the idea behind the construction of the tests. Finally I will deliver some reflections of why PISA has taken a central in the politics of education during the last decade.

THE DEVELOPMENT OF EDUCATIONAL TESTING AS MEANS FOR POLITICAL GOVERNING

Educational measurement techniques were developed in the nineteenth century. The revised code in England from 1862 is an example of an assessment and inspection system in which financial support to schools were linked to outcomes (cf. Musgrave, 1970; Lundahl & Waldow, 2009). This system of "payment by results" had also the ambition to govern the educational system.

The development of educational tests was early parallel to the progress of measurements of psychological faculties as intelligence with forerunners like Galton with his book *Inquires into human faculty* (1883), McKeen Catell's work *Mental tests* (1890) and of course Thorndike's classical book *Introduction to the theory of mental measurement* (1902).

With the development of the progressive movements in Europe and in the US in the late nineteenth and the early twentieth centuries the idea of evaluations as a base for educational reforms was established. In the beginning of the last century education became of decisive importance both for society and the individual. New governance and not the least the establishment of democracies demanded education. It was by education the future could be formed. For the individual education opened up the doors to a new life. A step from the given to choice.

Education was more and more linked to salaries and a position on the labour market. In this modern world it was important to have information about possible alternatives in order to make the best choice. The concept of evaluation became hence a part of modernity. Educational assessments became the main theme in educational evaluation. Or to talk with Ernest House (1980, p 16):

Modern evaluation is a direct descendant of modernism. Modernisation was liberation from tradition, a shift from the unquestioned reality given by tradition to a social context in which everything could be questioned and changed. It was a shift from 'givenness' to 'choice'.

In this modern education evaluation played a central role. Assessment techniques were developed in relation curriculum content (Tyler, 1950; cf. Kilpatrick & Johansson, 1994).

In the early decades of the last century we can see the first international cooperation for development of assessments being formed. One example is the International Examinations Inquiry (IEI), which was formed in the thirties (Lawn, 2008) aiming at an international cooperation in and for test development. This is an early attempt to build an international network around assessment.

INTERNATIONAL ASSESSMENTS FOR COMPARATIVE RESEARCH

The idea of international comparative assessments came twenty years later. The 4th of October 1957 the first satellite - Sputnik - was launched. The same year the 3rd of November Sputnik 2 was sent out in space carrying a dog – Laika. The Cold War and the competition in space escalated. The 12th of April 1961 Alexejevitj Gagarin was the first man in space. A month later president Kennedy promised that United States within a decade will land a man on the moon. The space race turned the search light on the outcomes of education, especially then the outcomes in mathematics and science. The year after the first Sputnik was launched the International Association for the Evaluation of Educational Achievement (IEA) was founded. My colleague Torsten Husén was one of the founders and acted during many years as chairman. The idea was to build a network of researcher that developed tests designed to be used in comparative studies. IEA was in its beginning a research endeavour, but with time educational administrations were involved.

I am not arguing that there is a simple causality, that the Cold War produced the interest for comparative international testing. The interplay is more complicated. But the international comparisons of results were easy to place on the political agenda in a time where strong voices were heard for competitive educational systems. The political interests interacted with the research interests.

In the fifties studies of economic growth and investments in education showed that investments in education were related to the growth in GNP (cf. Schultz, 1961), which in its turn strengthened the effort to find new roads for improving education and make it more effective. The Human Capital Theory was established. Two consequences are here discernable. One tendency was the focussing on cognitive processes for creating curriculum guidelines and didactic principles. A second tendency was to form an effective teaching technology. The Woods Hole conference at the end of the fifties became the starting point to a period of curriculum development in which the work of Piaget was give an important influence (Bruner, 1960).

It is interesting to note that researchers as Vygotskij had a similar position in Soviet Union as a basis for research of relevance for curriculum development (Jarosjevskij, 1974; Jarosjevskij & Lundgren, 1979).

These curriculum reforms emanating from the US had an impact in most industrialised nations. The work of IEA strengthens of course the internationalisation of curriculum development. The results of international assessment draw the political

view to how to govern goals and content in relation to measurable outcomes. Within education the idea of governing by goals and result was central for reforms long before the New Public Management was coined. Education and teaching always is a process formed by goals, content and results. In periods of change this is more evident than in periods of stability (Lundgren, 1988 and 2003).

When governing of education focus on measured outcomes the validity content of the items will be of specific interest. The Dutch mathematician Hans Freudenthal pointed out in the mid seventies that the content validity of the test in mathematics was problematic (Freudenthal, 1975) The Construction of the items was adjusted to the Bloom taxonomy (Bloom, 1956) and not the content.

How does a national expert value a test which does not belong to any objective of his national instructional system for this or that population or for this or that grade, that is, which is not covered by any subject matter of the national programme? (Freudenthal, 1975, p. 164).

Furthermore there were obvious translation problems. Similar critical questions around content validity in assessment were raised by Urban Dahllöf (1971). Hence, the possibility to compare outcome from different educational settings and curricula was questioned. This criticism had later an impact in the discussion around the construction of tests in the PISA programme.

In the seventies the industrial world faced changed economical conditions. The oil crisis in 1973 and 1979 and the increasing international competition strengthened the pressure on the efficiency and the productivity of educational systems. The economist Schultz word from the early sixties – “Truly, the most distinctive feature of our economic system is the growth in human capital” (Schultz, 1961, p. 17) – become still more evident in the seventies. With a change in economic growth the space for reforms was limited and new reforms had to be financed by increased efficiency. International assessments became now more important in national policies and were broadened in scope and in participating countries. Bloom expressed the ambition of international testing as base for school improvement in the following way:

The IEA surveys provide baseline data for each country against which future changes in education may be appraised. The IEA instruments and the increased sophistication about evaluation in each of the countries provide methods and procedures for the systematic evaluation of the effectiveness of new approaches to education (Bloom, 1974, p. 416).

But the seventies was also a period of criticisms against quantitative methods. The Cambridge Manifesto of 1972 illustrates very well this criticism.¹ In this manifesto it was pointed out that too little research had been directed towards teaching processes and too much attention had been given student behaviours. The reason for this was a research climate that reinforced precision in measuring and concoctions of school problems and research questions. New models and methods were the solution to this state of affairs.

In the seventies the educational systems were under attack for failing in efficiency and productivity and the educational research was under attack for being too much devoted to statistics and psychometric.

A CHANGED PRODUCTION AND A CHANGED ECONOMY

When we entered the eighties another profound change took place. The dilemmas to governing large-scale welfare institutions were striking. There had been a continuous professionalization within welfare institutions, that is, more educated and professionally devoted personnel which became difficult to politically govern. The magnitude of reforms gave little of space for change and with less economic growth these conditions were accentuated. The political landscape in many countries changed with new parties entering the scene— like the green - following other political ideas than the traditional ones and not that easy to place within the right – left continuum (Granheim, Kogan & Lundgren, 1990).

The globalisation and the governability problems called for new solutions. Two main alternatives were on the agenda. One was to decentralise, the other to create more competition by opening up for choice of schools and opportunities to establish private schools. In many countries the arguments for decentralisation were renewed. It could be characterised as a frozen ideology, now melted, and in the first instance realised by local development work, school improvement projects and school based evaluation and in a change of the role of school leaders. It is here the New Public Management is entering as a “solution” (cf. Nyttell, 2006). Education became the arena for consultants with ambitions to increase efficiency and restructure management.

Decentralisation was one discernible solution. However, from a broad international standpoint the picture is not that clear. In the US, as well as in the UK, changes in educational policy can be understood as a change towards centralisation. In the US, the development of standards can be interpreted as federal governing of the national outcomes. In the UK, centralisation was discernible in the development of curricula, accountability, the choice of school and the development of inspection and control. These changes aimed creating visible outcomes reinforcing competition and facilitating the choice of schools.

These moves towards decentralisation were not limited to education alone (Weiler, 1988 and 1990). There was, irrespective of changes in direction of policy-making towards or from the centre, some basic alterations in the relationship between the state and general education, and also in the relationship between civil society and general education. These changes were discerned in the 80s and became central in the 90s, both in public debate and in how governance was performed.

THE KNOWLEDGE SOCIETY

One important change concerned the relationship between national policy and the control of the national economy. Production had transformed Capital and was now moving from being located in tools and machinery to be in human competencies. The power of the capital was moved to the owner of knowledge. (Schön, 2000, p. 521; my translation).

To move enterprises in which the main substance is human competence is easier than moving tools and machinery. To finance reforms by increasing taxes, which partly could be done during the period of expansion, was limited in a more global

economy. With an increasing dependence on the international economy, the possibilities to manage the national economy and the incentives for growth changed in nature. These changes accentuated one of the basic problems of the modern state, to have a profound basis for its legitimacy. A change in legitimisation in a situation of diminished economic control became, in some instances, the impetus for moving state reforms from cost-taking initiatives to a symbolic reconstruction of existing institutions.

As pointed out, the transformation from a labour market structured by industrial production to a labour market structured by service production, circulation of products, reproduction and above all the new information technology, created new demands and reforms. It can be argued that the traditional organisations constructed to handle the economy and the political economy of modern industrialised society was no longer suited to handle a late modern society. They could not mobilise support for action. Accordingly, state institutions such as schools could not attract and build on the interests of the clients or users. Governance had to take other paths. One such way out of the dilemma is to focus on outcomes and accountability making education more transparent.

The trend towards global competition meant that new reforms could not be financed by an increase in taxes. They had to be financed by economic growth. Here we have a dilemma. The development of production – in the knowledge society – demanded more of education. Increasing resources has been the circumstance for the expansion of education, but resources are limited, and in a more global economy, as said earlier, new resources are not that easy to mobilise by increasing taxation. Further expansion had to be financed in new ways and by higher productivity. And this in its turn means to control the outcome of education.

The expectations of increased efficiency and productivity called for concrete well-articulated goals and a steady direction. But what we could discern, in the 70s and 80s, was that the governing subject – the government and administration – became weaker and fragmented. One explanation for this is the splitting up into smaller political party fractions, thereby forcing fragile coalitions. It has been argued that the classical ability of a government to be strong, to be able to reject demands, was lost in the 70s (cf. Crozier, 1977). This, in turn, created an increasing sensitivity to lobbying and power pressure, which led to an overload of demands on decision-makers.

The political authority of a government and its administration is composed of two elements: its effectiveness and public consent. Effectiveness and consent are related, but they can be in conflict. In order to guarantee the consent of the electors, and increasing number of interest groups and associations have been formed. This has created new problems. The more organisations that are formed, the more negotiations are necessary to gain support for one line of action or for a reform. A co-operative negotiating context is formed. This can result in indifference with respect to participation: citizens become de-motivated (Rose, 1980).

These problems seem occur quite frequently in educational administrations at that time with the result that governing documents, like curricula, became abstract to allow for various interpretations. Thus, these forces act contradictory to what was necessary for reforms in a new political context: that is well-articulated goals

and a steady direction. And here we can see the context to the variation in directions of curriculum discussions and suggestions. So once again there is a paradox. Decentralisation calls for more of goal governing and more governing by results, but at the same time goals expressed in curricula become abstract and difficult to assess.

In addition, many of the changes indicated so far were only part of more complex changes in the conditions of political leadership. There are reasons – in this context – to draw attention to the differentiation within the state apparatus itself. To be able to control the move towards politically defined goals, the educational administration organisation must be capable of ranking goals, making priorities and identifying alternative actions that are best adjusted to given economic conditions. Heavy specialisation and division of labour in central governance was relied on as the basis for rational decision-making. This specialisation has as a consequence the splitting up of the organisation itself, with the risk of losing the overall perspective that is necessary for rational decision-making.

It has become more and more evident during the '80s and '90s that earlier planning models could not be used. During the expansion, specialisation of the administration was a practical solution. Faced with the need to take new types of decisions in a different societal context, the existing organisation seemed to be unable to act rationally. With limited resources, various sectors were forced to compete with each other. A consequence of this competition was, in some places, that goals for education were broadened in order to make the educational sector look as important or even more important as other sectors. This broadening of goals was reinforced by the necessity to satisfy various and often different demands. And once again we can see the contradiction between what was produced and what was needed. And again, goals became more abstract when more clearly stated goals were needed.

What many political scientists pointed out (cf. Wildavsky, 1976) in the 70s was that the governing subject – the political leadership – had problems taking the initiative for an active reform policy. We can see examples of a fragmentation of the educational administration, thereby creating problems concerning overall planning and the ability to master complex groups of interrelated problems. We can also see tendencies towards more policy-making carried out by the administration itself.

To meet these problems with decentralisation call for new ways for political governing. The basic characteristics of centralised systems are that they are governed by resources, i.e. the economic system, and thus strongly regulated and framed. The curriculum system is rather detailed curricula and in textbooks as well as in teacher education. Movement towards decentralisation or more market competition weakens governing by economic resources. By that follows a deregulation, or at least a re-regulation. What remains for the centre in a decentralised system is then to strengthen the curriculum system and the evaluation system, i.e. to perform governing by goals and results, if the educational system is to serve the purpose to promote equality and to reproduce a common value-base.

To govern education by expressing goals to be achieved and evaluating the achievements demanded new conditions for governing. To be a steering device,

goals have to be clear. Here a new problem or dilemma arises. If, as was said earlier, one of the problems of governing is that as a result of pressure from various interest groups, and by a fragmented and specialised sector, goals become more broad and abstract, then these processes are contradictory to the demands of steering by goals. One way out of this dilemma is to reorganise the administration and to renew steering documents. One further argument for that has to be added - it is the rapid change of knowledge.

With the new and rapidly changing economy and production, as well as globalisation, and the rather dramatic changes in the volume and structure of knowledge, we have to realise that it is becoming more and more difficult to centrally plan the content of education. More decentralisation means that we have to perform the governing of content in new ways. In moving from central governing towards more local governing, the question of who has the responsibility is sharpened. Thus a movement towards decentralisation focuses the professional ability of teachers and their professional responsibility.

The access to information is rapidly increasing. Schools as institutions were created in a society poor of information. The way curricula and syllabi had been constructed reflects that. In the information dense society, the gravitation point in curricula cannot any more be the organisation and order of content. We are approaching a Copernican turning point, in which curricula must be based on how knowledge is structured, and articulated in basic concepts, theories, models and competencies, which in their turn must be expressed in terms of goals. In performing such a change, curriculum construction and processes for curriculum construction have to be changed. This means new forms of specialisation within the administrative bodies that represent interests other than the ones linked to specific content and thus specific school subjects.

There is one fundamental argument for governing by other type goals and outcomes than before. Resources and rules can govern areas or sectors within which we have a profound knowledge or belief about the relations between goals and methods. If we know that there is a clear relationship between – to take a simple example from traffic policy – speed, conditions of roads and car accidents, we can execute governing by resources and rules. On the other hand, the less general knowledge there is of the relation between goals and methods, the more governing by goals is applicable. The same when the competencies for future working life are hard to predict. However, this, in its turn, demands qualified personal having the skills and knowledge to adjust methods to specific circumstances.

Up to this point I have tried to sketch the main lines in the changes of education during the seventies and the eighties. These changes and this discourse for about education is the background to the OECD project INES which will be discussed more in detail a bit later.

At the end of the eighties, the 9th of November 1989, the Berlin Wall fell into pieces. Three years later, January the 1st 1992, the Soviet Union ended as did the Cold War. The external threat of the superpower blocs toned down. Competition was no longer about domination over the territory. It turned more over to a competition about economic power and growth, a competition that also must adapt

to environmental changes. In the nineties ICT entered as a technology education and with Internet the asset to information and knowledge radically changed.

The dominating themes in the public discussion during the '90s were the professional role of the teacher, school management and educational leadership. This has to be understood as a consequence of the changes sketched above. To govern by goals requires clear goals. At the same time these goals must give space for interpretation and implementation. The essence of goals is that they are not formed as rules. Goals have to be owned by those who have the responsibility to implement them. Here the essence of goals meets the essence of professionalism in the sense of having a knowledge base to interpret and make goals concrete in relation to teaching and learning processes. And furthermore, it calls for a clear division of responsibility and, hence, accountability.

To summarize, the changes in production and economy created a pressure on handling an expanding welfare society. Movements towards decentralisation and privatisation can be interpreted as two ways of solving the problems discussed. Both these solutions demand changes in curricula and in evaluations. The contradiction I have tried to point at is that the change of the political landscape and in administration operated in a way that goals became more abstract. The change towards what can be described as the “third industrial revolution” called for new abilities and competencies that reinforced the difficulties to articulate goals in such in a precise way.

This change has changed the conditions for international comparisons by assessments. The German historian Reinhard Koselleck uses the concept “temporalisation” in his research on how concepts change meaning over time (Koselleck 1979, 2003, 2006). International assessment is a concept that has been temporalised. It has moved from the Cold War context to a world threatened by environmental change and conflicts between faiths and a global economy.

PISA IN CONTEXT

These notes about a emerging “knowledge society” has the intention to give a context to the development of the PISA programme. In 1968 OECD established a specific centre for Educational Research and Innovation – CERI (Papadopoulos, 2006). It is unnecessary to say that 1968 was a year of specific importance in the history of education. CERI became besides the Educational Committee as an important policy institute (Waldow, 2006).

During the seventies and the eighties I participated in several OECD activities including an evaluation of the school system in Norway. In the late eighties I was involved in the “Education Indicators Program” (INES). This very ambitious programme aimed at building a system for education statistics in order to enable comparisons between countries within the OECD. Such a statistical system had of course an impact on national policies. In a global world international indicators delivered support for arguments on competitive strength. The active advocate for an OECD statistic was the United States. The background was of course the

U.P. LUNDGREN

emerging knowledge society and the renewal of human capital theory (OECD, 1998, 2000, 2001).

In July 1991, I became Director General of a new Swedish government Agency – National Agency for Education – aiming at national evaluation and development of the school system in Sweden. The Agency replaced the National Board of Education that was established in 1919. The Director General for the national board of agency for education was also member of the board of CERI. I served for nine years, the last two years I was chairman. As being involved in INES I became a member of the steering group for INES. The INES programme had an impact of the statistics produced within the OECD countries. I mean that on the whole the quality increased substantially. At the same time it was obvious that the data collected also had a steering effect. Even if OECD not has the mandate to change policies they influence them. That steering device was one of several reasons why it was important a General Assembly to get a clear support and a mandate from the member countries. Every second year the General Assembly decided on the development of the programme. The statistics were published annually in Education at a Glance. With time it was obvious that political interest grew not at least demonstrated at the minister meetings.

One problem that followed the project from the beginning was how to report learning outcomes. The only available international data that existed were those collected by the IEA. After negotiations with IEA we got access to the data for the member countries of OECD. They could thus after being reworked be presented in Education at a Glance. However, this was not unproblematic affair. When INES got the data they had been published in other forms and had lost its novelty. The most essential was that when outcome data was published in “Education at a Glance” it had taken so much time that data were from a political point of little or no interest.

The IEA data was not possible to use over time as the test varied between collections. The number of participating countries varied also, which gave the comparative analysis various reference points depending on the various data collections.

The launching of an outcome study carried out by INES came up on several occasions. Tom Alexander, at that time director of CERI, argued for an OECD managed programme. I will not go into the rounds and the negotiations between the IEA and CERI. The decision was taken and a steering group was formed to formulate a specification of the assessment program to be required - PISA Program for International Student Assessment. I became a member of the steering group for PISA and worked with it up to 2000. As chairman for CERI I prepared to present the progress at the General Assembly in Tokyo in August 2000. Due to acute sickness I had to leave before the programme came in operation.

As PISA progressed the European Union started to argue for an own assessment programme. Two parallel test systems would have been too burdensome for the EU countries. We have not reached the end of that story.

The major problem to master was the construction of a test that allowed comparisons over time. Freudenthal’s criticism of the content validity of the test in mathematics used by IEA was important. What is also important to point out is the

changes in curriculum discussions in the eighties and nineties in which the concept competence came in focus. These discussions reflected changes in production and economy and not least a change of political governing of education stressing management governing by goals and results as been pointed out earlier.

The discussions we had in the steering group often centred around ongoing changes in educational policies. Walo Hutmacher, member of the steering group – professor in sociology at Geneva University – argued for focussing competencies. These discussions were nourished by the work at Educational Testing Service in the U.S. They developed a test measuring the reading “literacy” in a way that broadened the concept of literacy by covering not only the ability to decode and read but also to comprehend texts.

This “Literacy” concept began to increasingly appear in parallel and in interaction with the concept of competence. OECD/CERI ran a project where the Educational Testing Service designed this test of literacy for the measurement of adults' literacy skills - the International Adult Literacy Study - IALS. Statistics Canada handled the empirical design and data collection. In 1994 The International Adult Literacy Survey (IALS) was carried out including seven countries initiative was conducted. The basic idea was to study “comparable literacy profiles across national, linguistic and cultural boundaries”. It included also a survey on participation in adult education and training. The results pointed at a possible strong relation between literacy and the economic potential of a nation (Jones, Kirsch, Murray & Tuijnman, 1995). IALS was enlarged in two further data collections in 1996 and 1998 (including 16 countries). The IALS study had an impact on what kind of test to be used in PISA. It influenced also the discussion around competencies which resulted in an another OECD project – Definition and Selection of Competencies (DeSeCo) – in which Hutmacher had an active role. Another spin-off was to find indicators on life-long education and life-wide education. I was chairing a working group trying to find indicators with the aim to study relation between various types of formal and informal education and competencies.

Another other argument for tests that measured competencies and were “curriculum free” was to broaden the discussion around the results. Competencies in reading and in mathematics have to be continuously practiced. This means that the environment must offer possibilities to read and to calculate. The outcomes of PISA we hoped could stimulate a debate on learning outcomes not only from an educational perspective but also a broad cultural and social perspective. Rarely has a pious hope been so dashed. One decisive argument was to have results that could be compared over time. The cons with tests that are “curriculum independent” is just that. How to relate the results to the national curriculum?

PISA is now in its fourth data collection. When the first results came they got an impact that was not expected, not even dreamed of.

There is a general problem with any type of comparisons of educational outcomes. They are quickly translated through metaphors taken from sports. Just one will be a winner. That is true for all previous international measurements. With PISA, the results were a shock in as it seems all countries. Even if Finland was the

U.P. LUNDGREN

exception, they had their chock. As one of my Finnish colleagues said – “it was a shock to be the best”.

The way that I have structured this presentation has been to embed for a contextual explanation. It's the “zeitgeist” that explain the PISA effect. During the nineties, the world changed dramatically. A global society grew. New technologies are changing the production. The economy became global and thus intertwined. Two new world economies emerged with the development in China and India and a third is in its beginning in Brazil. In this strongly emerging knowledge society is the competition not longer linked to only natural resources but also to intellectual resources. Education has become an international commodity. In transformations of this kind, there is uncertainty and a concern or even fear for the future. PISA gave school systems a value on an international scale. Every minister of education realised or believed in the necessity to be better than Finland. Political governing of education became the control of outcomes. The consequence is that Curriculum restructuring will be directed towards test performance. PISA is maybe no longer a comparative project. It is a model for the governing of national school development in a global world.

This emerging control regime has been reinforced by the changing world around us. The enemy is not behind a wall, but among us. The terrorist attacks in September 2001 marked changed social control. Control and surveillance in various forms are part of the daily routine. This “zeitgeist” is part of the context where PISA got its political meaning.

TO FINISH

The title of my presentation was “PISA as a Political Instrument. One History behind the Formulating of the PISA Program”. What I wanted to emphasize was that the PISA project and the effect of the PISA project cannot be understood from an educational, psychometric or technical basis. It has to be understood as part of a context that has been historically shaped by changing social conditions, both material and ideological.

Measurement is one governing device that is the essence of public education. It is a more sophisticated technique than Elio Antonio de Nebrija the 18th of August 1492 presented Queen Isabel. It was the year when Columbus missed the way to India, but explored an enlarged and literary global world. PISA is an example of what in a global world nationally is perceived as the answer to what is going to be taught, who it is going to be taught and how will the outcomes of teaching be judged and used for control and political governing.

International knowledge assessments are currently one of the symptoms of a verification of the knowledge we do not know if we need to face in a future we cannot foresee.

NOTES

- ¹. An international conference at Churchill College, Cambridge University, 20th of December 1972 at which a specific manifesto was signed claiming for a broader repertoire of methods used within educational evaluation.

REFERENCES

- Aldrich, R. (2000). Educational standards in historical perspective. *Proceedings of the British Academy*, 102, 39 – 67.
- Bloom, B. S. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: David McKay Company.
- Bloom, B. S. (1974). Implications of the IEA-Studies for curriculum and instruction. *School Review*, 413 – 435.
- Bruner, J. (1960). *The process of education*. New York: Random House.
- Crozier, M. (1977). *The governability of west European society*. University of Essex.
- Dahllöf, U. (1971). *Ability grouping, content validity and curriculum process analysis*. New York: Teacher College Press.
- Freudenthal, H. (1975). Pupils' achievements internationally compared - the IEA. *Educational Studies in Mathematics*, 2(6), 127 – 186.
- Granheim, M., Kogan, M. & Lundgren, U. P. (1990). *Evaluation as policy making. Introducing evaluation into a national decentralised educational system*. London: Jessica Kingsley Publishing Company.
- House, E.R. (1980). *Evaluating with validity*. London: Sage Press.
- Illich, I. (1981). 'Vernacular values' *Shadow Work*. London: Oxford University Press. Open Forum Series.
- Jarosjevski, M. (1974). *Psichologia v XX stoletii. Teoretitjeskie problemy razvitija psihologitjeskoj nauki*. Moskva: Izdatelstvo Polilititjeskoj literatury.
- Jarosjevskij, M. & Lundgren, U.P. (1979). *1900-talets psykologi. Den psykologiska vetenskapens utveckling i Europa, USA och Sovjetunionen*. Stockholm: Wahlström & Widstrand.
- Jones, S., Kirsch, I., Murray, S. & Tuijnman, A. (1995). *Literacy, economy and society: Results of the first International Adult Literacy Survey*. Paris: OECD/Statistics Canada.
- Kallo, J. & Rinne, R. (2006). *Supranational regimes and national education policies. Encountering challenge*. Research in Educational Science 24. Finnish Educational Research Association: Helsinki.
- Kilpatrick, J. & Johansson, B. (1994). Standardized mathematics testing in Sweden: the legacy of Frits Wigforss, *NOMAD, Nordic Studies in Mathematics Education*, 2(1), 6–30.
- Kosseleck, R. (1979). *Vergangene Zukunft*. Frankfurt am Main: Suhrkamp Verlag.
- Kosseleck, R. (2003). *Zeitschichten. Studien zur Historik*. Frankfurt am Main: Suhrkamp Verlag.
- Kosseleck, R. (2006). *Begriffsgeschichten*. Frankfurt am Main: Suhrkamp Verlag.
- Lawn, M. (Ed.) (2008). *An atlantic crossing? The work of the international examination inquiry, its researchers, methods and influences*. Oxford: Symposium.
- Lundahl, C. & Waldow, F. (2009). Standardisations and 'Quick Languages': the shape-shifting of standardised measurement of pupil achievement in Sweden and Germany". *Comparative Education*, (on line publication), August 2009.
- Lundgren, U.P. (1988). Nuevos desafíos para los profesores y para la formación del profesorado. *Revista de Educación*, 285, enero-abril 1988, 291–328.
- Lundgren, U.P. (2003). The political governing (governance) of education and evaluation. In P. Haug & T. A. Scwandt (Eds.) *Evaluating educational reforms. Scandinavian perspectives*. Greenwich, Conn.: Information Age Publishing.
- Musgrave, P. S. (1970). *Sociology, history and education*. London: Methuen & Co.

U.P. LUNDGREN

- Nytell, H. (2006). *Från kvalitetsidé till kvalitetsregim. Om statlig styrning av skolan*. Uppsala: Acta Universitatis Upsaliensis. *Uppsala Studies in Education*, 114.
- OECD (1998). *Human capital investment: An international comparison*. Paris: OECD.
- OECD (2000). *Knowledge management in the learning society*. Paris: OECD.
- OECD (2001). *The wellbeing of nations: The role of human and social capital*. Paris: OECD.
- Papadopoulos, G (2006). The OECD and the evolution of national policies for education, 1960–1990: An Overview. In J. Kallo & R. Rinne (2006), *Supranational regimes and national education policies. Encountering challenge*. Research in Educational Science 24. Finnish Educational Research Association: Helsinki.
- Rose, R. (1980). *Challenge to governance: Studies in overloaded politics*. London: Sage.
- Saenger, P. (1997). *Space between words. The origins of silent reading*. Stanford Calif.: Stanford University Press.
- Schultz, T. W. (1961). Investment in human capital. *American Economic Review*, 51(March), 1–17.
- Schön, L. (2000). *Modern svensk ekonomisk historia. Tillväxt och omvandling under två sekel*. Stockholm: SNS förlag.
- Tyler, R.W. (1950). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.
- Waldow, F. (2006). *Wirtschaftszyklen und Diskurskonjunkturen: Zur Entwicklung der schwedischen Bildungsprogrammatik, 1930–2000*. Weinheim: Deutscher Studien Verlag.
- Weiler, H. (1988). *Education and power: The politics of educational decentralization in comparative perspective*. Stanford.
- Weiler, H. (1990). Decentralisation in educational governance: An exercise in contradiction? In M. Granheim, M. Kogan & U. P. Lundgren (Eds.) *Evaluation as Policymaking*. London: Jessica Kingsley Publishers.
- Wildavsky, A. (1976). *Speaking truth to power: The art and craft of policy analysis*. Boston: Little, Brown and Company.

Ulf P. Lundgren
Professor Emeritus
Uppsala University

THOMAS POPKEWITZ

PISA

Numbers, Standardizing Conduct, and the Alchemy of School Subjects

OECD's Programme for International Student Assessment (PISA) is part of the new toolkit for the management of school improvement. That management focuses on measuring expectations related to school performance and benchmarks rather than on school inputs, such as increasing teacher credentials and the allocation of resources (Hopmann, 2008). The international measurements of what students learn in schools are related to earlier OECD programs. What is an innovation of PISA is its international benchmark that compares students' practical knowledge across nations in literacy, science and mathematical ability. The official documents describing PISA suggests that its numerical assessments rank the "readiness" of nations' schools for the economical imperatives of the 21st century knowledge economies and Knowledge Societies. The assessment of the practical skills in everyday life situations is believed to be correlated to student's eventual participation in the labor market and being productive citizens.

I admit that the promise of PISA is daunting. The concern with practical knowledge necessary for the future is laudable. Yet anyone reading the history of social science and policy would recognize that predicting that future in the present is no easy task. The difficulty is compounded by the mind boggling effort to conceptualize practical knowledge in a world of dissensus rather than consensus. The challenge becomes more intimidating with the tag-along assumption about having foresight in defining the applied knowledge in a world of continual flux and with change as its singular constant. Classifying the future and taming chance to govern change are never a straightforward and practical errand!

Thus my task here is more modest than the goals of PISA. I examine the grid of practices that give intelligibility to PISA's organizing the knowledge of school subjects. PISA is treated as an historical event. Its study is to make visible the principles that order and classify the objects "seen" and acted on the "practical knowledge" of school subjects. The politics of PISA, I argue, are in the principles that order what children should know, how that knowing is made possible, and issues of inclusion and exclusion embodied in these practices.

The first section historically traces the making of numbers as "facts", a presumption that makes the comparisons of PISA possible. Categories of equivalence are established to give uniformity among diversity. The uniformity and diversity, however, entail particular technologies through which the "facts" of numbers are produced through the very methods that are designed to measure children's knowledge. In the second section, I turn attention to the principles of school

T. POPKEWITZ

subjects that order and classify the “facts” of PISA’s measurements. The notion of alchemy is to consider the translation of disciplinary knowledge into the pedagogical knowledge. The practical knowledge about science, mathematics, and literacy education measured by PISA, I argue, has little or nothing to do with the practices of disciplinary fields! Pedagogical knowledge is concerned with the ordering conduct. The internal rigor in PISA’s measurement practices is, I argue, built on a chimera; an illusion that has consequences. The third section explores the consequences. The pedagogical models inscribed in PISA assessment of learning science and mathematics generate principles about who the child is, should be, and who is not that child. The principles embody cultural theses about modes of living that are named, as one gesture, as the lifelong learner in the Knowledge Society. These notions of the individual and society, I argue, are not only about a particular kind of person and community. The pedagogical style of thought instantiates a comparative style of thought that differentiates and divides populations in its practice to include.

The strategy is to study PISA as an historical event. It numbers and magnitudes are placed in a grid of practices that give its pedagogical distinctions intelligibility. I use the notion of grid to draw attention to the notions of practical knowledge and the lifelong learner as not “things” or concepts to measure. The objects seen, thought about and acted on in PISA are given plausibility and reasonableness through the scaffolding of different social and cultural practices. The kind of human named as lifelong learner is analogous to a recipe for baking a cake. The cake is made through ingredients mixed together. The outcome is “the cake”, an object or a determinant category that appears as having its own ontological existence! The subject of PISA – the practical knowledge of the lifelong learner – is as the cake, determinate categories about the present and future in which different principles come together to order what is thought and acted. The particular grid that makes possible this kind of person is no longer visible. The task of this inquiry is to make visible the grid assembled and its limits in contemporary reforms.

NUMBERS, PISA, AND REFORMING THE FUTURE SOCIETY BY MAKING PEOPLE IN THE PRESENT

PISA is part of a relatively new industry of international comparisons of educational institutions. The international comparisons of pre-tertiary schooling entail, for example, the Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) that are used in approximately 60 countries (http://nces.ed.gov/surveys/international/pdf/brochure_USparticipation.pdf).¹ In addition there are a host of comparative measures that rank higher education. These include The ARWU list – often called the Shanghai list, The Times Higher Education list (THE-QS), and The Webmetrics, and The Professional Ranking of World Universities (Lindblad & Foss Lindblad, 2009).

PISA, among these, has a particular importance. The 2007 Executive Summary, for example, describes PISA as involving nations that include “90% of world economy. 400,000 students in 57 countries, 30 OECD and 27 partner countries,

national representative sample representing 20 million 15 year olds.” PISA sponsorship by OECD and its comparison of the students’ “practical knowledge”, in the words of the program, is to measure school systems’ contribution to the competitiveness of the nation in the new global economic demands. The significance of PISA, Grek (2009) suggests that other international organizations (IOs), the OECD has become part and parcel of the internationalizing and globalizing and thus converging policy processes that have been commented on by many scholars in relation to education... While it is primarily concerned with economic policy, education has taken on increasing importance within that mandate, as it has been reframed as central to national economic competitiveness within an economist human capital framework and linked to an emerging ‘knowledge economy’ (p. 24).

The question of this section is, how can the numbers of PISA be seen as “facts” and as a way of “telling the truth” about society, schooling, and children be historically understood? That is, my concern is not with the internal validity or reliability of the test items but with the conditions which make possible the style of thought embodied in PISA. These conditions are more about the making of the citizen and moral economy than about learning particular work skills or the disciplinary cultures in which science and mathematics are produced.

PISA’s narratives about the present and future are premised on numbers as “facts” that tell the comparative truth about national schooling and the progressive/erosion of societies. The importance of numbers is not only in PISA but part of contemporary societies. This is easy to demonstrate, ironically, by citing numbers. If we focus on the U.S. gross national product, measuring people and things absorbs 6% of the U. S. (Porter, 1995, p. 28). But at a more general layer, it is almost impossible to think about schooling without numbers: children’s ages and school grades, the measuring of children’s growth and development, achievement testing, league tables of schools, and identifying equity through statistical procedures about representation and success rates of populations.

To historicize this making of numbers as “facts”, I turn to cultural and social histories. In an important book about numbers and social affairs, Theodore Porter (1995) begins by asking, “How are we to account for the prestige and power of quantitative methods in the modern world?”... “How is it that what was used for studying stars, molecules and cells would have attraction for human societies?” To consider these questions, Porter continues that only a small proportion of numbers or quantitative expressions have any pretence of describing laws of nature or “even of providing complete and accurate descriptions of the eternal world” (pp. viii-ix). Numbers, he argues, are parts of systems of communication whose technologies create distances from phenomena by appearing to summarize complex events and transactions.

The privileging of numbers as a way of telling the truth about social life and people can be expressed through various and historically recent qualities and characteristics in the construction of modern life.

First, quantification is a technology of social distance. The numbers of PISA provide a common universal language about equivalences. Census data about

populations, data about gross national products, and measurement scores about practical knowledge in science, for example, are such categories of equivalence. The number forms a space of governance through the standardization and technologies that transform cognitive schemes of statistics and scientific thinking into spaces of equivalences.

The seeming rigor and uniformity of numbers appear as transported across time and space so as to not require intimate knowledge and personal trust. The comparing inscribes a seeming naturalness to answers in different national settings. As placed in the perennial struggles of sciences and policies against subjectivity, numbers appear to exclude judgment. The mechanical objectivity of numbers appears to follow *a priori* rules that project fairness and impartiality, excluding judgment and mitigating subjectivity.

Second, the objectivity and the sense of equivalence in numbers have become part of the narratives of democracy. In the 18th century, prior to the French Revolution, the philosophers argued for the metric system to replace the vague and local systems of measurement by feet, hands, wheel barrows. An equal measurement system was deemed necessary for equality itself. By the 19th century, numbers defined a space for standardizing its subject and producing an object that seems merely technical, and its proper calculation to enable giving all an equal chance and representation.

Third, the claim of objectivity for numbers was itself instantiated historically in social processes. Any domain of quantified knowledge is artificial through creating uniformity among different qualities of things (Porter, 1995, p. 6). That uniformity gives social authority to particular norms and cultural narratives that are themselves embodied in social science and policy. Numbers embodied in educational discourses, for example, are instantiated by moral and political discourses. The debates about intelligence testing and eugenics have illuminated that the numbers of measurement in schooling never stand outside of its social spaces of production and realization. PISA, for example, is not merely about numbers and comparison about “practical knowledge”. Practice is itself a theoretical notion that is system of reason that orders and classifies what is seen, talked about, and acted on. The practical knowledge measured in the formulations of PISA embodies distinctions and differentiations about, for example, children’s capacity to solve and interpret problems, and “motivation to learn, their beliefs about themselves and their attitudes to what they are learning”. These categories about problem solving and motivation, however, are not merely descriptions of what children do but theoretical qualities from which equivalences and differences are produced to guide the measurement of conduct. Numbers are not merely numbers.

In the above sense, numbers are “actors”. The technologies of comparing through numbers are navigational tools that standardized a particular universe of capabilities to enable comparisons (Lindblad, 2008). If I return to PISA, the categories of equivalence – the practical knowledge measured across nations – create a new reterritorialization and scaling of the relation of individuality, the city, and state (Brenner, 1999; also see Stråth, 2002). In the EU, PISA re-envision the heterogeneity of cultural and political plurality in its member states through a category of “European”. The categories of equivalence seem to bring coherence

and consensus among differences for building a European space that is spoken about as competitive and cohesive (Grek, 2009; also see, Delanty, 1995). The relating of children's achievement to PISA becomes part of a unified space in which European education is to become a "world best" system. Grek, Lawn, Lingard, Ozga, Rinne, Segerholm & Simola (2009), for example, trace how the data production circulates through different European institutions such as OECD as an actor that crosses border positions. The new actor is made into a technology called "International Comparisons Programmes Manager" (p. 15).

If we think further historically about numbers, it becomes apparent that the appearance of numbers as facts is made through the making of those facts. This may sound as an odd way of thinking about numbers and what PISA does, almost to the point of an extreme relativism. But that is not what I am getting to. Rather it is to understand how abstractions are made into "things" that enter into daily life as principles governing reflection and action.

This double sense of the inscription of "facts" through making "facts" can be illustrated with the notion of "markets". Markets are a classification that circulates to explain and critique much contemporary policy and thus a useful example of this phenomenon in modern social science. The category of markets presupposes the notion of systems brought into social theory by Scottish Enlightenment historians and experimental moral philosophers. Smith's *Wealth of Nations* (1776), for example, wanted to probe the effects of the metaphor of system to see how the theoretical entities of philosophy (and moral economy) could actually work by measuring and quantifying things such as rents, profits, and wages as influenced by commodity prices (Poovey, 1998, p. 237).

The heart of Smith's moral economy was the "market system". Markets, however, was not something there to uncover its "reality" in order to appropriate and gauge human interest and/or its processes to bring progress. Markets were a method of thought, a grid of economic and sociological analysis, an imagination, and a method of governing. Numbers were applied to create a way to think about the system to which numbers were applied that "embodied [Smith's] *a priori* assumptions about what the market system *should be*" (Poovey, 1998, p. 216, italics in origin).

Numbers as magnitudes to compare differences was to express the "invisible hand" of wealth and society that connected the individual pursuit of profit and the growth of collective wealth; and to show the incompatibility between economic development and the governmental procedures (Foucault, 2004/2008, p. 321). Numbers did not exist prior to Smith to prove the abstraction of markets. Smith set up ways of measuring and calculating as if they did exist, to say something about wealth and governing (Poovey, 1998, pp. 240-1). The sciences of markets would "solve" the problem of studying the particulars observed so as to standardize phenomena in a manner that could be projected into the future. The historical schema focused on the intersection of subjectivity and sociality. It gave importance to domesticity, manners, women, and commercial society as "the most sophisticated incarnation of human sociality through which the human mind would be collectively revealed" (Poovey, 1998, p. 227).

T. POPKEWITZ

The categories and their magnitudes provided by the numbers became an historical agent of ‘human nature’, a philosophical universal that could be named and quantified to determine the effects of the abstraction of markets (Poovey, 1998, p. 247). The abstraction of markets performed as a cultural thesis about certain kinds of people. Its “second order abstractions such as labor and happiness... was no longer a universal claim but a “non-rhetorical (nonsuasive) place for a kind of representation that described what *could be* as if this potential was simply waiting to materialize” (Poovey, 1998, p. 248).

My focus on markets and numbers is to draw attention to how theoretical inscriptions given as facts are made into facts. Viewing PISA in this context its collection and aggregation of numbers participate in a “clearing” or space where thought and action can occur (Rose, 1999, p. 212). Numbers standardize and relocate the local and the personal in abstract systems of knowledge that at the same time operate in the spaces of personal knowledge.

Further, the measurements provide constant performance indicators in a continual process of locating one’s self in the world that are analogous to global positioning systems (Simons & Masschelein, 2008). PISA globally positions the child and nation through a style of thought that differentiates and divides through creating categories of equivalence among countries. The categories of equivalence (or sameness) function as an identity to represent difference. What now needs attention is how numbers do not act alone but act as they are inscribed in a grid of practices that give intelligibility to kinds of people. The “facts” enlisted through PISA’s measurements of practical knowledge are not merely descriptive of something “practical”. They are assembled historically in a manner that creates a cultural space that shapes and fashions modes of living.

PISA IN A GRID OF PRACTICES: THE ALCHEMY OF SCHOOL SUBJECTS

Numbers, I have argued, embody particular styles of thought that establish categories of equivalence that seem impartial, objective and democratic. But the numbers are not merely categories of equivalence. What constitutes the practical knowledge of PISA’s testing of reading, mathematical and scientific literacy is bound to a particular system of reason that translates disciplinary knowledge (physics and mathematics) into school subjects. The translations are assumed as merely copies of the original, that is, the disciplinary fields of knowledge and cultures. OECD asserts, for example, that PISA measures the practical ability to apply skills in everyday life situations linked to economy and labor and not, in effect, about learning science and mathematics.

But when examined more closely, the descriptions of what children learn are classified through psychologies of the learning sciences. Central are concepts of childhood, the working of the mind and social communications to which “content” knowledge is made subservient. That psychology and pedagogy have purposes other than those concerned with the pedagogies of learning disciplines.² The categories of learning, for example, are not derived from thinking about the processes, cultures and their interactions that lead to the generation of disciplinary knowledge. The measurements about practical knowledge PISA are about the

conduct of daily life. To draw on PISA's descriptions, the "practical" knowledge are related to children's attitudes, the extent to which they are aware of the life opportunities that given competencies may open, and the learning opportunities and environments which their schools offer. These knowledges are placed in the categories of science learning but they are more than that and possibly not even that. The object of the interpretation of numbers is the psychological and sociological categories about the capabilities of the child, the school, and the family ordered and classified through the learning sciences. The outcome measures are placed in relation to factors about school contexts, instruction, students' access and use of computers, and parental perceptions of students and schools, and performances changes in reading and mathematics. The relating of students' performance and data on the student, family and institutional factors is to explain differences in performances.

The learning sciences are part of the grid in which PISA's numbers constitute school subjects. Other elements of that recipe or assemblage that form the commonsense of school subjects can be pursued through the notion of alchemy. Like the medieval alchemists who tried to change lead into gold through chemical processes, pedagogy is the process of moving "things" from one space (disciplines) to another (school subjects). Pedagogical "tools" move academic classifications, ordering practices and cultural machinery (e.g., notions of laboratories, technologies, academic departments, and professional structures) into the school curriculum (theories of learning, age and grade organizations of children, didactic practices, among others). The notion of alchemy directs attention to the transportation and translation "tools" of the school curriculum. Schools require alchemic practices as children are not physicists or mathematicians. The alchemy then is not the issue at hand. Translations are never merely copies of the original. They are acts of creation. If school subjects are creations and not copies of the disciplines that are their namesake, what is produced through curriculum models? This question is posed as the knowledge systems of school subjects form the commonsense of PISA's measurements.

First is to consider that the pedagogical translations inscribe rules and standards for recognition and enactment (participatory structures) that give school subjects their identities as objects to know. The pedagogical models also provide the conditions for the operation to know that knowledge, the latter talked about as instructional processes of teaching.

This leads to the second observation. What is classified and ordered as disciplinary knowledge and, how that knowledge is made knowable and acted on in pedagogy have little to do with the patterns of interaction and communication of the academic fields (Popkewitz, 2008). The translation tools of curriculum are cultural theses about who the child is and should be.

This seems a difficult claim but one that requires unthinking the "trust" given to PISA. That "trust" is that PISA in fact measures disciplinary knowledge through drawing from the pedagogical models that constitute school subjects. This validity of this trust is what is questioned through thinking of pedagogy as an alchemic process.

To explore this briefly, school subjects are ordered through psychological “eyes”, whether we call that “eye” constructivist, social interactional, pragmatic, or behaviorist. When transporting discipline fields into curriculum, the different psychologies are not practices invented to think about the pedagogies to learn disciplinary cultures and their production of knowledge. The psychologies of pedagogy are related to making the child as the future citizen. The principles of the development and growth of the child form cultural theses about how the child is to live and should live as “a reasonable” person. This life is named as the lifelong learner in PISA and more generally in educational policy and reforms. It is a kind of person that embodies particular norms and values that link individuality to collective belonging and “homes”. While I discuss the lifelong learner as a kind of person below, my purpose here is to that the numbers of PISA are never merely numbers. They are inscribed in a grid of practices that take-for-granted the pedagogical models that produce school subjects. The curriculum practices signified as “practical knowledge” in the categories of measurement inscribe cultural theses about how life is and should be lived as the lifelong learner. The classifications and distinctions of teaching science and mathematics are directed to this cultural task of making particular kinds of people.

The translations of teaching mathematics education are illustrative. The learning of disciplinary knowledge is subservient to social and cultural values about the citizen and is not a pedagogy to learn the disciplinary norms and values of mathematics. Mathematics standards reform research in the US, for example, is underwritten by constructivist pedagogies. These psychologies historically are designed as a technology of governing the rules and standards of conduct. The curriculum is directed to the processes and practices through which the child is to order and judge actions in everyday life through abstract mathematical sets of rules and standards. But the symbolic structures of mathematics in the school curriculum are more than learning formulae and mathematical ways to reason. Sutherland and Balacheff (1999), for example, assert that mathematics education is the “modern” social answer to enabling children to become citizens – that is, “members of a society who have access to both a shared culture and who are empowered with intellectual and emotional tools to face problems within the workplace and everyday life” (Sutherland & Balacheff 1999, p. 2). The social answer is about the construction of the self. Brousseau (1997) argues that mathematics education is to develop in children the capacity to ‘be able to’ (Brousseau 1997, p. 12). The autonomy and agency assigned to the child as problem solver is assembled through social and cultural narratives.

The translations of disciplinary knowledge into school subjects thus have a double quality. First, it is to govern conduct through the insertion of particular rules and standards about thought and action. When science “literacy” is examined internationally, there is a dramatic shift to emphasize greater participation and increased personal relevance, and emotional accessibility in the science curriculum (McEneaney, 2003). That participation, however, links the child’s “expertise” in solving problems to the iconic stature of professional knowledge and to national images of its subject/citizen. Children’s participation and problem solving are to learn the majesty of the procedures, styles of argument, and symbolic system that

assert the truthfulness of the expertise of science. The conclusions of academic expertise are boundaries that enclose children's questioning and problem solving.

Second, the ordering practices that classify and constitute practical knowledge embody moral qualities about modes of living. If we take the term "motivation to learn" in PISA, for example, the notion of motivation inserts a particular way to "see", think, and act in designing the interior of the child's desire (Danziger, 1997). Early psychology did not provide explanations of everyday conduct. It was with the emergence of mass schooling that there was an interest in removing children's "fatigue" in learning through calculating and influencing the children's will, motives, interests, needs and desire. This treatment of inner "thought" brought about ways to classify experience itself as objects of administration.³ Motivation became a key player in this administration; its deployment is part of the organization and ordering of conduct in work.

What is deemed as the practical knowledge of PISA, then, is not practical in any pure or natural way. As Tröhler argues in this book, that knowledge is not built on an empirical examination of students' practices and uses of the curriculum in daily life. Further and to return to the discussion of markets, its notions of practice are built through an abstraction whose ordering and classifying procedures construct its "facts" through the making of facts. The facts embedded in the statistical categories are notions of school subjects that are drawn from the alchemy that inserts particular psychologies in governing who the child is and should be. The grid that gives intelligibility to these "facts" serve as "a map" for structuring what is to constitute "experience" and thinking about what is practical and useful. The limits of PISA measures require exploring further the cultural theses about the child produced in the alchemy of school subjects assumed in the assessments.

PISA AS COMPARATIVE CULTURAL THESES: THE LIFELONG LEARNER IN "THE KNOWLEDGE SOCIETY" AND THE DANGEROUS POPULATIONS

My purpose in this discussion is to explore how PISA is possible as a way to talk, think, and act in the field of educational reform. To engage PISA in this manner it to consider the grid in which its numbers, magnitudes, and categories of equivalence are given intelligibility. The privileging of the particular pedagogical psychologies as the translation "tools" for school subjects give focus to a particular kind of person who has the requisite "practical knowledge", what I earlier gave reference to as the lifelong learner who is to live in "the Knowledge Society". The lifelong learner is a kind of person that, however, entails a double gesture. It generates principles about who the child is, should be, and the child who threatens the envisioned future. The double inscription of the capabilities of the lifelong learner and the child feared as dangerous to the future are part of the same phenomenon. The practices through which curriculum models are enacted, measured, and judged are processes of inclusion, exclusion, and abjection.

T. POPKEWITZ

The Lifelong Learner: The Space of Freedom

PISA is signified as an inclusionary process whose measurements are “*relevant to lifelong learning*”, a phase about people who become the agents in the new global social, cultural and economic patterns called variously “The Knowledge Society” and “The Information Society”. The indicators are designed as measures of the abilities of the citizen who can “participate in [society and in the labour market]”. The indicators of students’ are signified as embodying the mode of life of the citizen who through demonstrating the science competencies “will enable them [as citizens] to participate actively in life situations related to science and technology” (OECD, 2007, p. 3). The *lifelong learning* is the determinant classification, like the cake earlier, that is given ontological status as who is and should be that competent citizen.

As said earlier, it would be nice if the future could be predicted and what is progressive, good, and virtuous ensured through these predictive strategies. But alias, the kind of person embodied in the lifelong learner is not merely a descriptive account of the future society and its inhabitants. Its classification and distinctions generate principles to structure experience and order what constitutes what is practical and useful in daily lives (see, e.g., Rose, 1999). The cultural thesis of the lifelong learner assessed in PISA, for example, is a particular life given to continual “learning”. Yet as explored below, that mode of life is generated through liberal and cosmopolitan political notions of the citizen, moral qualities related to particular Protestant notions of salvation and morality, and cultural patterns that reduced to signify the new economies (Popkewitz, 2008). Ironically, the dispositional qualities of this kind of person have no direct relation to the economy, and to the practices of science and mathematics.

What is this cultural thesis of the lifelong learner? Summarizing different literatures related to policy and research in education, the lifelong learner is a particular cultural thesis about modes of life (see, e.g., Fejes & Nicoll, 2007; Lawn, 2003; Popkewitz, 2008). The lifelong learner embodies enlightenment qualities of reason and rationality (science) as a mode of life (re)visioned to express individuality as a life of never-ending processes of making choices, innovation, and collaboration. Individual agency is the self actualization and self motivation to a life of choice.

Individual agency, however, is not about freedom from social constraints and restraints. What constitutes choice is shaped and fashioned by pedagogical theories designed to calculate and administer the rules and standards for reflection and action. Life is to be designed as the continual processes of rationally planning and organizing daily events whose capabilities are historically linked to a particular northern European notion of the modern “mind” (see, e.g., Popkewitz, 2008; Wu, 2006). Personal responsibility is the self-management of one’s risks by continually maximizing the correct application of reason and rationality in a never ending process of innovation. The fragility of this life is, however, tamed through the procedures assigned to define action by learning “problem solving” and “communication skills”, among others. The only thing not a choice being making choices.

Whatever the merits of this problem solving life and living as a “learner”, they are not merely descriptive of some natural reasoning of the child that curriculum, research and testing recoups. The lifelong learner recalibrates the political aspirations and collective belonging through principles generated about community, participation, and collaboration. The lifelong learner is given agency through problem solving and collaborating in multiple communities – communities of learning, discourse communities. Choice in individual life is sanctioned and acts by working collaboratively.

Community and collaboration are narrated to tell the collective obligation of the generalized global community of humanity. That global community, however, is in fact locally produced. The notions of learning and knowing inscribed in PISA, for example, relate to particular cosmopolitan notions of the enlightened citizen that intersect with secularization of salvation themes of the Reformation and the formation of modern republicanism that occurs between the 18th and beginning of the 20th century. The contemporary commonsense principles about diversity, self-emancipation and social progress that are related to particular a historical time and space that is not universal.

This historicizing of PISA’s criteria of knowledge provides a way of considering Simons & Masschelein (2008) argument about the emergence of the new individuality embodied in the lifelong learner. It entails the shift from earlier notions of emancipation to empowerment in which individual life becomes a continual learning process. Individuality is in learning as the capacity for appropriations that engage the uncertainties of the present. Virtue is managing effectively the limits and opportunities of the environment through steering one’s performances in a continual feedback loop of self-assessment.

The numbers of PISA that assess students’ knowledge and skills, then, are assembled and connected to a number of historical practices that become obscured in its naming of the “practical” knowledge children know. The numbers do not stand alone. They are embodied in a set of practices that generate a cultural thesis about who the child is and should be. This human kind is made through the data of numbers but is not only of PISA’s making. The principles generated through the alchemy of school subjects are about rescuing the nation through making the child. The psychological distinctions that PISA uses to talk about the child’s “motivation to learn”, “beliefs about themselves and their attitudes to what they are learning”, and solving problems that will “open life opportunities are practices about modes of living. The curricular competences are about the govern conduct.

Spaces of Exclusion and Abjection

If the notion of the lifelong learner is the cultural thesis about the spaces of freedom in the fiction of world of “the Knowledge Society”, its cultural territories are double gestures in which difference, divisions and abjections are inscribed.⁴ Let me explore this through a commonplace of school reforms in discussing equity. Equity is given expression in the term “all” – “all children will learn programs for all children”, and “education for all”. The Education for All Movement, for

example, is stated as “a global commitment to provide quality basic education for all children, youth and adults”. The program is endorsed by UNESCO, UNDP, UNFPA, UNICEF and the World Bank to provide an “expanded vision of learning” that creates a universal primary education to “massively reduce illiteracy by the end of the decade” (<http://www.unesco.org/en/efa-international-coordination/the-efa-movement/>). The reduction of illiteracy is shaped and fashioned through the narratives and images, spoken and unspoken, of lifelong learning/lifelong learner and its “others” recognized as different but to be provided with “equitable access to learning programmes” that include through adult literacy, gender parity, and quality education. The commitment is to ensure that there is no child left behind as *all* children will be equal.

When the “all children” is examined, there is no universal and undifferentiated “all” but a particular continuum of value that differentiates and divides. The “all children” implies a unity from which identities of difference are generated. As quickly as reforms state that the purpose is for “all children to learn”, however, the discourse shifts to the child who is different and divided from the space of “all children”. The different child is to be rescued and saved from his or her unliveable spaces. The space of the all children is the space of a difference and abjection that cases the Other into unliveable spaces.

The space of belonging and differences entails a complex relation that is not one of a dualism or a binary. Often unspoken in contemporary school reforms, the qualities and capabilities inscribed in the category of “all” children are those of the lifelong learner. That is, the lifelong learner is, discursively, the “good” child of the present and future. The child who does not belong to the category of “all” is recognized for inclusion but that recognition, paradoxically, inscribes difference. The difference operates in the in-between space of that can be categorized as the urban child in the US and the UK, the gendered child, and more generally the child who is classified as poor, disadvantaged, and immigrant/ethnic. Policy and programs are to re-design that child who does not fit; yet the processes of rescue and redemption inscribe difference that makes it not possible for the child to ever be “of the average” or as “all” children. My placing of the lifelong learner and its Others as part of the same phenomenon is to recognize that the unity of “all” entails a double gesture that instantiates difference. That difference is through assigning identities that universalize particular kinds of people in the cultural spaces of “all children”. If I use the American notion of the urban child, it embodies a cultural thesis and not a geographical place. American cities, for example, are spaces with great wealth and a cosmopolitan urbaneness that coexist with the spaces of poverty and racial segregation. Children who live in the high-rise apartments and brownstones of American cities appear as urbane, without classifications in school discourse and who do not live in the spaces of urban education and the urban child.

The divisions of the urban child, it should be apparent, are not about place but cultural capacities and capabilities. The cultural distinctions of urban child are used to differentiate children who live in suburbia and rural areas as well as in the “city”. Discursively and practically, urban and rural children are categorized and classified by the same sets of distinctions and differentiations (Popkewitz, 1998).

The distinctions that give intelligibility to the urban-ness of the child are formed in a grid of psychological categories about the child's, for example, low expectations, lack of self-esteem and motivation, and learning through "hands-on" experiences rather than abstract knowledge. The psychological categories are linked with social categories about 'dysfunctional families, school dropouts, teenage delinquency, drug abuse, among other. The assembly and connections of these qualities and capabilities make a human kind different from the characteristics of the lifelong learner (Popkewitz, 1998).

If we now return to the comparisons inscribed in the categories of equivalence in PISA, they make "sense" in a system of comparative thought that has nothing to do with any natural sense of practical knowledge. PISA taking the alchemy of school subjects as its commonsense is to insert the double gestures of its pedagogical principles: the hope of the cosmopolitan society that circulates in the notion of the Knowledge Society and fears of those qualities and characteristics of the child that threatens its present and/or future actualization. In 19th century thought, the inscription of differences was assigned to populations ordered in continuums of civilized/non-civilized. The ordering principles and distinctions about achievement, access, learning, among other categories, inscribe differences and divisions through languages of sciences in the policies of planning people. The differences are given expression in gestures of rescuing and redeeming those populations that are inscribed as different. The simultaneous process of producing the "other" in one's self is not of intentional but occurs under the banner of consensus about what is practical. The processes are instantiated in the very style of thought through which the distancing and immediacy are established.

CONCLUDING THOUGHTS

My focus on numbers is to make visible the system of reason through which OECD's PISA technologies and classifications are made intelligible. PISA is neither purely descriptive of some abstraction called "practical knowledge" nor can it be adequately understood outside of the grid of its ordering, classifying and differentiating system. Numbers as magnitudes and categories of equivalence are never merely numbers when inscribed in social life. The measurements of PISA do not act directly on people but act as part of a grid through which spaces are cleared for reflection and action.

Numbers are inscribed in a field of practices that, in the instance of PISA, entails the alchemy of school subjects that translates disciplinary knowledge into principles to govern conduct. PISA takes the commonsense of school subject and its pedagogical translation tools to make the categories of equivalence that constitute its comparative methods. The rules and standards of the "reason" of PISA constitute domains of people and render them stable in order to calculate, deliberate about, and act on.

The pedagogical translations, I argued, are gestures about modes of living. I used the notion of grid, analogous to the cake recipe, to explore how different principles of numbers, equivalences, and the alchemy of school subjects circulate

and overlap in making possible certain kinds of people. Among the grid, I argued, are numbers as “facts” shaped and fashioned by differences of unlike orders. The magnitudes about children’s knowledge regularize and govern the (im)possibilities of relations among social and psychological components. The social and psychological capacities and characteristics are given as universal but are historically tied to particular times and spaces. Further, the cultural theses generated about equality and education for “all children” instantiate a style of thought that excludes and abjects in its impulses of inclusion.

The “practical knowledge” in PISA, then, is not practical in the sense of natural to the phenomena of working of everyday life. The practical knowledge measured to rank people and society in PISA entails cultural theses about modes of living and principles about a coherent, unitary, and uniform world which the psychometric sciences can apprehend and policy can administer

The strategy of this analysis has been to view PISA as an event whose conditions are made possible through particular assemblages, connections, and disconnections. The notion of an event is to consider the conditions that make possible the commonsense (PISA) as a system of reason. The issue at hand is the ways in which recognition, representation, and identity are produced in the sciences of education and the policies of change.⁵ The limits of the “reason” of PISA, then, requires thinking about its rules and standards for ordering, classifying and dividing that is not “solved” or fixed through more subtle and efficient item construction.

One further aspect of contemporary policy analysis that needs to be problematized is the manner in which reforms are rhetorically positioned in relation to economics. This is evident in PISA’s statement of purpose to create the child for the new knowledge economies. If what I argued above is appropriate, the economic rhetoric stands as part of a cultural practice that is not merely about “economy”. The differentiation of economy as a determinant category separate from other spheres of social and cultural life is itself an invention of the 20th century and related to governing. Perhaps it is useful to reread Adam Smith, among others, who alerted us to the complex and subtle intertwining of the wealth of nations to moral and political philosophy, and to Foucault’s discussion of economy. With different intentions, the sciences of wealth gave focus to issues of the economy as not merely about labor but in the management of life and the production of moral subjects and subjectivities (my contemporary take). To “see” economy as an ontological “thing” outside of its moral and cultural inscriptions, as stated in contemporary policy and its instantiations in PISA, loses site of the grid of historical practices that provide the conditions of labor. Marx recognized this well. It is a historical amnesia that creates memories by forgetting that Ford, for example, could only produce assembly line production of Fordism when there was the (re)vision of the subjectivities of the US coach makers that preceded that mode of work. The assembly line and the modes of working in “high tech” industries today are not merely about “work” but the intersection with social and cultural rules and standards through which the “high tech” work becomes possible as a mode of life and as a way by which one thinks and acts.

NOTES

- ¹ An additional one is planned for adult competences, called program for the International Assessment of Adult Competencies (PIAAC).
- ² I use pedagogy here to refer to the ways that people learn about the practices and processes of engaging, for example, in the disciplinary work of history, social science, and the sciences. The pedagogies of school subjects and what is called practical knowledge that provide the foundations to the measures of PISA have different pedagogical purposes.
- ³ I recognize the “ontic” but am differentiating the things of the world from how they are responded to and are brought into discourses that give epistemological and ontological qualities to experience.
- ⁴ See Kristeva (1982) and Butler (1993) for use of the term through psychoanalytic theory; and Shimakawa (2002) for a more sociological approach. My interest in the notion of abjection is through its systems of reason and a social epistemology discussed below.
- ⁵ I discuss this in Popkewitz (2008) and as in relation to comparative studies of education in Popkewitz (2009). Also see Deleuze (1964/1994), Foucault (1968/1973) and Derrida (1997).

REFERENCES

- Brousseau, G. (1997). *Theory of didactical situations in mathematics. Didactique des mathematiques, 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland & V. Warfield, Trans.). Dordrecht: Kluwer.
- Brenner N. (1999). Globalisation as reterritorialisation: the re-scaling of urban governance in the European Union. *Urban Studies*, 36(3), 431-451.
- Butler, J. (1993). *Bodies that matter: On the discourse limits of “sex”*. New York: Routledge.
- Danziger, K. (1997). *Naming the mind: How psychology found its language*. London: Sage.
- Delanty, G. (1995). *Inventing Europe: Idea, identity, reality*. London: Macmillan.
- Deleuze, G. (1968/1994). *Difference and repetition*. P. Patton, trans. New York: Columbia University Press.
- Derrida, J. (1997). Architecture where the desire may live. In N. Neil (Ed.), *Rethinking architecture. A reader in cultural theory*. (pp. 319–323) London: Routledge.
- Fejes, A. & Nicoll, K. (2007). *Foucault and lifelong learning: Governing the subject*. London: Routledge.
- Foucault, M. (1966/1973). *The order of things. An archaeology of the human sciences*. New York: Vintage Books.
- Grek, S. (2009). Governing by numbers: The PISA effect in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Grek, S., Lawn, M., Lingard, B., Ozga, J., Rinne, R. & Segerholm, C. (2009). National policy brokering and the construction of the European education space in England, Sweden, Finland, and Scotland. *Comparative Education*, 45, 5–21.
- Hopmann, S. (2008). No child, no school, no state left behind: schooling in the Age of Accountability. *Journal of Curriculum Studies*, 40(4), 417–456.
- Kristeva, J. (1982). *Powers of horror: An essay on abjection* (L. Roudiez, Trans.). New York: Columbia University Press.
- Lawn, M. (2003). The “usefulness” of learning: The struggle over governance, meaning, and the European education space. *Discourse*, 24(3), 325–336.
- Lindblad, S. (2008). Navigating in the field of university positioning: On international ranking lists, quality indicators, and higher education governing. *European Educational Research Journal*, 7(4), 438–450.
- Lindblad, S. & Foss Lindblad, R. (2009). Transnational governance of higher education: On globalization of international university rankings lists. In T. S. Popkewitz & F. Rizvi (Eds.), *The*

T. POPKEWITZ

- globalization and the study of education* (pp. 180–202) (National Society for the Study of Education Yearbook, vol. 108 n. 2). Malden, Massachusetts: Wiley-Blackwell.
- Lindblad, S., Johannesson, I., & Simola, H. (2002). Education governance in transition: An introduction. *Scandinavian Journal of Educational Research*, 46(3), 237–245.
- McEneaney, E. (2003). The worldwide cachet of scientific literacy. *Comparative Education Review*, 47(2), 217–237.
- OECD. (2007). *Executive Summary: PISA 2006: Science Competencies for Tomorrow's World*. Paris: OECD (The Programme for International Student Assessment [PISA]).
- Popkewitz, T. (2008). *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*. New York: Routledge.
- Popkewitz, T. (1998). *Struggling for the soul: The politics of education and the construction of the teacher*. New York: Teachers College Press.
- Popkewitz, T. (2009). The double gestures of cosmopolitanism, globalization and comparative studies of education” in R. Cowen & A. M. Kazamias (Eds.), *International Handbook of Comparative Education, Vol I*, 379–395. Springer Science.
- Poovey, M. (1998). *A history of modern fact. Problems of knowledge in the sciences of wealth and society*. Chicago: University of Chicago Press.
- Porter, D. (1991). *Trust in numbers: The pursuit of objectivity in science in public life*. Princeton, NJ: Princeton University Press.
- Simons, M., & Masschelein, J. (2008). From schools to learning environments: The dark side of being exceptional. *Journal of Philosophy of Education*, 42(3–4), 687–704.
- Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge, MA: Cambridge University Press.
- Shimakawa, K. (2002). *National abjection: The Asian American body onstage*. Durham, NC: Duke University.
- Stråth, B. (2002). Introduction: The national meanings of Europe. In M. Malmberg & B. Stråth (Eds.), *The meaning of Europe: Variety and contention within and among nations*. (pp. 1–26) Oxford: Berg.
- Sutherland, R., & Balacheff, N. (1999). Didactical complexity of computational environments for the learning of mathematics. *International Journal of Computers for Mathematical Learning*, 4, 1–26.
- Wu, Z. (2006). Understanding practitioner research as a form of life: An Eastern interpretation of Exploratory Practice. *Language Teaching Research*, 10 (3), 31–350.

Thomas S. Popkewitz
University of Wisconsin-Madison

CLARA MORGAN

CONSTRUCTING THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

INTRODUCTION

Administered by the Organization for Economic Co-operation and Development (OECD), an international governmental organization of elite industrialized states (Salzman & Terracino, 2006), the Programme for International Student Assessment (PISA) encapsulates a process established by OECD member states to measure the quality of their school systems and to monitor levels of student achievement using standardized achievement tests and survey questionnaires. Such assessments produce indicators that serve to address public policy objectives and bureaucratic ends and to sustain a certain political economic rationality. The process of constructing indicators involves attempts at standardizing, classifying and categorizing objects to fit them into the representational schemes that make up our worlds. However, the PISA is a fragile entity that is susceptible to contestation because it is founded on the socially-constructed science of educational measurement.

The chapter proposes a conceptual framework that draws from the political economy, international relations and sociology disciplines. The political economic context situates the PISA within the broader political rationality of neoliberalism. The international relations lens sheds light on the role that American influence played in the PISA's construction and more generally, in the governance of international organizations. Sociological theories drawn from the sociology of science and technology and from Michel Foucault's conceptualization of the power bloc formation provide an understanding of how the PISA 'works' and how it is used to exercise power.

The PISA is viewed as an intrinsic component of the global architecture of education¹ in which various agents such as the OECD and other international organizations, states and experts are involved in constructing and reconstructing knowledge and in legitimizing the discourse and material practices of societal and economic progress (Chabbott, 2003). The PISA, as an instrument of educational governance, can be used by educational policy makers to legitimize what counts as knowledge through the codification and measurement of an object called 'literacy.' 'Literacy' is central to how the OECD objectifies the skills and competencies of the future worker in a knowledge-based economy.

CONSTRUCTING KNOWLEDGE

Research in the public policy field has illustrated the important role played by international organizations in transmitting and constructing knowledge (see for example Sahlin-Andersson, 2000; Porter & Webb, 2004; Mahon, 2005). International organizations such as the OECD provide policy advice, prescriptions, and ideas. The knowledge they produce becomes a “guide to future directions in the reproduction and development of practices that shape an increasingly harmonized global political and economic system” (Porter & Webb, 2004, p. 1). Researchers have analyzed the supranationalization of policy and pointed to the increased trend towards regulation, redistribution and provision occurring at global levels of governance (Deacon, 1997). They have examined the reordering of the world through processes of transnational regulation (Djelic & Sahlin-Andersson, 2006; Jacobsson & Sahlin-Andersson, 2006).

Researchers have also studied the OECD and its role in transnational governance (Mahon and McBride, 2009). Henry et al., (2001) provide a critical analysis of how the OECD influences educational policy-making and the effects of globalization on national and international educational policy. More recently, scholars such as Martens et al., (2010) examine how international educational programs such as the PISA have transformed national educational policies (Martens et al., 2010).

The chapter builds on these studies by proposing a multi-disciplinary approach to studying the PISA that draws on political economy, international relations and sociology. The PISA is described as a power bloc that, at this historical moment, has adjusted as a regulated and concerted system. As Foucault explains,

In a given society there is no general type of equilibrium between finalized activities, systems of communication, and power relations. Rather there are diverse forms, diverse places, diverse circumstances or occasions in which these interrelationships establish themselves according to a specific model. But there are also “blocs” in which the adjustment of abilities, the resources of communication, and power relations constitute regulated and concerted systems.

The PISA reflects a similar pattern of interrelationships that Foucault identifies as a power bloc. Three types of relationships emerge while studying the PISA. They are relations that build technical capacity, relations of communication and relations of power. These relationships “always overlap one another, support one another reciprocally, and use each other mutually as means to an end” (Foucault, 1983, p. 218).

Members of a community of practice build technical capacity by creating tools and methods for their work, by tinkering, and by rendering their problems doable (Fujimura 1987). In order for experts to continue building on these technical capacities, they require an infrastructure for communicating with one another. They are linked together through *relations of communication*. The PISA community members coordinate their activities across levels of government. They create an infrastructure for communicating with one another across institutional structures,

CONSTRUCTING THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

geographic boundaries and levels of government. Infrastructures facilitate the exchange of knowledge, ideas, tools, and practices and encompass relations of work (Bowker & Star, 1999; Star, 2002). *Relations of power* are created within the power bloc. Analyzing these relations helps one develop an understanding of how the PISA governs the conduct of individuals and schools and the way in which these relations are then connected to the global architecture of education.

The analysis draws on the author's doctoral research work.² A range of primary sources, such as OECD documents and conference proceedings, meeting minutes and newsletters, were consulted. Interviews took place with various senior officials, middle-managers and experts involved in the construction of international educational indicators, assessments, and statistics. Most of the interviews were conducted over the telephone and a few were face-to-face interviews.³ The interviews are cited by a coded number and date.

FROM KEYNESIANISM TO NEOLIBERALISM

The OECD's role in the construction of an international student achievement assessment needs to be understood within the broader context of the adoption of neoliberal ideas by its member states. Neoliberalism⁴ promotes the market instead of the state as the regulator of the population and creates the entrepreneurial, active subject. Neoliberal policies minimize government intervention in the operation of the markets, conceive individuals as "active in making choices" and promote the exercise of choice by free individuals who are viewed as partners with the state (Rose, 1996, p. 142-143).

During the post World War II era until the 1970s, OECD economists and educational researchers defined educational issues in terms of manpower planning and investment in education. It was understood that states needed to invest in their manpower since human capital contributed to economic growth. In the educational policy sphere, the OECD focused its efforts on the collection and harmonization of educational statistics for educational planners, developing mathematical models for forecasting manpower needs of the economy, building the knowledge base for the field of economics of education and conducting regular reviews of national systems of education. OECD educational activities were geared towards expanding educational opportunities and enrolment in secondary and post-secondary schools (Rubenson, 1999, p. 11). They were governed by an understanding that investment in education improved the quality of human capital.

Neoliberalism began to take hold in the late 1970s and early 1980s as OECD member states turned away from Keynesian policies that were instituted in the post-war era (Harvey, 2005, p. 39). Neoliberal policy reforms were characterized by the retrenchment of the welfare state, privatization of the government sector, and the deregulation of state industries. At the same time, the degree to which neoliberal reforms took hold in OECD member states varied. In the 1980s and 1990s, as part of the neoliberal reform agenda, governments adopted the New Public Management theory and implemented private sector managerial techniques in the public policy sphere.

C. MORGAN

Under neoliberalism, OECD education policy focused on implementing accountability and performance measures, improving educational quality and monitoring of educational systems. As state-run institutions, educational systems came under criticism for their poor performance. This was particularly the case for American schools which were blamed for the poor performance of American students in international student assessments in the 1980s. Government officials argued for more reliable measures of student achievement to assess future American competitiveness in the world economy.

Neoliberal ideas were transferred to the governance of education systems. Attempts were made to turn public schools into quasi-markets where students are consumers of educational services. This competitive market model encompassed a set of policies that ranged from privatization of schools and the institution of charter schools, to changing school funding formulas and school ranking according to standardized test results. Competitive accountability required measurable outcomes. Standardization of inputs and outputs was essential to measuring and quantifying outcomes (Morgan, 2006).

The role of education was increasingly viewed in instrumental terms: to reduce unemployment rates by developing the appropriate skills and competences in students for 'working life.' Equipped with the right skills, students entering the labour market could easily adjust to a technologically-driven and knowledge based society. OECD educational activities were concerned with the development of a creative and highly skilled competitive labour force (OECD, 1983). As Ball notes, such practices "serve and symbolise the increasing colonisation of education policy by economic policy imperatives" (1998, p. 122).

INTERNATIONAL ORGANIZATIONS AND AMERICAN INFLUENCE

By creating the PISA, the OECD has positioned itself as a leader in the international educational assessment field and as an integral node in the global architecture of education. This global structure includes several key international governmental organizations such as the World Bank, the International Monetary Foundation (IMF), and the United Nations. Regional organizations such as the European Union (EU) are also connected to it. In addition, there is the OECD's 'rival' in educational measurement – the International Association for the Evaluation of Educational Achievement (IEA) which was conceived under the auspices of the UNESCO Institute for Education (UIE).

The World Bank and the IMF are multilateral economic institutions that propagate the "ideals of economic globalization" (Jones, 2006, p. 49). They invest in infrastructural educational projects. The World Bank assesses and monitors these investments by using international student assessments to gauge progress. It also makes use of the PISA data as a proxy for measuring learning quality and for developing recommendations based on such results (World Bank, 2007). The OECD complements the work of the World Bank and the IMF by spreading free market ideas.⁵ The United Nations has several agencies that are heavily involved in education (Jones 2006).

CONSTRUCTING THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

An important regional organization in the educational global infrastructure is the European Union (EU). The European Commission, the executive body of the EU, has observer status at the OECD. The European Commission's statistical agency, EUROSTAT, produces educational statistics and gathers educational data collaboratively with the OECD and UNESCO. Since the 1990s, the European Commission and the European Council, represented by the Heads of State, have taken an active interest in national educational policy directions and emphasized the need for a competitive and skilled workforce in a knowledge-based society (Jones & Duceux, 2006; Mitchell, 2006).

Evidence points to the important role American officials have played in influencing the governance and policies of these international organizations. American governments have attempted to dominate the international policy sphere since World War II and have influenced policy directions within international organizations, including the OECD (Djelic & Sahlin-Andersson, 2006, p. 397). At times, they have used multilateralism as an instrument for perpetuating American interests and for maintaining a certain international political economic order conducive to American prosperity and capitalist growth (Karns & Mingst, 1990; Woods, 2002; Foot, MacFarlane & Mastanduno, 2003).

THE EMERGENCE OF AN EMPIRICAL SCIENCE FOR EDUCATIONAL RESEARCH

With the Russians launching Sputnik in 1957, the American government was determined to produce a highly qualified scientific and technical cadre capable of beating that of the Russians. In addition, American policymakers began to question the merits of their own educational system. Multilateralism was one of the instruments the Eisenhower Administration used to enroll European allies into its Cold War strategy to build a highly qualified scientific and technical cadre. In 1958, it gave the OECD's precursor, the Organisation for European Economic Cooperation (OEEC)⁶ a \$500,000 grant in order to establish the Office for Scientific and Technical Personnel (OSTP). The work practices and the tools that were developed by the OSTP helped create the technical capacity and the infrastructure for future OECD work in the area of education.

Even though educational reforms had been initiated in the early 1950s, Sputnik expedited the implementation of these reforms to ensure American schools produced more scientists and engineers (Bybee, 1997). In this context, American comparative educationists felt that there was a need for a 'scientifically' based approach to studying education comparatively rather than a cultural or narrative approach "largely concerned with the exchange and collation of descriptive material" (Postlethwaite, 1966, p. 356).

During the same time period, educational researchers began to meet to discuss issues related to student evaluation and problems facing educational systems (Husén & Postlethwaite, 1996; Bottani & Vriгдаud, 2005). Most of these individuals were academics who specialized in sociology or educational psychology and who were affiliated with university research centres. Held on the premises of the

C. MORGAN

UNESCO Institute for Education,⁷ the meetings were a forum for exchanging information cross-nationally and cross-culturally.

It was during a 1957 meeting devoted to educational evaluation issues that several members decided to meet in 1958 at the UNESCO Institute for Education to “consider the possibility of undertaking a study of measured outcomes and their determinants within and between systems of education.” Instead of relying on graduation rates as a measure of educational productivity, these educational researchers proposed to measure children’s learning which might “yield a very different productivity’ measure” (Husén & Postlethwaite, 1996, p. 129). They hoped that this form of international comparative educational research across educational systems would reveal “important relationships that would escape detection within a single educational system” (IEA, 2007). More specifically, they wanted empirical rather than qualitative evidence on the quality of educational systems in an era when Americans and Europeans were still recovering from the Sputnik shock (Foster, 1991). This group of individuals would become the founding members of the IEA.

BUILDING TECHNICAL CAPACITY FOR INTERNATIONAL STUDENT ASSESSMENTS

Members of the educational assessment field have struggled over the years to render their science factual. They faced constantly the problem of quantifying concepts that were not physical properties. As an example, the quantification of intelligence occurred as eugenicists, educational researchers, statisticians, and psychometricians made this science doable. The construction of a quantifiable object called intelligence was not an easy process to undertake and required significant tinkering by members of these communities of practice. Controversies arose as psychometricians and psychologists quantified intelligence. The racial and cultural biases of intelligence tests were difficult to eliminate since they reflected the test creators’ prejudices and values. Among psychometricians, there were disagreements on the application of testing theories (see Morgan, 2009, Chapter 3). Similarly, these issues would later arise in international student assessments as technical capacity was built for an empirical science of comparative education.

With the founding of the IEA, an international student assessment community of practice was able to launch its first pilot study in 1960. The study confirmed that such an international comparative study was feasible and that valid inferences could be drawn from the data of an international student assessment (Foshay et al., 1962; Postlethwaite, 1966; Husén & Postlethwaite, 1996). At the same time, several problems were reported in the final report of the pilot study which are still encountered today. These included the reliability of sampling procedures, translation problems and the validity of cross-cultural comparability (Foshay et al., 1962). More specifically, problems of comparability were to plague future studies that the IEA undertook such as the First International Mathematics Study and the Six Subject Study. The fragility of the science of international student assessment can be observed as members of the IEA struggled to address problems of comparability among educational systems.

CONSTRUCTING THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

The IEA produced knowledge and tools that fitted the needs of its community of users, enabling them to analyze the test data for national and international purposes. The IEA was very productive in disseminating and communicating its results. It incorporated new informational technology and tools that rendered its work more doable and used computers for data collection, storage, processing and analysis. This community of practice continued to develop and adapt to new technologies.

In the 1980s, neoliberal educational reforms influenced the direction of international student assessments as the IEA studies tried to report indicators that closely addressed these accountability requirements. The IEA was aware of the increased interest in educational accountability and in international educational indicators that was being voiced by several member states and by international organizations such as the World Bank. Yet, because the IEA did not have a strong organizational structure, it was not equipped to address the reporting requirements for international educational indicators in educational achievement.

Even though the IEA did become a more professionally run organization, it remained a non-governmental institution that neither had the political strength nor the financial resources to maintain its leadership position in the international student assessment arena. International organizations such as UNESCO were willing to collaborate with the IEA, but others, such as the OECD, preferred an arrangement in which more control could be exerted on “the ownership and timing of the data” generated by the IEA studies (OECD/INES 1995, p. 4). The IEA was criticized as an inappropriate venue for assessing students internationally. Its data collection system was deemed to be too fragmented and its data collection practices were not adapted to those of national educational systems (Owen, Hodgkinson and Tuijnman, 1995). Dissatisfaction with the IEA led to the problematization of a new international student assessment, the Programme for International Student Assessment.

In studying the emergence of the PISA, one observes the creation of a new community of practice whose members were originally involved with the IEA. The OECD became a forum for developing a new approach to studying educational achievement. Curriculum experts, educational researchers and psychometricians were pursuing new areas of research work and measurement procedures. There was an interest in assessing student knowledge in science based on a scientific literacy framework that incorporated an understanding of concepts, processes and values of science (Bybee, 1997). New American mathematics standards were adopting a quantitative literacy approach to student knowledge (Schoenfeld, 2002). Adult functional literacy skills were being assessed using new frameworks such as the International Adult Literacy Survey. Among psychometricians, a group of practitioners adopted item test measurement techniques that were developed by the Danish statistician, Georg Rasch.

However, the IEA was not able fully to accommodate these new problem domains, work practices and tools that practitioners involved in large scale assessments were eager to expand on. Furthermore, as more educational systems adopted standardized testing for addressing their accountability frameworks,

C. MORGAN

demand for this type of expertise also grew, stimulating competition among research agencies. The creation of an OECD-based student assessment facilitated an intersection of lines of work and tools, resulting in the formation of the PISA.

CREATING THE PISA

Under pressure to respond to domestic political events, officials from the U.S. Department of Education approached the OECD and insisted on the creation of a system of international educational indicators. These officials believed it was in America's national interest to compare itself against other industrialized states but it had to do so according to a common framework that contained clearly defined indicators (Interview Respondent #6, 9 June 2006). The U.S. Department of Education decided that the OECD was an appropriate venue for the implementation of such a project and proceeded to fund an infrastructure for the production of reliable and comparable international educational indicators which came to be known as the International Indicators and Evaluation of Educational Systems (INES) Project.

The American model for educational accountability that was adopted by the OECD provided a set of educational indicators that could be used by American policymakers and politicians to compare themselves to other industrialized countries. At the same time, Department of Education officials began to fund an international student assessment for science and mathematics (TIMSS) that was to be implemented by the IEA. Paralleling these developments, there was a growing interest in the linkages between adult literacy skills and human capital formation which culminated in the implementation of the first International Adult Literacy Survey (IALS) in 1994.

The INES Project brought together educational researchers who work inside government agencies, universities and research centres. The two key founders of the INES Project, the OECD and the National Center for Education Statistics (NCES), laid a plan for the collection of indicators through a progressive series of phases: an exploratory phase, an indicators development phase, and a production phase. In 1992, the INES Project began publishing *Education at a Glance* which became a "flagship publication" for the OECD (OECD, 1996).

The INES Project was organized into several networks. One of the networks, Network A, was responsible for developing indicators on learning outcomes. Network A members gathered data from various international assessments for the publication of *Education at a Glance*. Learning outcome data were compiled and derived from the IEA studies, the International Assessment of Educational Progress (IAEP), and the IALS. Network A members were dissatisfied with the learning outcome data they were receiving from the IEA studies. They turned to developing another data strategy which was presented in 1995 to the members of the INES Project.

A Strategy for Producing Student Achievement on a Regular Basis encompassed the foundational elements for an OECD international student assessment. It involved the creation of a new data source for compiling regular indicators on student outcomes. The new data strategy also made organizational recommendations

including the creation of a decentralized governing body responsible for administration and consensus-building; the contracting of evaluation work to an agency or to a consortium of agencies through a tendering process; and, the collection of national data to be conducted by participating states (OECD, 1995). The final version of the strategy was approved in 1997. Soon after, the Board of Participating Countries (BPC)⁸ was created to manage the project for collecting “indicators on the knowledge, skills and competencies of students in reading, mathematics and science” (OECD/INES 1997: 12, s.16).

In October 1997, the BPC launched its tendering process. It received three proposals – one from the University of Bourgogne, a second from the Australian Centre for Educational Research (ACER)-led consortium and a third from the IEA-Boston College-led consortium. ACER won the contract for administering the assessment. The ACER consortium proposed to generate new knowledge that was more oriented towards life skills and literacy rather than curriculum content (Interview Respondent #14, 27 August 2006).

The most significant tasks the founders of an OECD-based assessment faced were the development of the assessment domain frameworks and the selection of subject matter and technical experts. The frameworks were to be created and published in 1999. The first OECD PISA cycle was scheduled to take place in 2000, the second in 2003 and the third in 2006. Each cycle would test a major domain: reading was the major domain for 2000, mathematics in 2003 and science in 2006.

The legitimacy of the OECD as an authority in the international assessment field rested on the expertise of communities of practice which were enrolled into its network of associations. Three expert groups were formed, one for each of the assessment domains. The Chairs of these expert groups were chosen because of their domain expertise and their leadership in their respective fields. There was also a geo-political motivation for the choices made so that representation included an American, a European and a British Chair (Morgan, 2009, p. 135). Technical Experts were invited to become members of the Technical Advisory Group. Representatives included members from ACER-led consortium and leading technical experts who were also involved with the IEA studies (Morgan, 2009, p. 136).

The success of the PISA is largely due to the technical capacity and relations of communications that had been built by experts involved in international student assessment community. It is also attributed to the relations of power that are created through the application of the PISA approach to defining learning outcomes. The PISA instrument can be used by educational policy makers to legitimize what counts as knowledge through the codification and measurement of an object called ‘literacy.’ This new approach to measuring literacy was conceived in the IALS where literacy was associated with a set of skills that adults needed to have to function in the information age. Literacy skills were viewed as an “element of human capital” which contributes both to “personal development” and “to aggregate economic and social performance” (OECD and Statistics Canada, 2000, p. 61). The IALS constructed a new technique for quantifying human capital.

C. MORGAN

Instead of relying on educational attainment as a measure for human capital formation, workers could be directly assessed for their levels of functional literacy. Furthermore, the IALS required the coordination of several agents within the global architecture of education that included the OECD, EUROSTAT and the UNESCO Institute for Education.

The PISA builds on this approach to measuring human capital. The PISA's first publication which introduced the assessment framework explained that "the indicators are designed to contribute to an understanding of the extent to which education systems in participating countries are preparing their students to become lifelong learners and to play constructive roles as citizens in society" (OECD, 1999). The discourse was concerned with governing the conduct of students as lifelong learners and as citizens. The underlying objective that unified OECD member states was to be able to measure the international competitiveness potential of their labour force in a knowledge based economy. The assessment framework quantified the literacy of the knowledge worker in terms of reading literacy, mathematics literacy and scientific literacy.

Whereas the IEA measured student performance according to national curricula, the OECD created an assessment that measures student competencies for the global economy. The test items developed for the PISA enforce a new curriculum template that is competency-based rather than curriculum-based. Instead of schools producing workers for the local labour market, schools are now conceived as producing workers for the international labour market whose skills are to be assessed at age 15.

Despite the effects produced by these relations of power, the PISA remains a fragile entity. The statistical analyses, codification, categorization and data production work of the PISA requires that its data be perceived as of high quality, reliable, valid, legitimate and relevant. It is precisely this measurement work that contributes to the PISA's fragility. The same controversies and debates that plagued the IEA in its studies continue to plague the PISA, and more generally, the empirical science of student assessment. These areas include cultural bias, student exclusions from tests, quality of the assessments, expansion of the assessment to include less developed countries, over-interpretation of results and the construction of test items.⁹

CONCLUSION

One of the interview respondents described the PISA in the following manner: "[The] PISA is an accountability engine. It tells you hardly anything about teaching and learning. It tells you that there is a problem but it doesn't tell you how to fix it" (Interview Respondent #7, 9 June 2006).

PISA, as an accountability engine, provides a measure for evaluating the performance of both students and schools. PISA results are reported in the media in the form of league tables and rankings. They arouse public interest. The PISA results provide policymakers and politicians with a lever for enacting educational reforms and for bolstering support for such reforms. However, the PISA and other similar international assessments remain fragile. They are plagued with technical

difficulties, controversies and debates which are (temporarily) resolved by expert groups. The challenge remains the quantification of psychological constructs that are not physical properties.

The PISA reflects a power bloc formation that at this time in history has evolved into a concerted system. This model arises as three overlapping relationships – technical capacity relations; relations of communication and power relations – intersect. The PISA in its current formation serves the needs of politicians, policymakers and international and regional organizations as an accountability engine for governing education in the 21st century.

NOTES

- ¹ By global architecture of education, I follow Jones (2006) who describes it as a “system of global power relations that exerts a heavy, indeed determining, influence on how education is constructed around the world” (43).
- ² Morgan, C. (2009). *The OECD Programme for International Student Assessment: Unraveling a Knowledge Network*. Saarbrücken: Verlag.
- ³ All interviews took place in compliance with Carleton University’s Ethics Committee regulations.
- ⁴ Neoliberalism originated in the writings of intellectuals and academics such as Friedrich Hayek, Milton Friedman and various economists from the Chicago School (see Rose, 1996, 1999; Brown, 2003).
- ⁵ Another organizational actor that helps to put into practice ideas of free trade is the World Trade Organization (WTO). Educational services are covered under the WTO’s General Agreement of Trade in Services (GATS).
- ⁶ The OECD’s predecessor, the Organisation for European Economic Cooperation (OEEC), was created after World War II. It was assigned the mandate of distributing \$12 billion in American aid provided by the Marshall Plan from 1948–1952.
- ⁷ The UNESCO Institute for Education (UIE) was established in 1951/52 in Hamburg, Germany (then the Federal Republic of Germany) and was initially tasked with post-war reform of the German educational system. The UIE’s mandate gradually shifted to include “international comparative education” and “East-West cooperation in educational research” (UNESCO, 2005). In July 2006, the UIE changed its name to the UNESCO Institute for Lifelong Learning.
- ⁸ In March 2004, the BPC was renamed and became the PISA Governing Board (PGB).
- ⁹ For a more detailed discussion, see Morgan (2009), Chapter 9

REFERENCES

- Aubrey, H. (1967). *Atlantic economic cooperation: the case of the OECD*. New York: Frederick A. Praeger Publishers.
- Ball, S. J. (1998). Big policies/small world: An introduction to international perspectives in education policy, *Comparative Education*, 34(2), 119–130.
- Bottani, N. & Vrignaud, P. (2005). *La France et les évaluations internationales*. Paris: Haut Conseil de l’évaluation de l’école.
- Bowker, G. & Star S.L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, Mass: MIT Press.
- Brown, W. (2003). Neoliberalism and the end of liberal democracy. *Theory & Event*, 7(1).
- Bybee, R. (1997). *Achieving scientific literacy: from purposes to practices*. Portsmouth: Heinemann.
- Chabbot, C. (2003). *Constructing education for development: International organizations and education for all*. New York: Routledge Falmer.

C. MORGAN

- Deacon, B. (1997). *Global social policy. International organizations and the future of welfare*. London: Sage.
- Djelic, M.L. & Sahlin-Andersson, K. (2006). Introduction: A world of governance: The rise of transnational regulation. In M.L. Djelic & K. Sahlin-Andersson (Eds.) *Transnational governance: institutional dynamics of regulation*. Cambridge, UK: Cambridge University Press.
- Downing, J. & Dalrymple-Alford, E. C. (1974–1975). A methodological critique of the 1973 IEA Survey of Reading Comprehension Education in 15 Countries. *Reading Research Quarterly*, 10(2), 212–227.
- Foot, M., MacFarlane, S.M. & Mastanduno, M. (2003). Introduction. In M. Foot, S.M. MacFarlane & M. Mastanduno (Eds.), *US hegemony and international organizations: the United States and multilateral institutions*. Oxford and New York: Oxford University Press.
- Foshay, A., Thorndike, R., F. Hotyat, D. Pidgeon & D. Walker (1962). *Educational achievements of thirteen-year-olds in twelve countries*. Hamburg: UNESCO Institute for Education.
- Foster, P. (1991). C. Arnold Anderson: A personal memoir, *Comparative Education Review* 35(2), 215–221.
- Foucault, M. (1983). Afterword. The subject and power. In H.L. Dreyfus & P. Rabinow (Eds.) *Michel Foucault: Beyond structuralism and hermeneutics*. Chicago: University of Chicago Press.
- Fujimura, J. H. (1987). Constructing 'do-able' problems in cancer research: Articulating alignment, *Social Studies of Science* 17(2), 257–293.
- Harvey, D. (2005). *A brief history of neoliberalism*. Oxford and New York: Oxford University Press.
- Henry, M., Lingard, B., Rizvi, F. & Taylor, S. (2001). *The OECD, globalisation and education policy*. London: Pergamon Press.
- Husén, T. (1973). Foreword In L.C. Comber & J.P. Keeves (Eds.) *Science education in nineteen countries. An empirical study*. Stockholm: Almqvist and Wiksell; New York: Wiley.
- Husén, T. (1979). An international research venture in retrospect: The IEA surveys. *Comparative Education Review*, 23(3), 371–385.
- Husén, T. & Postlethwaite, N. (1996). A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, 3(2), 129–141.
- IEA (2007). Brief history of the IEA. IEA website: http://www.iea.nl/brief_history_of_iea.html, last accessed April 27, 2010.
- IEA Publications (1974) *Comparative Education Review* 18(2), 327–329.
- Jacobsson, B. & Sahlin-Andersson, K. (2006). Dynamic of soft regulations. In M.L. Djelic & K. Sahlin-Andersson (Eds.) *Transnational governance: institutional dynamics of regulation*. Cambridge, UK; New York: Cambridge University Press.
- Jones, P. (2006). *Education, poverty and the World Bank*. Rotterdam and Taipei: Sense Publishers.
- Jones, K. & Duceux N. (2006). Neoliberalism in the schools of Western Europe. *Our Schools/Our Selves*, 15(3), 93–111.
- Karns, M. & Mingst, K. (1990). The United States and Multilateral Institutions: A Framework for Analysis. In M. Karns & K. Mingst (Eds.), *The United States and Multilateral Institutions. Patterns of Changing an extra space may be needed here. Instrumentality and Influence*. Boston: Unwin Hyman.
- Mahon, R. (2005). The OECD and the Reconciliation Agenda: Competing Blueprints. Paper prepared for “Challenges and Opportunities Faced by European Welfare States: The Changing Context for Child Welfare”, a conference held 7–8 January 2005 St. Anne’s College, University of Oxford, Oxford, UK.
- Mahon, R. & McBride, S. (2009) (Eds.). *The OECD and transnational governance*. Vancouver: University of British Columbia Press.
- Martens, K., et. al. (2010). *Transformation of education policy*. New York, NY: Palgrave Macmillan.
- Mitchell, K. (2006). Neoliberal governmentality in the European Union: education, training, and technologies of citizenship. *Environment and Planning D: Society and Space*, 24(3), 389–407.
- Morgan, C. (2006). A retrospective look at educational reforms in Ontario. *Our Schools/Our Selves* 15(2):127–141.

CONSTRUCTING THE OECD PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT

- Morgan, C. (2009). *The OECD Programme for International Student Assessment: Unravelling a knowledge network*. Saarbrücken: Verlag.
- OECD (1983). *Activities of OECD in 1983*. Report by the Secretary-General. Paris: OECD.
- OECD (1995). Data strategy for network A. Presented at the General Assembly of the INES Project, Lahti, Finland, 12–15 June. [GA–95–8]
- OECD (1996). *Education at a Glance. Analysis*. Paris: OECD.
- OECD/INES (1995). Network a meeting record. Network a plenary meeting. March 28–31.
- OECD (1999). *Measuring student knowledge and skills. A new framework for assessment*. Paris: OECD.
- OECD/INES (1997). International indicators of educational systems. Network a strategy for student achievement outcomes. March. [DEELSA/ED/CERI/CD(97)4]
- OECD and Statistics Canada (2000). *Literacy in the information age*. Paris: OECD.
- Owen, E., Hodgkinson, G.D. & Tuijnman, A. (1995). Towards a strategic approach for developing international indicators of student achievement. In OECD, *Measuring what students learn*. Paris: OECD.
- Papadopoulos, G. (1994) *Education 1960–1990. The OECD Perspective*. Paris: OECD.
- Porter, W & Webb, M. (2004). The role of the OECD in the orchestration of global knowledge networks. Paper prepared for the International Studies Association Annual Meeting, Montreal.
- Postlethwaite, N. (1966). International project for the evaluation of educational achievement (I.E.A.), *International Review of Education*, 12(3), 356–369.
- Rose, N. (1996). Governing “advanced” liberal democracies. In A. Barry, T. Osborne, & N. Rose (Eds.), *Foucault and political reason: liberalism, neo-liberalism, and rationalities of government*. Chicago: University of Chicago Press.
- Rubenson, K. (1999). Adult education and training: The poor cousin. An analysis of OECD reviews of national policies for education. *Scottish Journal of Adult and Continuing Education*, 5(2), 5–31.
- Sahlin-Anderson, K. (2000). *National, international and transnational constructions of New Public Management*. Stockholm: Stockholm Center for Organizational Research.
- Salzman, J. & Terracino, J. B. (2006). Labour rights, globalization and institutions: The role and influence of the Organisation for Economic Cooperation and Development. In V.A. Leary & D. Warner (Eds.), *Social issues, globalisation and international institutions*. Leiden/Boston: Martinus Nijhoff Publishers.
- Schoenfeld, S. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher*, 31(1), 13–25.
- Star, S. L. (2002). Infrastructure and ethnographic practice. Working on the fringes. *Scandinavian Journal of Information Systems*, 2002, 14(2), 107–122.
- Vinovskis, M.A. (1998). *Overseeing the Nation's Report Card. The creation and evolution of the National Assessment Governing Board (NAGB)*. Washington D.C.:U.S. Department of Education.
- Woods, N. (2002). Global governance and the role of institutions. In D. Held & A. McGrew (Eds.), *Governing globalization. Power, authority and governance*. Cambridge: Policy Press.
- World Bank (2007). *Toward high-quality education in Peru standards, accountability, and capacity building*. Washington, D.C.: The World Bank.

Clara Morgan
Department of Political Science
Carleton University

ANTONIO BOLÍVAR

THE DISSATISFACTION OF THE LOSERS

Pisa Public Discourse in Ibero-American Countries

In Buddhism the term “Duhkha”, related to suffering, is difficult to translate. It consists of a feeling of discontent, dissatisfaction, discomfort, disillusion or frustration, normally produced by not having something one desires or misses. We can say that the reaction to the OECD Programme for International Student Assessment (PISA) in the Ibero-American countries has been one of “dissatisfaction”. The population is discontented or dissatisfied with appearing in the successive PISA Reports, which help to determine their respective policies, as “losers” in comparison with other countries. It seems as if the official educational policy intended to improve the situation by showing – with the legitimacy granted by an external international test – that the situation is quite bad.

The use of the word “loser” may be debatable, but – in the popular experience and discourse – PISA is perceived primarily as a ranking or horse race, in which some countries have not been able to compete appropriately, or they have ended up in much lower positions than they expected. The news media prioritize the results of the ranking, reinforcing the dichotomy of “winners” and “losers”, between countries that offer a good education and others that provide a mediocre or poor one. Thus, a Spanish newspaper, referring to PISA 2006, used the title: “once again the PISA report puts Spain, in questions of education, at the end of the line of the developed countries” (*El Mundo*, 11/12/07).

On the other hand, feeling dissatisfaction is a necessary state that precedes improvement. The educational administrations, in order to reduce this dissatisfaction, find themselves obligated to present declarations or reports showing that, in reality, the results are not all that bad. A well-known Spanish professor (César Coll) titled his commentary about the PISA 2006 data, “far from the social aspirations” (*El Periódico*, 06/12/07), indicating that the real question is that the results “do not correspond to the desires and expectations of our society, and they reveal a situation that is stagnant or has worsened since the beginning of the cycle of PISA studies in the year 2000”.

The analysis of the discourse in the Ibero-American press shows that PISA has a high impact on the way teachers, parents and governments look at education. The results show that the majority of the articles about PISA simply refer to it as an instrument to “measure” the quality of the education. Conversely, governments

A. BOLÍVAR

increasingly use references to the PISA results to add legitimacy to their projected actions to increase efficiency. Thus, they frequently present this argument: “we are going to propose such and such reforms, because, as shown in PISA, our country is lagging behind”

COMPARISON AS A WAY OF GOVERNING

PISA does not want to be limited to a comparison, as it also attempts to compare the “performances” of the different educational systems in order to improve their efficiency in a globally competitive world. In fact, the original goal was to provide countries with measures of their strengths and weaknesses, by analyzing their situation in light of the performance of other systems. The instrument, by establishing one unique pattern on which to base the measurement of the different countries, produces a distribution where the educational systems occupy a relative position. Specifically, the way the information is presented, by classification tables rather than in alphabetical order by countries, leads to a superficial reading, limited to comparing the rankings (Mortimore, 2009). Moreover, according to the study by Figazzolo (2009), the news media have generally adopted the perspective of the rankings when referring to PISA, sometimes blaming the teachers for the poor results.

It is debatable whether to compare countries like those in Ibero-America, which – with the exception of the Iberian Peninsula (Spain and Portugal) – are at a disadvantage in economic development, rate of schooling or educational level of families compared to the mean of the OECD. The ability of a country to reach certain levels of education and quality depends, to a large degree, on its level of development. However, PISA is much more than a ranking of countries: it reveals interesting data about correlations between students’ achievement and their socioeconomic backgrounds, about the most appropriate way to organize schools, or about how to increase the equity of educational results. This information should lead to a search for alternative ways to improve teaching, for example, by training the teachers or improving their work conditions, in order to achieve more egalitarian educational systems while increasing quality. Andreas Schleicher (2006), Head of the Indicators and Analysis Division (Directorate of Education) of the OECD, outlines the PISA goals with regard to educational policy as follows:

The purpose of PISA goes far beyond the mere supervision of the current state of students’ learning in the national educational systems. The information provided by PISA should allow the politicians in charge to observe what factors are associated with educational success, and not limit themselves to making comparisons between results in an isolated way (p. 23).

Within a more global framework of harmonizing the educational globe (Tröhler, 2010), PISA situates educational policies in a worldwide competitive space. In doing so, it becomes a specific way of regulating education. This comparative framework leads to the search for solutions with regard to what works in other countries, learning from the competitors. Professor Fernández Rizo (2006), former

director of the National Institute for the Evaluation of Education in Mexico, pointed out:

In complex topics like those related to education, there should be various referents, as none of them would be adequate alone. Therefore, it is not irrelevant to compare ourselves with the more developed countries, as distant referents, to a certain extent ideal; but it is necessary to complement this comparison with others, for example, with similar countries, with our own situation in the past, and with the goals that have been set for the future (p. 162).

After pointing out the deficits, it is possible to converge gradually with the more advanced countries. Through the publication of its results, PISA wanted to provide the necessary knowledge to make political decisions about the most appropriate educational reforms for improvement. Schleicher (2007, p. 351) argues that “PISA can provide policy makers and practitioners with effective tools to improve quality, equity and efficiency in education”. In this way, it acquires a growing influence in defining the national educational policies. This knowledge, stemming from the evaluation of competencies, becomes an instrument for governing.

Linked to the emergence of processes of “transnational regulation” of education (Barroso, 2006), PISA has become a regulation mechanism that inscribes the educational questions in a new space: international and “objective” measures of results. The main purpose was to provide indicators about the efficacy of the education system of each country in comparison with the others. Therefore, it has contributed to making the public aware, that is, creating a social construction, about the relative position of each country in international space. All of this makes PISA much more than a survey and a set of associated reports, as it organizes very wide-ranging social worlds and uses sophisticated means of policy and knowledge coordination (Carvalho, 2009). PISA has become a tool produced within the scope of research that supports and “takes part” in the task of coordinating public action in education.

The *Knowledge & Policy* Research Project studies PISA “as part of the construction of a multilateral space for the creation and exchange of “knowledge for policy” (Carvalho, 2009, p. 5). The PISA texts can be read as narratives that produce knowledge for politics. More specifically, the construction and diffusion of PISA is a *Knowledge-based Regulation Tool*, understood as “technical-social instruments (based on knowledge and generating knowledge) that disseminate a particular kind of knowledge in order to shape the behavior of actors in a given policy domain”. PISA contributes to organizing specific relationships between those who emit the reports and the potential recipients: “capitalizing on and participating in the construction of this normative setting, it institutes comparative logic as a form of administration and government” (Barroso & Carvalho, 2008, p. 78). PISA can be considered as an organization capable of spreading and legitimizing a certain type of knowledge (the comparative evaluation of competencies) for the regulation of educational policies and action. It becomes,

A. BOLÍVAR

then, a regulation instrument, based on the expert knowledge of an organization like the OECD (Noaksson and Jacobsson, 2003; Mahon and McBride, 2009).

PISA, furthermore, has been the vehicle for a specific type of curricular knowledge: an approach based on competencies, which, in the long run, has caused the different Iberian-American educational policies to be oriented within this framework. As pointed out (Afonso & Costa, 2009): “Defining PISA as a *Knowledge-based Regulation Tool* means that it is conceived as an example of the complex and circular relationship between knowledge and policy: PISA as a policy instrument produces knowledge; PISA as a research instrument produces policy” (p. 6).

PISA A MEDIA DISCOURSE: NARRATIVES OF DISCONTENT

In the Ibero-American setting, PISA has managed to occupy the public educational space like no other type of report or survey. This media success is due, without doubt, to the design of the instrument itself: worldwide application, attention paid to the validity of the instrument, periodic regularity of the survey, generation of its own data, participation that depends on the respective public authorities, policy based on evidence, tests focused on competencies and not on curricular contents (innovative concept of “literacy”), etc.

Diverse analyses have been performed about how the *mass media* have presented PISA in the different Ibero-American countries (Ravela, 2006; Ferrer et al., 2006; Massot et al., 2006). The association of trade unions that make up the International Education (IE) carried out a study in 2008 (Figazzolo, 2009) to analyze the impact of PISA 2006 in the debate on educational policies, focusing especially on how the media reported on the PISA study, the conclusions reached by the governments, and the reactions of the trade unions. Instead of creating awareness in the different sectors and encouraging them to get involved in improving the quality of the education, in Spain and Latin America “the effects of spreading the results generally involve looking for someone to blame, discrediting the educational system in general, and the feeling that no matter how much is being done, it is not getting them anywhere” (Ravela, 2006, p. 298). Thus, it is common to use the results as a means of confrontation between the government and the opposing party.

Validating the saying that “good news is not news”, in the Ibero-American world the PISA reports have been presented – with a certain degree of sensationalism – as the dissatisfaction of the losers in a race: “last in line”, “failure”, “the worst in sciences”, etc. One of the studies points out: “the information offered by the media did not involve an appropriate rational analysis and produced a catastrophic distortion that was quite negative for schools” (Massot et al., 2006, p. 390). An informative simplification based on slogans or, in some cases, manipulation by not presenting all the data, distorts the reality and conditions public opinion. As Figazzolo (2009, p. 26) points out: “media have often advocated the implantation of those features that characterize high-performing school systems into low-performing education systems, regardless of the various contextual factors”. This role increases when, as has occurred

generally, instead of going to original sources of information, the public limits itself to what the press says.

In general, the Ibero-American public's perception of PISA is conditioned by their prior political and ideological position and especially within the framework of the educational debates at the time in the country in question. Rather than performing an in-depth analysis of the results from the PISA study and their implications, the desires and aspirations of each group, including the education administration itself, are projected in them, producing a biased or incomplete interpretation. This type of interpretation has meant that the PISA results have not had any relevant pedagogical repercussions. Thus, each of the PISA reports has been received from a political and ideological duality, serving the educational policy that interested each ideological group, and producing an ideological manipulation of the results.

In contrast with other countries like Germany (Kotthoff & Pereyra, 2009), where they led to a widespread debate that questioned the educational policy and system, in Spain the first PISA results were hardly noticed, receiving at most a self-serving political use. The news media also gave superficial information, but only during fifteen days in December. Thus, the newspaper *El País* presented the headline: "Spanish secondary students, among the worst in the developed countries – 16% of young people in Spain reach the age of 15 with difficulty in reading correctly", and even –in an alarmist way– talked about a "catastrophe in education". As one study concluded (Massot et al., 2006), there was a lack of analysis and diffusion of the results of the PISA study, a lack of a culture of evaluation in the educational setting and, finally, an excessive politicization of the interpretation of the PISA 2000 results by the representatives of the educational community. In the preparation of an educational counter-reform to the LOGSE Law, the Organic Education Quality Law (LOCE), the effects of PISA 2000 were quite limited: increase the hours dedicated to reading and mathematics (Royal Decree 3473 of 29/12/2000), just a few days after the PISA results were published.

Something similar occurred with PISA 2003. For a brief look at the newspaper headlines, we can examine the newspaper archives from December of 2004 "Spanish students at the tail of the OECD in mathematics, science and reading" (*El País*, 7/12/04). An editorial in the newspaper *ABC* (8/12/04) said, "The report shows, moreover, that we have a stagnant system, with a tendency to worsen in areas as relevant as mathematics and reading comprehension". Furthermore, the editorial page of *El País* on the same date (8/12/04) talked about the same thing: "The x-ray of the state of education in the OECD a translation would use OECD countries, the PISA 2003 report, has sounded the alarm about the poor results obtained by the Spanish system. Although the knowledge and skills of Spanish students in mathematics, reading comprehension and scientific culture already appeared at the tail of the developed countries in the year 2000, the macro-study made public on Monday reveal a worrisome stagnation and even a worsening in reading".

This situation was repeated in December of 2007 with the same reiterated discourse: "While Spain receives failing marks from the PISA report, Finland gets

A. BOLÍVAR

the highest grade” (*El Periódico*, 24/12/07), and “Spanish education moves backward. The level of reading comprehension of 15-year-old Spanish students suffers the worst decline in the OECD” (*El País*, 05/12/07). A former Spanish Education minister (Ortega Díaz-Ambrona, 2007) commented:

Every three years the OECD insists on ruining our Christmas with their PISA report. They did so in December of 2001 and 2004. Now they are back with the bad news that our educational system is still poor. We live with hope for a few years without the report, convinced that non-university education is improving thanks to the efforts of the government at the time, but then party pooper PISA arrives and with the PP or the PSOE, we end up looking bad. We seem to be trapped in a sinister vicious cycle that makes it difficult to diagnose the problem and its solution.

And he ended “meanwhile, in 2010, the swallows will be back as well as the PISA report. Will we have better results or continue this never ending story? We’ll see”. In part, this has been the experience of PISA in the Spanish public discourse: bad news that is quickly forgotten, at the most a hurricane passing by.

In the meantime, the objective of the official educational policy has been to counteract the effects, reducing the criticism with official declarations or reports. The respective governments finance participation in PISA and, in exchange, feel they have the right to use the data in their own way. The case of the Evaluation Institute in Spain is paradigmatic of the use for political interests, rather than to generate an enlightened and public use of educational knowledge. Argumentation strategies are developed to use the PISA data to support the respective political policy which, in fact, existed before PISA. Thus, the mediocre results of PISA 2000 were used, among other purposes, to criticize the previous educational law (LOGSE) and thereby justify the new reform (Educational Quality Law, LOCE). Similarly, PISA 2003 arrives at another moment of political change. The new government has stopped the application of the previous law (LOCE) and is in the process of preparing a new educational law. In this context, logically, the report will be used to criticize the previous proposal and support the new one: design the curriculum according to competencies, increase the number of hours for reading and mathematics, and introduce Citizenship Education into the curriculum.

IBERO-AMERICA IN PISA

As far as education and culture are concerned, Ibero-America (Latin America, Spain and Portugal) has become a reality, as shown by The Organization of Ibero-American States for Education, Science and Culture (OEI) and the “Ibero-American PISA Group” (or GIP) network. Moreover, the Ibero-American countries participating in PISA are significantly different in their sizes and in the percentages of school coverage in the age groups being evaluated. Latin America faces problems inherited from its colonial past, and it is considerably behind compared to the industrialized countries. The societies have complex demographics, with profound inequalities and differences between countries in the access to education, and with insufficient quality.

With regard to their participation in PISA, México and Brazil are two Iberian-American countries (together with Portugal and Spain) that have participated in the three editions of PISA. Furthermore, in 2006 Argentina and Chile participated, as they participated in PISA Plus, but not in 2003; Uruguay participated in PISA in 2003; and Colombia did so for the first time in 2006. Some countries stop participating (Argentina, Chile) because of the political consequences of continuing to receive low scores for the current government, or due to the economic cost of participating, especially during a period of economic crisis. In the 2009 edition, other participants were: Panamá, Perú (participated in PISA plus) and the Dominican Republic.

México, together with Brazil and Spain, promoted the creation in 2005 of the *Ibero-American PISA Group (GIP)* network. It was created due to the need to join forces to facilitate cooperation, reflection and mutual help, contribute to improving operational capacity, and encourage a more active participation in the project. Having a common language makes it possible to more effectively share interests, problems, experiences and, especially, initiatives within the *Ibero-American PISA Group*. The GIP is a mutual support group designed to contribute to a better performance on the PISA, by means of continuous technical training and exchanging best practices. The cooperation and support are organized by the National Project Managers (NPMs) and their respective technical groups at the national centers. Which language – the spelling decision again Thus, they have carried out rounds of technical review for the reactive units on reading for each country, before sending them to the ACER (Australian Council for Educational Research). In this way, the PISA 2009 report includes ten items proposed by this Group, which also has a representative in the group of experts who advise the OECD on elaborating the next test. From mere participants (providing a sample to the study), they have taken on responsibilities in designing the instruments and studies (up until now directed exclusively by Anglo-Saxon countries). The GIP has published the book *Ibero-America in PISA 2006. Regional Report* (GIP-OECD, 2009), which emerged from an International Seminar.

Table 1 shows the data from the seven countries participating in *PISA 2006*. They are the countries on the American continent with the largest educational systems, with Brazil at the head, followed by México, Argentina and Colombia; then come Spain, Chile, Portugal and, finally, with the smallest system, Uruguay. There are great differences between these countries in terms of educational coverage of young people between 15 and 16 years of age, where Spain has 100% coverage, followed by Argentina with 87.4%, while México only covers 62.9% and Colombia 60.6%.

The average obtained by the Ibero-American countries is 426, while the Latin American average is 408, far from the OECD average (500 points). With regard to the levels of competencies, Spain is at level 3, Portugal, Chile, Uruguay and Mexico at level 2, and below the minimum level of scientific literacy, at level 1, are Argentina, Brazil and Colombia. As Fernández Rizo (2006, p. 166) rightly states, “the results of Mexico and other Latin American countries on PISA should not surprise us: they lie in the range of what would be expected, due to the weight

A. BOLÍVAR

of the socioeconomic factors and the resources the schools have access to”. Likewise, the results should be interpreted with caution, given that the percentages of school coverage of the 15-year-old population are lower than in other countries. Precisely those youths who do not take the PISA tests, as they do not attend school, pertain to the lowest layers, which means the performance is overestimated. Or the authors could use: Those youths who do not take PISA tests often do not attend school and would thus lower the scores – which means performance is overestimated.

Table 1. Populations of 15-year-old youths in the Iberian-American countries in PISA 2006

<i>Country</i>	<i>Total Young people 15 years old</i>	<i>15 year olds enrolled in school</i>	<i>Percentage of coverage</i>
<i>Argentina</i>	662.686	579.222	87.4
<i>Brazil</i>	3.390.471	2.374.044	70.0
<i>Chile</i>	299.426	255.459	85.3
<i>Colombia</i>	897.477	543.630	60.6
<i>Spain</i>	439.415	436.885	99.4
<i>Mexico</i>	2.200.916	1.383.364	62.9
<i>Portugal</i>	115.426	100.816	87.3
<i>Uruguay</i>	52.119	40.815	78.3

In the case of Spain, there are not many students in the lower levels of the scale or in the upper levels either. There is a certain controversy about whether a low dispersion of results (295 points between the best and worst results, compared to 311 in the OECD countries), means in principle that there is more equity, or whether it is only an “artificial” effect of the homogeneity of the results of Spanish students (Bolívar, 2008). It is possible, then, to interpret as egalitarian that which is only an effect of the narrow dispersion of the results. The Spanish Report on PISA 2006 states (IE, 2007) among its conclusions that “the Spanish educational system is comparatively one that offers greater equity to its students, close to that of the Nordic countries” (p. 100).

Chile is a country that had fairly poor results in the year 2000, decided to withdraw in 2003, and improved significantly in 2006. The Education Minister of Chile (Mineduc) titled the presentation of the results (04/12/2007) in this way: “Chile leads the results of the international PISA test in Latin America”. Naturally, this outcome was used to argue that the Chilean educational policy had been the correct one: “we are drawing closer to the developed countries. We have the best educational results in Latin American”, said Minister Provoste. A more scientific analysis (Cariola et al., 2009) shows that the Chilean students, trained completely in the reformed curriculum, generally had better results than their Latin American counterparts, and that they improved significantly in Reading compared to five years before. However, a lack of equity is also shown by the socioeconomic and academic segregation of the educational system and the disparity in the performance of men and women.

With regard to Argentina, a recent study (Rodrigo, 2009) of the poor performance of its students on PISA found that external factors alone (socioeconomic background of population, economic investment on the educational system, coverage of the Argentinean educational system, public vs. private schools, etc.) are not able to

Table 2. Average results of Iberian-American countries on PISA 2006 countries

	<i>Sciences</i>	<i>Countries</i>	<i>Sciences</i>
<i>Spain</i>	488	<i>Mexico</i>	410
<i>Portugal</i>	474	<i>Argentina</i>	391
<i>Chile</i>	438	<i>Brazil</i>	390
<i>Uruguay</i>	428	<i>Colombia</i>	388

explain these low results. There is no causal relationship between student learning and external factors such as the level of spending on education. It is at the school level where the courses taught shown how the processes depend, firstly, on the learning achieved. According to research on school effectiveness, the most relevant factors in explaining the differential between Argentina and other countries – like Spain – are internal factors: organization of teaching, working conditions of teachers, and levels of academic requirements. The way it works, and how school experiences are organized, depends mainly on: working conditions of teachers (low salaries, multiple employment, training, teacher absenteeism), which explains the low school time devoted to teaching, as well as the low quality of education. Similarly, modes of transmission of knowledge and the limited employment of textbooks show that the level of functioning of Argentinean secondary schools does not guarantee the development of effective teaching-learning processes for all students, which, from a comparative international analysis, would ultimately be one of the reasons for Argentina's low level on PISA.

On the other hand, México was, together with Brazil, a country that dared to participate in PISA 2000. Fernández Rizo (2006) recalls that when the results were published in Mexico “some of the news media gave a negative simplistic version, saying that the country was in the penultimate place in the world with regard to the quality of its education” (p. 158). However, PISA seriously put into doubt whether the management of the educational system was oriented toward educational achievement. The increase in the number of young people in secondary education had not also guaranteed sufficient minimum quality, for example, in reading. PISA, as a distant voice, with its legitimacy and recognition as an international test, came to show the public opinion that “there are very serious problems that should be taken care of and that, otherwise, the country would have to face the consequences in terms of lagging behind, inequality and lack of competitiveness” (Zorrilla, 2009, p. 80).

Brazil is the only country that, without belonging to the OECD, decided to participate from the first edition. In the context of a “managerialist vision” (Motta, 2008), during President Cardoso’s second term (1995-2002), his Education Minister, Pablo Renato Souza, decided to put the Brazilian educational system to the test with

A. BOLÍVAR

PISA 2000, even knowing that the results could be disastrous. But the mediocre educational system in Brazil is not thought to be due to a lack of resources, but rather to a lack of measurement using comparative standards. Far from pretending to compete in quality with richer countries, this evaluation was defended because it is considered one of the most serious and respected instruments in the world to identify the deficits of educational systems. Retrospectively, Souza explains (Souza, 2006) that it was a brave decision to “enter into a study where the most advanced countries are”, as an expression of a country’s maturity to comply with the methodological requirements and analyze the information obtained from a scientific document. He added: “some countries decide not to divulge the research data considered less favorable. Our decision shows a degree of maturity of the education authorities, and it shows the will to calmly analyze the state of our education and extract lessons and implications for making political decisions. [...] The results from PISA will house style? serve as a basis for perfecting the evaluation and follow-up of the effectiveness of the educational systems, using the international patterns as a parameter and drawing ambitious horizons for how to formulate public policy in all the areas in our country” (pp. 156-7).

Therefore, although the results, as expected, were poor, PISA played a high level strategic role in returning them to the actors, especially the teachers (Souza, 2005). Thus, in part, the teachers received pressure, as responsible parties, for future improvement. In this regard, the Basic Education Evaluation System (SAEB) made an effort to report the results in a way that would be understood by teachers. The Minister used to argue that the improvement in the academic results does not depend on the resources, given that certain Federal states have more resources, but similar results.

DISCUSSION: IBERIA-AMERICA AND PISA

The question of why Latin America scores so low on PISA can have different answers depending on the ideological position from which they are perceived and formulated. Thus, from a neoliberal perspective, Jeff Puryear (2007) points out:

There is no single answer, and most Latin American countries clearly can't expect to do as well as wealthy countries such as Finland any time soon. Still, it is telling that Latin America has failed to put into place a number of components that are common in the world's high-performing school systems. These include: world-class standards (especially in reading and math); incentives to get and keep first-rate teachers; mechanisms to make teachers effective instructors; special attention to students who fall behind; and universal pre-school.

But it is senseless to compare the performance of the Latin American systems – plagued by gaps of inequality— with the performance of systems like Finland and the other more socially egalitarian countries. Latin America will gain nothing right now by imposing high standards –internationally competitive— if they are not accompanied by an “internationally competitive” spending, especially for poor

students, and if these students are not supported, together with their families from day one.

The investment in education in Latin American, even though it has increased in the past few years (before the current crisis), is insufficient. The percentage of GNP spent on education now represents about 4 %, a level similar to that of the OECD countries. But the spending per student is five times lower in Latin America, as the school age population represents between a quarter and a third of the total, compared with less than a fifth in the OECD. It is necessary to improve both the quantity and the quality of the public goods and services. Spending more is important, but the way this money is spent is even more relevant. The Latin American governments continue to invest little in those policies that have a greater impact on the performance of the students. If a greater expenditure in education is quite important, when there is little money available, improving the quality of this spending becomes even more important, making it more effective and better focused.

Latin America and the Caribbean have made (the same year as the third PISA, in 2006) their own evaluation of the performance of Primary level students, in one of the most ambitious attempts, known as the *Second Regional Comparative and Explanatory Study* (SERCE), carried out by the Latin American Laboratory for Assessment of the Quality of Education (LLECE) of the UNESCO (2008). Seventeen Spanish-speaking countries participated with third and sixth grade primary students in the areas of mathematics, language and sciences. A good analysis of the results, as well as their implications for educational policy and classroom practices, can be seen in Murillo and Román (2009) and Llece (2010). In this sense, SERCE offers information that complements PISA, as can be observed in the joint analysis performed by the OEI (2008). Although the influence of the social, economic and cultural environment of the student and the school increases with the grade level, PISA and SERCE also show that good work by teachers and schools is a determining factor in improving the education of all young people.

In the context of the commemoration of the 200 years of Latin America's independence from Spain, the General Secretariat of the Organization of Ibero-American States for Education, Science and Culture (OEI) has proposed a collective commitment in education to deal with the current challenges and demands of the Iberian-American people and as a way of betting on the future. Thus, the project "*Education Goals 2021: the Education that we want for the Bicentenaries' Generation*" arose; twenty-seven specific goals were established which, in turn, were specified in the form of 38 indicators. These indicators express the criteria that will be used to evaluate the advances made toward the goals, which are expected to be reached in 2021. The achievement levels are formulated in different degrees in order to adapt them to the initial situation of each country. An important future task will be for each country to define the level of specific achievement it plans to reach (OEI, 2008).

However, if it is clear that PISA has become the most advanced and complete system of international evaluation to date, it is also true that it is limited to a selection of three basic competencies, so that, according to Schleicher (2007: 350),

A. BOLÍVAR

“it is obvious that PISA cannot capture the *entirety* of competencies that will make young people successful”. The DeSeCo project, also funded by the OECD, wanted to provide a comprehensive view of competencies “for personal, social and economic well-being”, but up until now it has not been reflected in PISA. In the work mentioned above, Schleicher recognizes that PISA will not be able to evaluate interpersonal dimensions of competencies, which are of increasing importance, such as the ability of students to relate well to others, to manage and resolve conflicts, or to respect and appreciate different values, beliefs or cultures. Therefore, a serious limitation is that it only evaluates part of what is taught in school. Furthermore, there are other basic dimensions where PISA does not enter: coverage, level of long-term efficiency, evaluations of students, teachers and schools. In a clearly coordinated way, each of them must fulfill its role in order to improve the educational quality.

Ibero-America has gradually assumed more responsibility for the results of the students. However, on the whole, these results have not improved progressively. If broadening the schooling to the entire population was the goal of the 1980's and part of the 1990's, the current challenge, as well as universalizing Secondary education (World Bank, 2005), is to improve the quality, understood as providing the entire population with the necessary skills. On the whole, Latin America, on the successive tests (both TIMMS and PISA), obtains lower results than the countries in Europe and Asia. In addition to revealing great distances from other OECD countries, the Latin American countries present more unequal distributions. The reforms undertaken in the past few decades, once the entire population had access, have not been able to affect the hard core of teaching: the qualitative improvement of the teaching in the classroom. In one context, in some cases, quite focused on the country itself or with populist governments, an external international test like PISA enjoyed a legitimacy that served to question the management of the respective education systems, demanding quality focused on educational achievement in competencies.

The commentary about PISA has been mobilized to legitimize a certain educational policy. Rather than extracting lessons from the results, like –for example– which methodological changes could favour which language and which spelling, again learning the competencies, any changes made have been based on a prior political, ideological or educational position, and the data from PISA have been instrumentalized toward this end. The data have been used to justify the changes made or to provide support for educational policies already in place. Instead, the diffusion of the reports should contribute to a “rationalization of the public action” in education (Maroy & Mangez, 2008).

REFERENCES

- Afonso, N. & Costa, E. (2009). *Use and circulation of OECD's "Programme for International Student Assessment" (PISA) in Portugal*. Knowledge and Policy Project. <http://www.knowandpol.eu>
- Barroso, J. (Ed.). (2006). *A regulação das políticas públicas de educação: Espaços, dinâmicas e atores*. Lisboa: Educa-Unidade de I&D de Ciências da Educação.

- Baroso, J. & Carvalho, L. M. (2008). PISA: un instrument de régulation pour relier des monde. *Revue Française de Pédagogie*, 164, 77-80.
- Bolívar, A. (2008). Alto rendimiento y equidad. Conseguir un sistema equitativo sin bajar el nivel de los resultados. *Critica*, 956, 33-37.
- Cariola, L., Cares, G. & Lagos, E. (2009). *¿Que nos dice PISA sobre la educación de los jóvenes en Chile?* Santiago de Chile: Ministerio de Educación. Unidad de Currículum y Evaluación.
- Carvahlo, L.M. (2009). *Production of OECD's Programme for International Student Assessment (PISA)*. Knowledge and Policy Project. <http://www.knowandpol.eu/>
- Ferrer, F.; Ferrer, G. & Massot, M. (2006). *Percepciones y opiniones de la comunidad educativa sobre los resultados del proyecto PISA*. Madrid: CIDE/MECD.
- Figazzolo, L. (2009). *Testing, ranking, reforming: impact of Pisa 2006 on the Education Policy Debate*. Brussels: Education International.
- GIP (Grupo Iberoamericano de PISA)-OECD (2009). *Iberoamérica en PISA 2006. Informe regional*. Madrid: Santillana.
- IE (Instituto de Evaluación) (2007). *PISA 2006. Informe español*. Madrid, MEC.
- Kotthoff, H.-G. & Pereyra, M.A. (2009). La experiencia del PISA en Alemania: Recepción, reformas recientes y reflexiones sobre un sistema educativo en cambio. *Profesorado. Revista de Currículum y Formación del Profesorado*, 13(2), 1-24.
- LLECE (2010). *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe*. Santiago de Chile: Unesco y Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.
- Mahon, R. & McBride, S. (Eds.) (2009). *The OECD and transnational governance*. Seattle: University of Washington Press.
- Maroy, Ch. & Mangez, C. (2008). Rationalisation de l'action publique ou politisation de la connaissance? *Revue Française de Pédagogie*, 164, 87-90.
- Martínez Rizo, F. (2006). PISA en América Latina: lecciones a partir de la experiencia de México de 2000 a 2006. *Revista de Educación*, Núm. Extr., 153-167.
- Massot, M., Ferrer, G. y Ferrer, F. (2006). El estudio PISA y la comunidad educativa. Percepciones y opiniones sobre el proyecto PISA 2000 en España. *Revista de Educación*, núm. Extr., 381-398.
- Mortimore, P. (2009). *Alternative models for analysing and representing countries' performances in Pisa*. Brussels: Education International.
- Murillo, J. & Román, M. (2009). Mejorar el desempeño de los estudiantes de América latina algunas reflexiones a partir de los resultados del SERCE. *Revista Mexicana de Investigación Educativa*, 41, 451-484.
- Motter, P. (2008). *The Role of the media in educational policy formation and legitimation in Brazil: 1995-2008*. University of Wisconsin-Madison, Dissertation.
- Noaksson, N. & Jacobsson, K. (2003). *The production of ideas and expert knowledge in OECD*. Stockholm: SCORE (Score Rapport Series). www.score.su.se/pdfs/2003-7.pdf
- OEI (Organización de Estados Iberoamericanos) (2008). *Metas educativas 2021. La educación que queremos para la generación de los Bicentenarios*. Madrid: Secretaría General de la OEI. Disponible en <http://www.oei.es/metas2021/indice.htm>
- Ortega Díaz-Ambrona, J.A. (2007). La educación: el cuento de nunca acabar. *El País*, 13/12/07.
- Puryear, J. (2007). Monthly "Human Capital" column. *Latin America Advisor*, December 17.
- Ravela, P. (2006). Los resultados de las evaluaciones educativas en la prensa. In P. Arregui (Ed.) *Sobre Estándares y Evaluaciones en América Latina* (pp. 295-384). Santiago de Chile: PREAL.
- Rodrigo, L. (2009). *El bajo rendimiento escolar de los estudiantes argentinos en el programa PISA. Un análisis de los factores explicativos en perspectiva internacional*. Madrid: Universidad Complutense, Dissertation.
- Schleicher, A. (2006). Fundamentos y cuestiones políticas subyacentes al desarrollo de PISA. *Revista de Educación*, Special Issue, 21-43.
- Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, 8(4), 349-357.

A. BOLÍVAR

- Souza, P. R. (2005). *A revolução gerenciada: educação no Brasil, 1995-2002*. São Paulo, SP: Prentice Hall.
- Souza, P.R. (2006). Avaliação a serviço da qualidade educativa. In E. Santos (Ed.), *Reescrevendo a educação: propostas para um Brasil melhor* (pp. 148-168). São Paulo: Editoras Ática e Scipione.
Disponível: <http://oei.es/pdfs/reescrevendo.pdf>
- Tröhler, D. (2010). Harmonizing the Educational Globe. World Polity, Cultural Features, and the Challenges to Educational Research. *Studies in Philosophy and Education*, 29(1), 7-29.
- UNESCO (2008). *Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer Reporte SERCE*. Santiago de Chile: OREALC/UNESCO
- World Bank (2005). *Expanding opportunities and building competencies for young people*. Washington: The World Bank.
- Zorrilla Alcalá, J.F. (2009). El efecto de los resultados de PISA sobre la política educativa de México. En A. Teodoro y A. Montané (Coords.). *Espejo y reflejo: Políticas curriculares y evaluaciones internacionales*. RIAIPE Ibero-American Network of Educational Policies

Antonio Bolívar
University of Granada

SECTION II

PISA AND SCHOOL KNOWLEDGE

DAVID C. BERLINER

THE CONTEXT FOR INTERPRETING PISA RESULTS IN THE USA

*Negativism, Chauvinism, Misunderstanding, and the Potential
to Distort the Educational Systems of Nations*

The context in which results from PISA tests (the Program for International Student Achievement) are interpreted is exactly the same as that for the results of TIMSS (Trends in International Math and Science Study), or the results of the NAEP (the National Assessment of Educational Progress) and our many state and national test of achievement.

The context for score interpretation is an uncritical acceptance of the scores on the tests as valid indicators of

- a) what goes on in classrooms,
- b) how good teachers are,
- c) how good the school or district or state or nation is, and
- d) what the future holds for the examinee, the state, or the nation.

It must be pointed out, however, that there is little or no evidence that the tests we commonly use in any of these assessments have the powers attributed to them. It is not far from the truth to call the scores “*talismanic*” (Haney, Madaus & Kreitzer, 1987). That is, for many people test scores have special powers, particularly of prophesy.

It is in this context of trust in the validity of the scores reported that almost all newspapers, politicians, business people and school administrators bemoan the poor performance of America’s children on the tests. The nation, a state, or some school district is in dire trouble whenever the scores are released *no matter what the scores are*. Since almost all the tests are norm-referenced about 50 percent of the test takers will appear to the general public to be below average. This is, of course, a state of affairs that is not to be tolerated, even if it is true by design and stipulated through definition.

ANALYZING THE INTERPRETIVE CONTEXT IN 2006

The 2006 PISA study looked primarily at scientific literacy but the interpretive context for those results would have been the same were the scores for reading and mathematics. The news about US performance on this particular science test was greeted almost with glee by newspapers and politicians of all affiliations.

*M.A. Pereyra et al. (eds.), PISA Under Examination: Changing Knowledge,
Changing Tests, and Changing Schools, 77–96.
© 2011 Sense Publishers. All rights reserved.*

D.C. BERLINER

Newspapers apparently like bad news, not good news, and that is how the 2006 results looked to the press.

There are politicians of both major parties in the USA who do not like public schools because public schools require taxes that corporations and many citizens do not like to pay. Our allegedly poor performance is, to those politicians and citizens, just more proof that public education is a failure and that choice in schooling (more charter, voucher and private schools) needs to be promoted. Many of the critics of public education have divestiture in mind. They would like to see a small public school system and a much larger private school and charter system develop.

Many of these same test interpreters look at the US performance vis-à-vis other nations and demand a national system of standards because some top-scoring nations on PISA have national standards. A project of this kind has just been completed in the USA and states are being pressured into “voluntarily” complying with the federally promoted national standards and accompanying testing program. The logic of this endeavor is befuddling since many of the poorest scoring countries in the world also have national systems of standards, so this policy clearly is not likely to be a magical solution for the allegedly poor performance of US students. Furthermore, the Organization for Economic and Development (OECD) has discovered that school principals who have more autonomy in choice of curriculum and textbooks have schools that perform better on PISA tests than those that do not (Pont, Nusche & Moorman, 2008; Pont, Nusche & Hopkins, 2008). So the USA, by building a linked system of standards, textbooks, and tests, is now developing a homogenous system of education in which authority for some decision-making will be taken away from its educational leaders. This is the opposite response to that suggested by the researchers who design the tests on which the US wants to do better. Sadly, current US policy is reminiscent of the South Sea cargo cults at the beginning of the 20th house style on all numbers century. At that time, those who wanted certain goods (food, radios, clothing, etc.) engaged in bizarre ritualistic behavior that resembled that of European and Japanese visitors, those who possessed the coveted goods.

Not only did we in the USA score lower than the Finns and a dozen other nations, but some of the nations that beat us are, in American thought, “piddly little nations,” adding insult to injury for the politicians and news columnists. “Really,” say some of these lawmakers and influential writers with no sense of geography or history, “how can we be beat by Lichtenstein, Slovenia or Kyrgystan”? These are places most Americans never heard of! Clearly American chauvinism is part of the framework for interpreting PISA scores.

A Receptive Environment for Accepting Test Results

People in the USA, in general, have a positive opinion about the role of testing so they take test results from PISA very seriously. Part of that positive attitude is because a culture of competitive sports has developed in the USA. The testing of children, schools, and nations promotes the desired competition and fans, even rooting, develop in exactly the same way as that occurs in football, soccer, or

tennis. It is not coincidental that a “test” is what they call a cricket match in England and its former colonies. This competitive culture in the US has combined with a strong influence on the schools by American business. Both trends – competitiveness and a business mindset – bring to discussions about schooling a firm belief in the value of quantitative indicators to monitor school processes. Batting averages, goals scored, time in the 100-meter dash, widgets produced per hour, and so forth, are then seen as similar to PISA scores. Thus there is an acceptance of accountability systems that are highly quantifiable and there is faith that numbers from tests are as believable as those derived from games of cricket, football, and from monitoring the stock market.

Fears About Test Results

In the interpretive context for PISA scores in the USA, the worriers repeatedly say that our competitiveness will be hurt: that the nation, as we know it, will cease to exist. This is particularly galling to many Americans because we are brought up to believe that we are somehow entitled to be number one in sports, economic productivity, the arts and sciences, our standard of living, our material possessions, health care, the proficiency of our military, and so forth. In the minds of many politicians, the press, and the business community, America should expect to be number one forever. That position is a god given right. Some version of this kind of chauvinism has characterized every recent president’s messages about US scores on international tests.

Held along with those beliefs is the surety that the schools are an important mechanism to ensure a good life for the children of American citizens, the surest way to rise in social class standing. In this interpretive context it is easy to manipulate the public when scores come in far from the top, as do the USA scores on PISA. Fear is easy to stimulate under these conditions and it is why, in part, PISA gets so much attention. With our illogical belief that we should do well in everything, our repeatedly modest scores on PISA shake American beliefs in our superiority in all things.

Our modest scores perpetuate the myth of a failing school system and a failing nation, a position now held by many Americans. David Brooks, an influential columnist with the New York Times, our most influential newspaper said recently:

America’s lead over its economic rivals has been entirely forfeited, with many nations surging ahead in school attainment. ... The skills slowdown is the biggest issue facing the country. ... [t]his slow-moving problem, more than any other, will shape the destiny of the nation (Brooks, 2008).

Tom Friedman, also an influential New York Times columnist and author of the bestselling *The World is Flat* (2007) said:

In the 2006 Program for International Student Assessment that measured the applied learning and problem-solving skills of 15-year-olds in 30 industrialized countries, the U.S. ranked 25th out of the 30 in math and 24th in science.

D.C. BERLINER

That put our average youth on par with those from Portugal and the Slovak Republic, rather than with students in countries that are more relevant competitors for service-sector and high-value jobs, like Canada, the Netherlands, Korea, and Australia (Friedman, 2008).

Because there is pressure for newspapers and newscasters to respond to PISA data releases immediately, and because such hurried responses are also accompanied by the need to sensationalize headlines in the newspapers and produce sensational leads on TV, the reporting of US performance on PISA is always a problem. More critical and positive analyses of the PISA results always come later when greater thoughtfulness can be used in the interpretation of data. When PISA data are finally digested and understood better, and less sensational conclusions are reached than those that occurred with the original reporting of results, those analyses are *rarely* reported. It is argued by the press that the latter analyses are too complex, that the reports are too late to be newsworthy, and the analyses may even provide “good” news, which, therefore, is not news

What Was not Reported

On the 2006 Science scales some things *did not* make the news. The OECD average was 500 and Finland scored at 563, Canada at 534 and Japan at 531 (OECD, 2009). The US scored at 489 and the worriers began their triennial PISA warnings of dire results for the nation. Hardly any of the persons who interpret the findings for national consumption noticed that the US had 10 percent of its 15 yr olds scoring in the top two categories of performance, while the OECD average was 9%. Ten percent out of over 50 million public school children suggests that the economy can be well supplied with gifted science students. This number is more than enough to meet the needs of the US labor force, which for some time has been unable to absorb all the scientific talent graduated from our schools.

Finland had 21% of its 15 year olds in these top groups, and the USA had only 10 percent, so we have reasons for wanting to do better. But the numbers involved in the two nations are vastly different. There are about 1.4 million youth in school in Finland at all levels, and 74 million youth in school in the US at all levels. That means that Finland has 240,000 scientifically talented youth while the USA has about 7.4 million scientifically talented youth. And since the US has been one of the world’s leaders in the percent obtaining a college education over the last 50 years, we can assume that a huge additional pool of scientifically talented people is already in the work force.

So sometime after we learn about our mediocre performance when PISA results are released, cooler heads prevail and we learn that we are in no danger of being without scientific talent. In fact, we have the opposite problem, finding jobs for all this talent. But neither the size of the talent pool nor the problems of obtaining employment is mentioned much in the first PISA press reports.

When we look at the distribution of 15-year-olds who score at level 5 and 6 in the PISA nations, the pool of scientific talent among youth in most of the

industrialized nations of the world, the USA looks remarkably like the powerhouse it aspires to be (See [Table 1](#)).

The US has 25% of the worlds’ most talented youth (OECD, 2009, p. 21). That number should be of some consequence if American business and the American government can find ways to provide work for that pool of talent. The cooler heads often notice that well paying jobs for talented youth who work hard to master various fields in science and engineering are crucial to keeping our talent focused and productive as scientists. But the US has not been doing that. Lowell & Salzman (2007) noted that in 2001 the USA produced 758,300 Bachelor’s degrees in science and engineering. But when surveyed afterwards, one-third of these technically sophisticated people were neither employed in engineering or science, nor were they continuing their education. Of the 160,100 Master’s degrees given in science and engineering that year, the percent *not* employed in engineering or science, *nor* continuing their education, rose to about two-thirds!

Benderly (2010), in a premier science journal, quotes experts who say: “There is no scientist shortage in the USA.” The great lack in the American scientific labor market is *not* top-flight technical talent but attractive career opportunities for the approximately 30,000 scientists and engineers, about 18,000 of them American citizens, who earn PhDs in the U.S. each year.

Table 1. The percentages of top performers (level 5 and 6 in PISA testing) in different nations and across different economies (OECD, 2009, p. 21, from the PISA 2006 Database)

Nations with .5% or less of the top performing youth: Hungary, Turkey, Israel, Chile, Slovak Republic, Denmark, Norway, Mexico, Greece, Portugal, Slovenia, Thailand, Lithuania, Argentina, Croatia, Bulgaria, Estonia, Latvia, Romania, Columbia, Indonesia, Serbia, Jordan, Uruguay, Macao/China, Iceland, Luxembourg, Tunisia, Lichtenstein, Qatar, Azerbaijan, Krygyzstan, and Montenegro.	6% of the total of talented youth
Nations with 1% of the top performing youth: Austria, Switzerland, New Zealand, Sweden, Brazil, Hong Kong/China, Belgium, Finland, Czech Republic, and Spain	10 % of the total of talented youth
Italy	2%
Netherlands	2%
Australia	3%
Poland	3%
Taipei	3%
Canada	4%
France	5%
Korea	5%
Russian Federation	6%
United Kingdom	8%
Germany	8%
Japan	13%
United States	25% of the total of talented youth

Despite what is reported and believed, even by President Obama, there appears to be no shortage in the supply of scientifically talented American students and, what is more, our students' academic achievement has been increasing rather than declining in recent years.

Rarely noticed by politicians and news reporters concerned with our mediocre PISA performance, is that the students emerging from America's K-12 system are studying science and math subjects more, and performing better than they have in years. In fact, the number of Americans earning PhDs in science and technical fields has risen by 18 percent since 1985

Cooler heads also note, that with only a few exceptions, the differences between nations is really quite small when you look at the number of items answered correctly by each nations' 15 year olds. Most nations are bunched together in terms of raw scores. It is through scaling that nations close in raw score are made to appear far apart from each other on the scaled scores, so that a normal distribution is approximated. An example of this can be found in the 6 items released from a previous PISA study, as shown in [Table 2](#).

Table 2. Percentage of students correctly answering each of the PISA released questions (Lowell & Salzman, 2007, p. 21)

<i>Question Number</i>	<i>#1</i>	<i>#2</i>	<i>#3</i>	<i>#4</i>	<i>#5</i>	<i>#6</i>
<i>U. S. performance (scale score = 483)</i>	62.7%	74.6%	20.2%	67.8%	37.2%	39.8%
<i>OECD performance (scale score = 500)</i>	68%	72.9%	25.4%	73.9%	40.3%	32.2%
<i>U. S. performance Gap or advantage on a particular PISA item</i>	-5.3%	+1.7%	-5.2%	-6.1%	-3.1%	+7.6%

We see here how the US fared against the OECD nations. What is noted is that the US "gap" in performance on these 6 items, when weighed against the OECD average of all its students, is under 2%. This may result in scale score differences that will appear larger, but if the six items released are any indicator, then the USA is not nearly as poor a performer as it was made out to be. A 2% gap, compared to the OECD average, surely does not signal the end of prominence for the USA.

Social Class and PISA

In interpretations of PISA and other tests, the racial, ethnic and social class breakdown of scores get some attention, usually as a cause for alarm. The large difference in scores between wealthier and poorer students, however, is used to continue the attacks on teacher and administrator ineptitude. This is peculiar because PISA has been clear about what the test scores are indicators of: PISA is not an assessment of what young people learned during their previous year at

school, or even during their secondary school years. It is an indication of the learning development that has occurred since birth.

Improving quality and equity therefore require a long-term view and a broad perspective. For some countries, this may mean taking measures to safeguard the healthy development of young children, or improving early childhood education. For others, it may mean socioeconomic reforms that enable families to provide better care for the children. But in many, it can mean efforts to increase socio-economic inclusion and improve school offerings (OECD, 2003, p. 195).

Thus we see that every three years the US makes the wrong interpretation of its modest performance on PISA. The PISA test is, by design, *not* an evaluation of a country's school system. Unlike TIMSS or NAEP it is purposefully unrelated to the various curricula of the various nations. It is, instead, a reflection of a number of factors, many of them out-of-school House style on full stops. Here factors that affect school achievement. (Berliner, 2009). This is why Feniger, Livneh & Yogev (2007) were able to predict national scores on PISA from only two variables, neither of which had anything to do with quality of education, per se. They discovered that the variance accounted for in the PISA rankings was easily explained by gross domestic product per capita and the percentage of youth in the population. The multiple regression they used yielded an R^2 of .82. So PISA scores are seen in that analysis as reflecting the money a nation has to spend and the demands for that money by the percent of youth in need of education. High per capita GDP and low birth rate should yield a high PISA score. Finland and a number of other countries fit that bill quite well. The USA has a lower age population relative to other developed countries, probably due to immigration, often younger Latinos.

But more important for interpreting test performance is the degree of inequality in wealth within nations. Wilkinson & Pickett (2009) have documented that within wealthy nations inequality is greatest in the USA. As we like, we are number one. Their data also make clear that on a large set of variables that the USA might want to be better at, we are not. On an index made up of mathematics and literacy scores, as well as life expectancy, infant mortality, homicide rate, imprisonment rate, teenage birth rate, trust among the population, obesity, mental illness (including drug and alcohol addiction), and social mobility, the USA is the worst scoring among this group of relatively wealthy nations. The US scores are likely to remain modest or low in international comparisons not because of the quality of its teachers and administrators, necessarily, but because of its distribution of wealth and poverty and the associated social capital that exists in schools for the rich and poor. The inadequate or excellent performance of children within schools may well be a function of the social and intellectual capital that is present or lacking in schools that serve rich and poor children, an interpretation not at all at odds with the quote from OECD about PISA, given above. In fact, PISA scores were looked at from this perspective in Australia (Perry & McConnery, 2010).

The investigators categorized schools by quintiles based on both the average incomes of the children's families, and the average family income of the students in the school as an aggregate. The result is a 5 x 5 table as seen in [Table 3](#).

Table 3. Science scores on PISA in 2003 for students of various social classes in schools of varying social class standing (Perry & McConney, 2010)

Individual Student's SES	The SES that predominates at the school				
	<i>1st quintile</i>	<i>2nd quintile</i>	<i>3rd quintile</i>	<i>4th quintile</i>	<i>5th quintile</i>
<i>1st quintile</i>	455	457	471	497	512
<i>2nd quintile</i>	483	493	501	528	540
<i>3rd quintile</i>	496	500	512	541	558
<i>4th quintile</i>	520	524	531	557	577
<i>5th quintile</i>	555	544	550	582	607

Australia's average on this PISA science test hides incredible variation by the social class that predominates in the school and the social class of the students themselves. A lower class student in a lower class school (cell 1, 1) scores 455 in the PISA exam, but were that same child to get to a school that serves mostly upper class children (cell 1, 5) that child would likely score 512, over half a standard deviation higher. And were a wealthy student attending a school with mostly poor children (cell 5, 1), that child would score 555, rather than a score of 607 were he or she in a school that had mostly wealthy children (cell 5, 5). This also is a difference of half a standard deviation. The diagonal (cell 1,1 v. cell 5,5) shows that the greater the variation in a students family social class *and* the schools' predominant social class, the greater the disparity of the scores, in this case a difference of over one and one-half standard deviations! Reading and mathematics PISA test scores show the same pattern.

If the Australian data holds across countries, as I suspect it does, it informs us that the greater the number of poor children in a country, and the greater the number of schools that predominantly serve the poor, the lower the PISA scores will be. Conversely, were income distributed more evenly and access to schooling more open to members of different social classes, PISA scores for a nation would likely be higher. So in the end, PISA can be said to inform the USA of its troublesome income distribution and its troublesome housing segregation, resulting in schools that serve the poor, poorly, and the rich separately, and quite well. Of little interest to those who want schools blamed for poor PISA performance is that the number of schools attended by children whose families are impoverished has increased in the USA. Impoverished schools are those with 75% or more of its children eligible for poverty programs. The recent increase was from 12 percent of the schools in the USA having 75% or more of its student body impoverished in 1999–2000, to 17 percent attending such schools in 2007–08. (Aud et al., 2010). It is hard to imagine the international test scores of the USA getting any better with

these kinds of trends in the segregation of schools for the poor and the growth in the number of such schools. Most interpreters of PISA in my country seem oblivious to these issues. They seem not to notice that childhood poverty rates in the USA are around 25% while childhood poverty rates in Finland are about 3%. Failure to notice this motivates politicians and the business community to continue their belief that better teachers and administrators are the solution to the low achievement that they see. Such perseverance in the face of contradictory data is inappropriate. PISA 2006 results, reported in [Table 4](#), show why blaming of school personnel is so inappropriate.

Table 4. PISA 2006 scores on the Science Literacy Scale for US students, by race and ethnicity

Average of the OECD countries	500
White US students	523
Students of more than one race	501
Asian students	499
Native Hawaiian and Pacific islander students	483
American Indian students	436
Hispanic students	439
Black students	409

There, science scores are seen to be dramatically affected by race and ethnicity, strong correlates of social class in the USA (OECD, 2006). White US students still constitute over 50 percent of the students in the US public schools, though they will soon be a minority. But they do total about 25 million students. It is quite clear from these results that they are *not* scoring low in comparison to the OECD average. The problem the US has is with the gaps between the white students and some others. White students score well over one standard deviation higher than do black students, and almost a standard deviation above the scores of Hispanic and American Indian students. It is the large and poor black, Hispanic, and American Indian population of the US that brings the US averages on PISA tests down to their modest levels. Data like these suggest quite strongly that PISA is measuring the effects of social class on school achievement and may not be an indicator of much else.

For those that worry about the US it is worth pointing out that these 25 million undifferentiated white students scored 6th in the world in Science in the 2006 PISA tests reported in [Table 4](#). They tied Japan, one of the feared economic competitors to the USA. Add to this success the fact that America's private school children, totaling 6 million more students, are overwhelming white. Those students are likely to do as well or better on these kinds of tests were they to participate. Thus the talent pool in the US is enormous, despite the endless cries of those who see PISA scores as indicators of America's doom.

This relationship is even clearer when the TIMSS data is examined, as presented in [Table 5](#) (Gonzales et al., 2008).

The TIMSS average score for the USA was 529 in mathematics, but in schools where less than ten percent of the children were considered to be in poverty the

D.C. BERLINER

score was 583: only two nations out of 58 nations performed better. In schools where poverty among the students was between 10 and 24.9 percent, the score was 553: These students placed 4th out of 58 nations. Together these schools constitute

about 31 percent of our public school population, representing about 15 million students of all races and ethnicities who are lucky enough to attend schools that have low rates of poverty. These 15 million students out- performed students in 54 other nations in mathematics. On the other hand, in schools that had poverty rates of 75% or more, the TIMSS score was 479, about a full standard deviation lower than were the scores in schools that served the more affluent.

Table 5. Average mathematics scores of U. S. fourth grade students on TIMSS 2007, by percentage of students in public school eligible for free and reduced lunch (Gonzales et al., 2008)

<i>Percentage of students eligible for free and reduced lunch (a measure of poverty in the schools)</i>	<i>2007 TIMSS average mathematics score at grade four</i>
Less than 10 percent	583
10 to 24.9 percent	553
25 to 49.9 percent	537
50 to 74.9 percent	510
75 percent or more	479
U. S. Average	529
Average for 58 nations in TIMMS	500

TIMSS like PISA are tests that are highly sensitive to social class, limiting the inferences that can be made about many aspects of schooling. Invalid inferences, however, are common whenever the results of these tests are released. Poverty and wealth determine much of the variation in achievement test scores.

The nurturance of achievement in children is examined in PISA, but these data are not reported much because in the USA the focus is on the test scores themselves. Questions were asked about the possession of eight items, including a desk to study, a quiet place to work, a computer for schoolwork, educational software, an internet connection, a calculator, a dictionary, and school textbooks. The proportion of children reporting less than four of these educational items is used as an index of how well a nation nurtures achievement in its children. Data from Finland showed that only 15% of its 15-year- olds had less than four of these possessions while the US rate is 4 times that, a whopping 44.8 %. But because the test score becomes the focus of attention, the cooler minds, the better analyses, and the alternative explanations have trouble surfacing when the scores are released.

Although the people of the USA think the PISA scores measure schools and schooling, almost everyone who digs deeper knows that the PISA scores represent schools and society *in interaction*. And it is my impression that even though PISA analysts know this, they play to those that believe schools and schooling processes are the major cause of score differences. They should, instead, be making sure that

THE CONTEXT FOR INTERPRETING PISA RESULTS IN THE USA

PISA scores are seen as a function of the social settings in which schools in various counties operate.

In summary of this section, the interpretive context for most of the comparative data we get is fear about jobs for themselves and their children, fear of facing the childhood poverty issues that are now quite severe, and fear about the future of the US economy blended with my nations' preoccupation in being number one and the chauvinistic attitudes associated with that belief.

THE CONSEQUENCES OF AN INTERPRETIVE FRAMEWORK BASED ON FEAR AND HYPERCOMPETITIVENESS

Were PISA score envy among nations to increase, as it appears to be doing, then PISA tests will become high-stakes tests for some nations and some political parties. High-stakes testing has been in effect in US states for many years and lessons have been learned.

The state level tests that began in earnest in 2003 as a consequence of the federal law called the *No Child left behind Act* (NCLB) are the ones that are most instructive. If a school does not make a certain amount of growth on these tests it can be taken over by the state, teachers fired and replaced, or the school can be closed. These are high-stakes tests: the consequences of not doing well are quite severe. Teachers and administrators in schools that make little growth are shamed and blamed for that state of affairs. Under such pressure it is not uncommon to find that teachers and administrators cheat, or they break standardization procedures associated with the tests. These common acts destroy the validity of the tests, rendering them useless (Beardsley, Berliner & Rideau, 2010; Nichols & Berliner, 2007).

Teachers also engage in vast amounts of test preparation with their students, some of which also makes the validity of high-stakes testing quite problematical. Teachers and administrators also move students that they expect not to test well out of the schools, or treat them poorly hoping they will drop out, or they hold them back so they get to be a year older and maybe a little smarter before they have to take the tests, and so forth. All this is fully documented (Nichols & Berliner, 2007). Teachers and administrators in high-stakes environments behave in accordance with Campbell's law, which states that any time a social indicator (such as a stock price or a test score) takes on too much value, both the indicator and the people who work with that indicator are corrupted.

Lessons from the USA and elsewhere about Campbell's law and issues of validity are relevant to PISA testing because the PISA results are taken so seriously. National scores have become political as worries grow about each state's economic future, despite the fact that the tests cannot predict those outcomes well. When countries pay too much attention to PISA, one can expect responses to that pressure to resemble what we have documented in the USA. Some people will do what ever is needed to keep their jobs and status. The assessment system will corrupt them, and the indicator used will quickly become invalid.

While many responses to the pressure of high-stakes testing are immoral and unlawful, I also found more subtle ways a high-stakes test and a national system of

education can be corrupted. I found a great deal of curriculum narrowing in the USA. This may be the most important lesson to be learned by nations competing for high achievement scores.

In the US the NCLB law requires that reading and mathematics scores be used to determine if a school is making appropriate growth, called Adequate Yearly Progress (AYP). The federal government does not monitor performance in other school subjects. So a rational response by teachers and administrators is to increase the time spent in reading and mathematics a lot.

Data from a nationally representative sample of school districts, are provided in [Table 6](#) (Center on Education Policy, 2008).

Table 6. Changes since 2001-2002 in instructional time for elementary school English language, Arts and Mathematics, in districts reporting increases (Center for Education Policy, 2008)

<i>Subject matter examined</i>	<i>Average total instructional time spent pre-NCLB (in minutes per week)</i>	<i>Average total instructional time spent post-NCLB (in minutes per week)</i>	<i>Average increase in instructional time per week (in minutes)</i>	<i>Average increase in instructional time as a percentage of total instructional time</i>
<i>English language Arts</i>	378	520	141	47%
<i>Mathematics</i>	264	352	89	37%

It is clear that changes in the time allocated for teaching reading and mathematics in elementary schools were quite dramatic between 2002 and 2007. These are the years of the NCLB act and mandated high-stakes testing. The time allocated to reading has been increased, on average, over two and a third hours a week, while mathematics time has been increased, on average, about an hour and a half a week. What needs to be kept in mind when interpreting this table is that the “average” masks relevant information. School districts serving low-income students probably changed their time allocations a lot, while those serving wealthier students probably did not change much at all since their students were likely to make AYP.

In a previous study of the curriculum by these same scholars (Center for Education Policy, 2006), 97 percent of the school districts *not* making Adequate Yearly Progress (AYP) had changed their curriculum times and content. These are, of course, the districts where the pressures to score higher on the tests are greatest and where you expect such a response to that pressure.

Many would say the changes documented above may not be all bad, unless there is reason to believe that the increased time is used poorly and is having either no effects or detrimental effects on students. This is where things get interesting. Evidence exists to support the hypothesis that the increased time spent on reading and mathematics is *not* helping the schools produce better readers and mathematicians.

On our highest quality national test of student achievement, the NAEP test, scores were rising in both reading and mathematics at a greater rate before NCLB than after NCLB (National Center for Education Statistics (2009a; 2009b). When you increase time and reduce the growth in learning, something is seriously wrong. It is as if a fundamental law of learning is being violated, like the violation of a law in physics. When time is increased and rate of learning decreases, we are left to think either or all of the following: our curriculum is no good, we have killed students' motivation to learn, or our tests are no good. But something is wrong.

Further, an analysis by Lee pointed out that the achievement gap between wealthier and poorer students has not been closing at all on the National Assessment of Educational Progress, the best audit test America has about how states are doing with regard to their NCLB goals (Lee, 2008). Others argue that the gap is closing, but very little (Braun, Chapman & Vezzu, 2010). Nobody seriously argues that the achievement gap is closing a lot! This all suggests that something is quite wrong. Perhaps the increased time for learning in reading and mathematics results in a less interesting curriculum for teachers to teach, and for students to learn. If that is so then the results we get are sensible, though certainly disheartening.

England has also tried to reform its schools through an accountability plan similar to that of the USA. And England seems to have the same problem of failed policy as does the USA. Following up his research of the 1970s, the British classroom researcher Maurice Galton found that teacher-centered pedagogy, characterized by interactions of a very low cognitive level, managerial in their intent, had *increased* dramatically between 1976 and 1996. Pupils had fewer opportunities to question or to explore new ideas after the tests became the primary instrument that the government used to change the schools (Galton, 2007). Assessment pressures have resulted in 42% of teachers' time being taken up with whole class teaching, compared to 18% in 1976. In primary schools in England there are now few opportunities for expressing anything that resembles creative reasoning (Galton, 2007).

Galton & McBeath (2002) surveyed primary teachers in England who regret that time pressures no longer allow them to engage in informal conversations with individual children during lessons, or to allow pupils, at certain times, to pursue their own ideas and interests as part of topic and project work. Since the seventies this kind of time has decreased by nearly 50%. Hong & Youngs (2008) report similar findings from the city of Chicago and the State of Texas as they too responded to high-stakes testing. In Chicago, as in Great Britain, high-stakes testing seemed to narrow the curriculum and make it harder for students to acquire higher-order thinking, writing, and problem-solving skills. In Texas, as in Great Britain, it was found that schooling changed in ways that emphasized rote learning, not broad intellectual skills (Hong & Youngs, 2008; McNeil, 2000). Lipman (2004) also studied the Chicago schools and reports that the accountability program ensured that the more affluent students in Chicago received a much richer and more intellectually challenging curriculum than did the poor children in Chicago. Poor minority children, in particular, were required to memorize fragmented facts and information, and they were constantly taught simple test-taking techniques.

Apartheid is descriptive of what has happened in many school systems across the USA (Kozol, 2005).

The Rest of the Curriculum

Time added for reading and mathematics needs to come from somewhere else in the curriculum. [Table 7](#) presents data on *decreased* instructional time in some curriculum areas to provide increased instructional time in the areas of reading and mathematics.

In interpreting this table it is important to remember that the average masks the bigger cuts that some districts have made in these subjects. With this caveat in mind we see that the teaching of social studies, intended always to be part of youth development for responsible citizenship, is down, on average, over an hour a week.

Yet Americans of all political persuasions ask that the schools help to develop citizenship. So this trend in curriculum is in opposition to the aspiration that all Americans have for their school curriculum. Furthermore, school activities that might foster citizenship have been cut because of the need for more time in reading and mathematics, so cleaning up neighborhoods and parks, visiting nursing homes, going on field trips to the legislature, projects that examine pollution at a local level, and so forth, all have been jettisoned (Nichols & Berliner, 2007). Yet is through such projects that citizenship is learned. Fear about the results of high-stakes testing is the culprit here. Science, a field that probably will be even more important in the 21st century than in the 19th and 20th centuries, is down, on average, over an hour a week as well. Schools can ignore a lack of growth on science tests because no sanctions attach to that test. Therefore, science, like social studies has been robbed of minutes to expand time for reading and mathematics. Thus curriculum that might help insure American economic competitiveness in the future, and surely will contribute to intelligent citizenship in our science- and technology-rich future, has been sacrificed for the possibility of scoring a bit higher on a high-stakes test.

[Table 7](#) also documents that time for physical education is down, despite the fact that America's youth are more sedentary than they should be, are quite overweight, and Type 2 diabetes is becoming more common. It is easy to argue that physical education is more important today than ever before, and it is acknowledged as one of the most important ways to keep a nations' medical costs down. Yet physical education is sacrificed for the possibility of a few more points on state tests that are required to rise continuously.

Nationally, as seen in [Table 7](#), recess was found to be down, on average, about an hour a week. Nichols & Berliner (2007) even discovered a superintendent of schools who forbid naps for preschoolers and kindergartners because test scores needed to go up.

Art and music, nationally, are down an average of an hour a week. This is particularly troublesome because the USA never spent a lot of time in these subjects. In California, for example, 89% of its K-12 schools fail to offer music, visual arts, theatre, and dance that meet the states' own standards of instruction in

these areas. In fact, 61% of California schools do not have even one full-time-equivalent arts specialist. (Woodworth, Gallagher & Guha, 2007).

Table 7. Decreases in instructional time for various curriculum areas to accommodate increases in time for English language arts and mathematics (Center for Education Policy, 2008)

<i>Subject matter</i>	<i>Average minutes per week before NCLB</i>	<i>Average minutes per week after NCLB</i>	<i>Average decrease per week</i>	<i>Average decrease as a percentage of total time per week</i>
Social studies	239	164	76	32%
Science	226	152	75	33%
Physical education	115	75	40	35%
Recess	184	144	50	28%
Art and music	154	100	57	35%

The defense of the arts can be made on many grounds, but one stands out in terms of the needs of the 21st century, namely, that the arts are alternative ways to represent reality. Ideas expressed through the visual arts, dance, and music are not presented in the verbal or mathematical symbol systems that are in everyday use. So by cutting the arts we limit the ways our students can represent the world in which our students live and about which they may choose to comment. A reduction in curriculum for learning the arts, therefore, restricts our students' ways of thinking, limiting creativity.

The defense of the arts in schooling can be based on economic and cognitive psychological reasons, which add to other reasons for defending the arts as a natural expression of our humanity, and for occasionally providing humanity with works of indescribable beauty.

Because of the test anxiety felt by the teachers and administrators of poor children, The study of the arts in California makes clear that the arts are rationed: They are taught primarily to the wealthy and the middle class, but not taught to the poor. These data are provided in [Table 8](#).

Table 8. Percent of California students receiving instruction in various areas of the arts, by poverty level of the school they attend (Woodworth, Gallagher & Guha, 2007)

<i>Subject matter</i>	<i>Percent studying this subject in schools serving wealthy children</i>	<i>Percent studying this subject in schools serving middle class children</i>	<i>Percent studying this subject in schools serving poor children</i>
<i>Music</i>	45%	38%	25%
<i>Visual arts</i>	48%	44%	29%
<i>Theater</i>	17%	14%	8%
<i>Dance</i>	17%	14%	7%

D.C. BERLINER

Close to twice as many students in schools that serve the wealthy receive instruction in the arts as do the students in schools that serve the poor. This is another example of the apartheid system of schooling that has developed in the USA.

Teachers and high-stakes testing

Although teachers' voices are often dismissed, surveys of teachers reveal how the NCLB high-stakes testing culture affects the content of their courses. For example in Colorado teachers say (Taylor et al., 2003):

We don't take as many field trips. We don't do community outreach like we used to like visiting the nursing home or cleaning up the park because we had adopted a park and that was our job was to keep it clean. Well, we don't have time for that any more (p. 30).

A Florida teacher says (Jones & Egley, 2004):

Our total curriculum is focused on reading, writing, and math. There is no extra time for students to study the arts, have physical education, science, or social studies. Our curriculum is very unbalanced.

Hundreds of similar quotes are now in the literature. I would estimate that over 75% of the teachers in the USA are unhappy with the current assessment policies. But thus far politicians have not listened much to their concerns.

WHAT DO AMERICANS REALLY WANT AS CURRICULA?

Rothstein, Jacobson & Wilder (2008) identified curriculum areas that administrators, teachers and parents all agreed were important. After basic skills, and rated almost as important, was critical thinking. But this major curricula goal for American education, perhaps never taught well, is made even less likely to be included in the school curriculum because it takes inordinate amounts of time to do well and that takes time away from preparing for tests.

Six other curriculum areas were all close to being tied for third place among desirable curriculum goals. One of these was about developing social skills. But with the current tests and the pressure they engender, project-based, inquiry-based, and problem-based cooperative work groups, where such skills are learned, are less frequently seen in our public schools. Instead classes have become more like workplaces where individual effort is expected and it is individual effort that is rewarded, despite the desires of the public and the predictions about the need for social skills in the modern workplace.

The public also wants the schools to develop a work ethic in youth. This is hard to do with many poor and minority students. Those students are often punished by the high-stakes testing programs because they do not have the social and intellectual capital of their middle class peers. In a high-stakes testing environment these poor and minority students quickly learn they are a liability to the schools

that they attend, that they “haven’t got the right stuff,” and that schools are about winners and losers. In such a setting, instead of developing a healthy work ethic, many of these students develop despair. They often disengage themselves from schoolwork, leaving school before they graduate (Amrein & Berliner, 2002; McNeil, 2008).

Other curriculum goals show the breadth of what Americans desire. They want schools to develop in students' personal responsibility, an ability to get along well with others, especially others from different backgrounds. Youth were thought to need knowledge of how government works and of how to participate in civic activities like voting, volunteering, and becoming active in communities. The survey respondents believed our students should receive vocational, career, and technical education that could qualify youth for skilled employment that does *not* require a college degree. The survey respondents also wanted the schools to provide a foundation for lifelong physical health, including good habits of exercise and nutrition. They wanted our schools to develop in our students a love of literature and the capacity to participate in and appreciate the musical, visual, and performing arts. Finally, in the area of emotional well-being, our students were thought to need tools to develop self-confidence, respect for others, and the ability to resist peer pressure to engage in irresponsible personal behavior.

Many of these American curriculum goals probably overlap with what is wanted by the citizens of other nations. But these curriculum goals are downplayed or even sacrificed in systems where teachers see their role as primarily imparting knowledge to score well on tests that are consequential to them. Those teachers and schools are failing to meet the expectations of citizens who want more than what is being offered in their schools.

CONCLUSION

PISA items appear to tap thinking and problem solving, and do provide a picture of a country’s ability to raise children who are healthy in mind and body. That is good. But if, as is often the case, too much value is placed on a national score, then I expect schools in that nation to respond as they have in the United States. Campbell’s law predicts school personnel will cheat or break standardization, they will fudge sampling, they will remove some students from the data set, teach special test prep units after seeing the tests, and so forth. But most of all they will try to provide the curriculum that will increase scores on the test. In the case of PISA this is reading, mathematics, and science. Pursuit of higher PISA scores means increasing time spent in areas assessed and that means there will be reductions in other curriculum areas in which we want youth to be educated.

It is clear from the survey data and from the teachers themselves that in high-stakes testing environments the tests determine the curriculum rather than the other way around. Nations worried about their PISA scores might well be wary of what we have documented in the USA. If not, national and regional differences in education will be lost and similarities in the educational systems of each nation will follow. This may not be desirable from many points of view, the most pragmatic

being that we are sure that the 21st century economy will require from our work force a broad set of skills, not a narrow one. And high-stakes testing leads to a narrowing of the skill sets that students possess.

Diversity in the outcomes of the educational system ought to be a goal of each nation's educational system, not sameness. The philosophy of *Bildung* in Germany, the belief in the Danish way in Denmark, the *Baccalauréat* and *L'Agrégation* traditions of France, the early childhood philosophy emanating from Reggio-Emilia, the continental educational traditions of familiarity with philosophic and moral thought, may all change if PISA score envy becomes more prevalent. Each nation must decide if this is good or bad.

The most important lesson from the USA is that high-stakes testing narrows what it means to be an educated person, and that is a shame. Such definitions should be as broad and encompassing of different talents as possible.

No one really knows what 21st century skills are needed to foster success for individuals and nations. But developing critical thinking, engaging in activities that require problem solving and creativity, and doing individual and collaborative projects of complexity and duration, are all good candidates for helping each child and each nation to thrive. We need to keep in mind what has been found from almost all longitudinal studies of youth. It is the soft skills that determine success every bit as much as literacy and numeracy when looking at outcomes such as college completion, earnings, and a host of other outcome variables later in life (Deke & Haimson, 2006; Lleras, 2008).

The skills tested in NCLB and in PISA are necessary, but they are not sufficient for predicting success *either for individuals or for nations*. Thus we need to cultivate our student's talents, *whatever they may be*, so that they learn how deliberate practice and success go together in sports, drama, student government, music, geography, computer illustration, computer gaming, fashion design, cooking and more. We need to remember that when administrators and teachers concentrate their efforts on promoting attainment of only a few skills, they detract from the talent pool for individual and national success will suffer – if we have uniformity of outcomes – in an international economy that will demand adaptability

It is certainly foolish to be against all assessment programs, as some in my country are. But when the stakes are high, it is not foolish to ask questions about validity. As I have shown, some of the inferences made from the PISA assessments are misguided if not fraudulent. It is not foolish to monitor the effects of testing programs on the personnel who live in high-stakes testing environments. As I have shown, some are corrupted and many more are disheartened. And it is not foolish to monitor the curriculum and instruction provided in high stakes testing environments. The citizens of all nations must be sure that the tests are not determining the curriculum, as is happening in my country. Each nation needs to insure that it is getting the curriculum they really want for their children, rather than one that attempts primarily to raise test scores to satisfy some politicians' ego. Without vigilance high-stakes testing can quickly distort a schools' and a nation's education system in ways that are not healthy.

THE CONTEXT FOR INTERPRETING PISA RESULTS IN THE USA

REFERENCES

- Amrein, A.L. & Berliner, D.C. (2002, March 28). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). Retrieved January 13, 2009 from: Retrieved January 13, 2009 from: <http://epaa.asu.edu/epaa/v10n18/>
- Amrein-Beardsley, A. Berliner, D. C. & Rideau, S. (2010). Breaking professional law: degrees of cheating on high-stakes tests. *Educational Policy Analysis Archives*, 18 (14). Retrieved June 28, 2011 from <http://epaa.asu.edu/ojs/article/view/714>
- Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., Frohlich, L., Kemp, J., Drake, L. (2010). *The Condition of Education 2010 (NCES 2010-028)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved April 28 from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010028>
- Benderly, B. L. (2010, February 22). *Does the U.S. produce too many scientists?* *Scientific American*. Retrieved May 20, 2010 from <http://www.scientificamerican.com/article.cfm?id=does-the-us-produce-too-m>
- Berliner, David C. (2009). *Poverty and potential: Out-of-school factors and school success*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved May 2, 2010 from <http://epicpolicy.org/publication/poverty-and-potential>
- Braun, H., Chapman, L., & Vezzu, S., (2010). The black-white achievement gap revisited. *Education Policy Analysis Archives*, 18(21). Retrieved May 31, 2011 from: <http://epaa.asu.edu/ojs/article/view/772/855>
- Brooks, D. (2008, July 29). The biggest issue. *The New York Times*. Retrieved May 25, 2009 from: <http://www.nytimes.com/2008/07/29/opinion/29brooks.html>
- Center on Education Policy (2008, February). *Instructional time in elementary school subjects. A closer look at changes for specific subjects*. Washington, DC: Author. Retrieved April 15, 2009 from <http://www.cepcdc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=234dc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=234>
- Center on Education Policy (2006, March). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author. Retrieved May 3, 2010 from: <http://www.cepcdc.org/>
- Deke, J. & Haimson, J. (2006, September). Expanding beyond academics: Who benefits and how? Princeton NJ: Issue briefs #2, *Mathematica Policy Research*, Inc. Retrieved May 20, 2009 from: http://www.eric.ed.gov:80/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/28/09/9f.pdf Mathematicapolicy research Inc.
- Feniger, Y. Livneh, I & Yogev, A. (2008). Globalization and the politics of international surveys of educational achievement: The case of Israel. *Paper presented at the meetings of the International Sociological Association*, Barcelona, Spain, September 5, 2008.
- Friedman, T. L. (2009, April 21). Swimming without a suit. *New York Times*. Retrieved May 25, 2010, from <http://www.nytimes.com/2009/04/22/opinion/22friedman.html>
- Friedman, T. L. (2007). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Strauss and Giroux.
- Galton, M. (2007). *Learning and teaching in the primary classroom*. London, UK: Paul Chapman Publishing.
- Galton, M. & McBeath, J. (2002). *A life in teaching? The impact of change on primary teachers' working lives*. London: National Union of Teachers.
- Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and Science achievement of U.S. fourth- and eighth-grade students in an international context (NCES 2009-001)*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Educational Statistics, p. 25. Retrieved April 2, 2010, from [http://www.usinnovation.org/files/TIMSS2007Highlights\(Dec2008\).pdf](http://www.usinnovation.org/files/TIMSS2007Highlights(Dec2008).pdf)
- Haney, W., Madaus, G. & Kreitzer, A. (1987). Charms talismatic: Testing teachers for the improvement of American education. *Review of Research in Education*, 14, 169-238.

D.C. BERLINER

- Hong, W.-P., & Youngs, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students? *Education Policy Analysis Archives*, 16(6). Retrieved January 14, 2009 from: <http://epaa.asu.edu/epaa/v16n6/>.
- Jones, B. & Egley, R. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing, *Education Policy Analysis Archives*, 12(39). Retrieved June 15, 2009 from: <http://epaa.asu.edu/epaa/v12n39>
- Kozol J. (2005). *The shame of the Nation: The restoration of apartheid schooling in America*. New York: Random House.
- Lee, J. (2008). Is test driven accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78(30), 608-644.
- Lipman, P. (2004). *High stakes education: Inequality, globalization, and urban school reform*. New York: Routledge Palmer.
- Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research*, 37, 888-902.
- Lowell, B. L. & Salzman, H. (2007). *Into the eye of the storm: Assessing the evidence on science and engineering education, quality, and workforce demand*. Washington, DC: The Urban Institute.
- McNeil, L., Coppola, E., Radigan, J. & Vasquez Heilig, J. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives*, 16(3). Retrieved May 1, 2008, from <http://epaa.asu.edu/epaa/v16n3>
- National Center for Education Statistics (2009a). *The Nation's Report Card: Mathematics 2009 (NCES 2010-451)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- National Center for Education Statistics (2009b). *The Nation's Report Card: Reading 2009 (NCES 2010-458)*. Institute of Education Sciences, U.S. Department of Education, Washington, D.C.
- Nichols, S. L. & Berliner, D. C. (2007). *Collateral damage: how high-stakes testing corrupts American education*. Cambridge, MA: Harvard Education Press.
- OECD. (2003). *PISA 2003. Technical Report*. Paris: OECD.
- OECD. (2006). *Program for International Student Assessment (PISA) 2006*. Paris: OECD.
- OECD. (2009). *Top of the class: High performers in Science in PISA 2006*. Paris: OECD.
- Perry, L. B. & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112 (4). 1137-1162.
- Pont, B., Nusche, D. & Moorman, H. (2008). *Improving school leadership, Volume 1: Policy and Practice*. Paris: OECD.
- Pont, B., Nusche, D. & Hopkins, D. (Eds.) (2008, August). *Improving school leadership, Volume 2: Case studies on system leadership*. Paris: OECD.
- Rothstein, R., Jacobson, R., & Wilder, T (2008). *Grading education: Getting accountability right*. New York: Teachers College Press.
- Taylor, G., Shepard, L., Kinner, F. & Rosenthal, J. (2003). A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost. *CSE Technical Report 588*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Wilkinson, R. & Pickett, K. (2009). *The spirit level: Why more equal societies almost always do better*. London: Allen Lane/ Penguin.
- Woodworth, K. R., Gallagher, H. A., & Guha, R. (2007). An unfinished canvas. Arts education in California: Taking stock of policies and practices. *Summary Report*. Menlo Park, CA: SRI International. 2007. Retrieved January 14, 2009 from: <http://policyweb.sri.com/cep/publications/AnUnfinishedCanvasSummaryReport.pdf>

David C. Berliner
Arizona State University

DAVID SCOTT

PISA, INTERNATIONAL COMPARISONS, EPISTEMIC PARADOXES

INTRODUCTION

Two forms of knowledge can be identified (let us call them K_a and K_b). K_a represents those knowledge sets, skills, and dispositional states of a person, collectively known as capacities. K_b represents those knowledge sets, skills and dispositional states which allow this person to do well in tests, and in particular, high stakes tests. K_a and K_b have different characteristics. If an education system introduces high stakes testing, that is, testing in which there are significant rewards attached to success in the test for an individual, an institution, or even a nation, then there are two consequences. The first is that K_b becomes the dominant form of knowledge in the curriculum and the second is that K_a over time is transformed so that it becomes more like K_b , that is, it has more of its characteristics. Testers commonly conflate K_a and K_b , and in doing so make a number of false assumptions about knowledge and its assessment, with the consequence that these two forms of knowledge become indistinguishable in the minds of policy-makers, educational practitioners, students and other stakeholders. Furthermore, knowledge of an individual's or a group's (i.e. nation, age-cohort or category) capacities, as in international comparative systems of testing such as the Programme for International Student Assessment (PISA) (OECD 2000, 2001, 2006, 2009), is underpinned by a particular and specific geo-historical notion of comparison.

An alternative to this position is provided by the philosophy of critical realism. Critical realists make three initial claims: there are significant differences between the transitive realm of knowing and the intransitive realm of being; the social world is systemically open; and researchers and observers need to grasp the ontological depth of reality. The first of these then, refers to a distinction between the intransitive world of being (the ontological realm) and the transitive world of knowing (the epistemological realm), so that to conflate them becomes illegitimate, either upwards, resulting in the epistemic fallacy, or downwards, resulting in the ontic fallacy (cf. Bhaskar, 1989). There are two implications of this. Social objects, though real, constantly change, and it is therefore the changing object which is relatively enduring, even to the extent that the object has been so utterly transformed that it is barely recognisable in relation to its former self. The second implication is more significant, and this is that, in certain circumstances and within certain conditions, social objects from the transitive realm can penetrate the intransitive realm and be objectified. What follows from this is that in principle the

D. SCOTT

measurement of the capacities of an individual or set of individuals can activate emergent properties of the construct being measured and change that construct.

This also suggests that a disjuncture can occur between the two realms, with the result that they become unsynchronised. Bhaskar (1989) identifies four reasons for this: there are social objects in the world whether they are known or not; knowledge is fallible because any epistemic claim may be refuted; there are trans-phenomenalist truths which refer to the empirical world and discount deeper levels of social reality, i.e. the work of social mechanisms; and more importantly, there are counter-phenomenalist truths in which those deep structures may actually be in conflict with their appearances. If we conflate the two this leads to confusion and misappropriation.

The second claim is that the social world is systemically open. Closed systems are characterised by two conditions: objects operate in consistent ways, and they do not change their essential nature. Neither of these conditions pertains to open systems. In closed systems measured regularities are synonymous with causal mechanisms. Experimentation is therefore unnecessary because experimental conditions are naturally present. There are two alternatives: artificial closure and the use of methods and strategies that fit with systemic openness, including, but not exclusively, inferential judgements from the analysis of indirect evidence. The first of these alternatives, artificial closure, makes a number of unsubstantiated assumptions: cross-environmental transferences can be made even if the original knowledge is constructed in artificial conditions; and this original knowledge is correctly related to the constitution of the object, i.e. an assessment result is isomorphic with the capacity of the individual, whether this is expressed as a knowledge-set, skill or disposition. We are therefore left with those methods and strategies that conform to the principle of systemic openness.

The third claim is that social reality has ontological depth. Social objects are the real manifestations of the idealised types used in discourse and are the focus for any enquiry. They are structured in various ways, and because of this, they possess powers. The powers that these structures (or mechanisms) exert can be one of three types (cf. Brown et al., 2002). Powers can be possessed, exercised or actualised. Powers *possessed* are powers that objects have whether they are triggered by the circumstances or not. Their effect may not be evident in any observable phenomena. Powers *exercised* have been triggered and are having an effect in an open system, and as a result they are interacting with other powers of other mechanisms within their sphere of influence. These exercised powers may still not give rise to any observable phenomena as these other powers may be acting against them. Powers that have been *actualised* are generating their effects; within the open system they are working together with other powers, but in this case they have not been suppressed or counteracted. Embodied, institutional or discursive structures can be possessed and not exercised or actualised, possessed and exercised, or possessed and actualised. As a result, a causal model based on constant conjunctions is rejected and replaced by a generative-productive one, and objects and relations between objects (as in educational systems or testing regimes) have emergent properties.

Three propositions follow from this critical realist perspective. The first is that any descriptions we make of human agency and its capacities are dependent upon ‘intentional causality or the causality of reason’ (Bhaskar et al., 2010, p.14). Second, these descriptions need to take account of ‘synchronic emergent powers materialism’ (ibid.), that is, time-sequenced changes to the powers of objects, whether discursive or embodied; and thirdly, there is a need to acknowledge ‘the *evaluative and critical* implication(s) of factual discourse’ (ibid., my italics). These three principles have significant implications for developing a comprehensive explanation of cross-national and cross-cultural testing regimes, such as PISA.

FALSE BELIEFS

The default position taken by those working within the psychometric tradition of knowing other minds is that a person has a number of capacities (i.e. knowledge sets, skills, dispositions), which we can describe as the contents of that person’s mind, and which subsequently we can characterise using the methods of experimentation and testing. There is therefore potentially a true score for a person, and this true score represents in symbolic terms her capacity in the particular domain being tested. For a variety of reasons, errors may occur in the process of constructing that true score, but these are corrigible, i.e. they can be corrected by using different (and thus by implication better) methods and approaches. Errors may occur because the wrong type of instrument is chosen for determining the person’s true score or because her emotional and affective states are such that she gives a false impression of her capacities. In contrast, I want to suggest that there are a number of false assumptions being made here, perhaps best expressed as false beliefs.

The first of these is that a person has a knowledge, skill or dispositional set, which is configured in a particular way (i.e. it has a grammar), and it is this knowledge, skill or dispositional set, or at least elements of it, which is *directly* assessed when that person is tested. In contrast, any testing that is carried out with the purpose of determining whether these attributes are held, not held, or even partially held by an individual, always involves an *indirect* process of examination, where the additional element is a conjecture, logical inference or best guess. Furthermore, the required performance elicited during the test is specifically related to the testing technology, so, for example, if a multiple-choice test is chosen, the correct answer and therefore the correct construction of the problem are framed to fit this technology. In order to obtain a true measure of that person’s capacity (i.e. K_a), and not, it should be noted, a comparative measure of the construct being tested at the individual or group level (i.e. K_b), then a retroductive mode of inference would need to be used to identify what must have been the case in order to bring about the observed event (i.e. the testee answering a multiple-choice question in a standardised test).

A second false belief is that this grammar is organised into elements, there are relations between those elements, and each element can be scaled, which can then be directly investigated. This can be contrasted with a position which suggests that, in the application of the knowledge, skill or dispositional set, whether for the

D. SCOTT

purposes of testing or for use in everyday life, a range of other knowledge elements, skills and dispositions are invoked. This should not be conflated with the idea that the contents of the curriculum cannot be disconnected for the purposes of testing, leading to a belief in property holism (cf. Curren 2006, for a refutation). What, in contradistinction, is being argued for here is that in the application of a knowledge set, skill or disposition, whether for the purposes of testing or otherwise, a range of other types of knowledge and skill are needed, and the testee may not have sufficient knowledge of these matters or be sufficiently skilful in relation to them. For example, the application of higher-level mathematical skills, i.e. solving algebraic equations, assumes a knowledge of and a capacity in lower level mathematical skills, such as addition and subtraction.

There is, on the other hand, a set of factors which may result in construct-irrelevance variance (Messick, 1989), that is, variance amongst a population of testees as a result of factors which do not have anything to do with the construct being tested. Even if knowledge of or competence in the construct is equally distributed in this population, some testees will do better than others (i.e. on their actual scores) and this is not because they have greater knowledge or are more competent in the construct being tested. This might involve either construct-under-representation or construct-over-representation (William, 2010), and within the confines of the test itself it is impossible to determine which of these has occurred. The challenge for testers then is to eliminate such construct-irrelevance variance. However, this is not without its problems. First, we cannot say with any degree of certainty what the variance might be because we don't know what a true score for the individual or an aggregated true score for a group is, and therefore have nothing to compare it with. Analytical comparisons can be made, and in PISA are made, i) over time (between T_1 and T_2 , where T represents a time-point), ii) between different capacities (if an individual is expert at C_a , then she will also be expert at C_b , where C represents a capacity), iii) between different constructs (Co_1 has the same level of difficulty as Co_2 , where Co refers to a construct), iv) between different performative settings (S_1 is considered to be isomorphic with S_2 , where S refers to a setting), v) on the same test at two different time points (this is an external measure of reliability, R_a), vi) with different items on the same test at one point in time (this is an internal measure of reliability, R_b) and vii) on comparable tests at two different time-points (this is another external measure of reliability, R_c). With i) an assumption is made that no emergent properties of the construct being tested are activated, and moreover, that no learning takes place, as a result of the testing or otherwise, between T_1 and T_2 . With regards to ii) an assumption is made that expertise in specific capacities automatically transfers to expertise generally. With the third analytical comparison (iii), an assumption is made that all measureable constructs have an equal level of difficulty in their acquisition and in their application. The fourth of our measures (iv) seeks to confirm the validity of a score on a test by examining whether that aptitude can be applied to other spatio-temporal settings outside of the test setting. An assumption is made that the construct being tested has transferable characteristics and is not specifically connected to a particular performative setting. Finally, with v), vi) and vii), an assumption is made that if a score on a test is reliable then it is also valid. Each of

these analytical comparative forms is underpinned by assumptions or beliefs that in turn need verification, or at least can allow trust in their use. And thus a further rationale needs to be provided for each of these assumptions.

A second problem with eliminating construct-irrelevance variance is that it cannot be achieved by replacing a competency with a knowledge construct, despite this being the clear intention of PISA test constructors. For example, PISA 2006 attempted to assess three broad science competencies: 'i) Identifying Scientific Issues; ii) Explaining Phenomena Scientifically; and iii) Using Scientific Evidence' (OECD 2006, p. 12). This is because the problems associated with construct-irrelevance variance apply equally to knowledge *and* competence constructs, and in addition, with regards to the assessment of competence constructs, there is the problem of multiple interpretations being made. Traditionally, this is described as a problem relating to tester-reliability.

Test-constructors confronted by the problem of construct-irrelevance variance may seek to reformulate the construct, so that those matters which might be considered to be separate from the construct, such as the time element for solving a problem in a test, now become part of the construct, i.e. the assessment now relates to the capacity to solve the problem within a definite time period and not just to the capacity to solve the problem. This introduces a performative element into the construct itself. Once again, this move is beset with problems, since it weakens the idea that individual expertise in that construct can be transposed to other settings because it is now more context-dependent as an assessment. What has been weakened is the predictive validity of the assessment. In cross-national testing environments such as PISA some of those performative elements can be standardised, i.e. the tests are conducted in roughly similar conditions. However, what cannot be standardised is the relation between what is taught and what is being assessed, how this assessed knowledge relates to its usage in other environments, and the test-taking capacity of the individual or group.

A third false belief is that in the use of a knowledge-set, or in the performance of a skill, or in the application of a disposition, no internal transformation takes place. (In fact, both internal and external transformations are neglected within traditional psychometric accounts.) In contrast, within a person's mind two knowledge sets are being activated. The first is the original knowledge set (K_a); and the second is the transformed set (K_b). Further to this, K_b is not just the result of a causal mechanism at work but may also at different points in time influence and transform K_a ; that is, it has the capacity to bend back on itself and act recursively to change its original form.

There is also an external transformative process at work, and thus a fourth false belief is that testing a person's knowledge, skills and aptitudes has no washback effects on either K_a , the original knowledge construct, or K_b , the internally transformed knowledge set ready for testing. In contrast, the well-documented process of washback works in just this way (cf. Stobart, 2009), so that instead of the assessment acting merely as a descriptive device, it also acts in a variety of ways to transform the construct it is seeking to measure. Washback effects work on a range of objects and in different ways. So, for example, there are washback effects on the

D. SCOTT

curriculum, on teaching and learning, on the capacity of the individual and more fundamentally on the structures of knowledge, although these four mechanisms are frequently conflated in the minds of educational stakeholders.

Micro-washback effects work directly on the person, whereas macro-washback effects work directly on institutions and systems, which then subsequently have an impact on individuals within those institutions and systems. For example, at a global level, policy enactments may lead to changes in national curricula and national systems of testing, which in time will lead to changes in curriculum and assessment at the level of schools and thence to changes in what is learnt and what an individual considers to be performative knowledge. What is considered to be appropriate performative knowledge has therefore changed as a result of changes at global, national and school levels. Washback effects do not work in a deterministic way, since there are a large number of activities that have to be coordinated during the sequence of events to achieve the desired result, and mechanisms such as these have emergent properties because they operate in open systems (cf. Bhaskar, 1989).

The argument is therefore made by cognitive psychologists and test constructors that no internal or external processes of transformation occur when the knowledge, skills, or dispositions of the person are tested; i.e. that person knows A or has skill B or disposition C, and that in the act of displaying that knowledge or using that skill or allowing that disposition to be realised, no change occurs to the original knowledge construct, or skill set or disposition, in order for that person to respond in the appropriate manner to the situation confronting her. In contrast, I want to suggest that there is a transformative process and it may take a number of forms, i.e. accretion and thus retention of the original knowledge domain, skill or disposition; or subsumption, where the original knowledge domain is subsumed into a new domain and thus loses its identity; or subtraction so that parts are discarded to accommodate the contingencies of the new setting.

What this also points to is that in the process of determining whether a person knows this, or can do this, or has the necessary disposition, an inferential process is required so that the observer can move from evidence, i.e. the test result, to a description of an actual state of being. The assumption is made that if this person can do X in the test situation, then they can also do it in different situations, or if that person knows something in the test situation, then they also know it in other situations. It is, in short, the problem of transfer (from T_1 to T_2 or from C_1 to C_2 , where T refers to a moment in time and C refers to a context of application), and it is problematic because it is prospective and morphogenetic. A measure of predictive success to determine whether a person or group of people can do X in other settings outside the testing environment can be developed; however it is an unreliable measure for two reasons. Events, happenings and unplanned occurrences during the interval between the two time points (T_1 – the test setting and T_2 – the application setting) cannot be controlled for; and the two different activities are not comparable.

A fifth false belief is that the process of testing works in a unidirectional linear fashion. For example, a person knows X, that person is subjected to an examination which is designed to test for traces of X in a population of knowers with similar characteristics, and a score in relation to that construct is recorded indicating that

the person either knows it, doesn't know it or knows it to some extent. No consideration is given to bidirectionality, incorporating forward and backward flows, so that the taking of the test and the recording of the mark impact on and influence the original knowledge construct. This changes the structure (both quantitatively and qualitatively) of the construct, and its affordances, making the original determination of it and them unreliable.

A sixth false belief is that different types of knowledge, including those at different levels of abstraction, can be tested using the same algorithmic process. For example, testing a knowledge of facts and testing a capacity to synthesise basic facts are different processes. And this is because in the former case the test item refers directly to the construct being tested, whereas in the latter case it refers to an example of the construct, and successful mastery of the construct has to be inferred from successful mastery of the example. This latter process therefore additionally has to satisfy criteria such as relevance, quality and probative force for that inferential relationship between example and construct to be considered valid.

A seventh false belief is that the performance on the test represents to a greater or lesser extent (given that the person may have been distracted or constrained in some way or another) what the testee can do or show, rather than there being a qualitative difference between the performance on the test and the construct, skill, or disposition of the testee. An individual may have to reframe their knowledge set to fit the test, and therefore the assessment of their mastery of the construct is not a determination of their capacity in relation to the original construct, but a determination of whether the testee has successfully understood how to rework their capacity to fit the demands of the testing technology.

An eighth false belief is that a test can be constructed which is culture-free or free of those issues which disadvantage some types of learners at the expense of others. This mechanism works in a number of ways: test constructors may use background material which is unfamiliar to some testees but familiar to others; test items may have been taught in different ways to different groups of testees, that is, they have been given different values, or taught in a different order, or even not taught at all; and the testing technology may be unfamiliar to them because of factors which are peripheral to the articulation or use of the particular construct, but central to the testing technology used to assess it.

EXAMINATION TECHNOLOGIES

If no incentive is attached to the taking of a test, i.e. personal benefit such as gaining entry to a higher education institution, or monetary reward, or furtherance of a student's learning trajectory, or national advantage, then the student is not likely to treat it very seriously. The value that she attaches to it is always a matter of perception, rather than designation, and this means that different types of students will be motivated to do well to different degrees. Cognitive psychologists and test constructors argue that these individual characteristics of test takers are accounted for at the level of the group, and the argument is then made that these characteristics, i.e. propensity to lose concentration in a test or not give a true

D. SCOTT

account of their capacities because the examination technology offers them no incentive to do well, or having a presentational style which is at variance with the affordances of the examination technology, are randomly distributed amongst members of any group, and therefore do not effect scores at the group level. As a result, groups can be reliably compared with each other. However, the assumption that these characteristics of group members are evenly distributed is false (cf. Mac Ruairc, this volume), and in addition, this is a measure of reliability rather than construct validity. Furthermore, these characteristics may be the defining characteristics of the group.

As an example, let us take a multiple choice test. The technology only allows a limited range of answers; therefore there is a high probability of false negative and false positive errors (Wood & Power, 1987), despite misleaders being inserted as questions to allow reliability checks to be performed. Only a limited range of knowledge items and processes can potentially be tested because correct answers are being asked for, and those answers are framed in ways that do not allow discursive, equivocal responses. As a result, this technology has the effect of widening the gap between the capacity of the individual and her performance (both internally and externally), because the test is constructed so that it has few of the characteristics of the original knowledge construct and potentially its application. There is in short a limited discretion given to the person being tested and therefore in principle at least, multiple-choice testing has a greater propensity to washback onto the curriculum. Furthermore, the characteristics of the technology used for multiple-choice testing favour some groups in comparison with others, i.e. boys may have an advantage over girls.

A contrasting example is the use of a free-ranging essay format to determine the comparative capacity of a group. A wide discretion is given to each candidate, though marker unreliability effects may be high. The assessment is not focused on discrete facts but on general competencies, i.e. the ability to sustain an argument. Thus in principle it may be better able to measure higher level skills. Validity may be strong if this is understood as an alignment between the knowledge, skills and dispositions of the person and the description that is made of them. Because marker discretion is high and because the candidate is allowed more latitude in how she frames her answers, then the possibility of a significant washback effect is reduced.

A test is always a performance. The taker of the test frames their response to the test in terms of what they perceive to be the correct answer. This operates at the unconscious level, and it is unremarkable. When we have a conversation with another person, we frame our responses and our mode of responding to how we think our messages are going to be received. With regards to testing, there is a further element, which is that the testee frames their answers in terms of their perception of what they consider to be the correct response. If for example, there is some ambiguity in the question, the testee asks herself the question: what type of answer should I give which is likely to result in the award of the maximum amount of marks? Test constructors aim to write questions or construct problems to be answered with as little ambiguity as possible. This is achieved (though rarely successfully) by reducing the scope of either the question/problem to be solved or by reducing the response that

the testee is required to make, and this involves a reformulation of the knowledge construct, though it may still contain residues of its original form.

FAIR TESTS

PISA test constructors have chosen to measure competencies rather than knowledge sets on the grounds that the latter are specific to particular countries, whereas competencies have universal characteristics. There are two problems with this. First, those national and local features of knowledge domains apply in equal measure to skills, competencies (skills expressed as individual capacities) and dispositions (configurations of individual capacities which can be expressed as affordances). Second, there is a longer and more complex inferential chain involved in the measurement of competencies than there is in the measurement of knowledge acquisition, and there is therefore a greater likelihood of construct-irrelevance variance occurring.

PISA has attempted the difficult task of constructing curriculum-free tests; the most notorious example being the 11+ examination in the UK (cf. Torrance, 1981, for a critical evaluation). The reason for this is that making comparisons between the test performances of students from different countries, with different curricula and with different teaching methods and approaches, requires the selection of test items that do not reflect national curricula or national pedagogic methods. So these international comparative tests, and this includes items which refer to socio-economic conditions of the student and attitudinal data (as in the latest PISA Science-focused set of tests), are not a measure of their curriculum, nor what they have been taught, nor are they a measure of what they have learnt in any formal sense. This means that the content of the test items and the presentation of those test items are likely to favour some countries at the expense of others.

Cultural differences may take a number of different forms, such as, ascribing different values, and different strengths of values, to cultural items, or determining the nature, quality, probative force, relevance-value and extent of evidence, or focusing on practices which may be more familiar to people in some countries and less so in others. However, more importantly, cultural differences with regards to the selection of test items refer to the expression of the problem to be solved. If, for example, different national idioms, different national ways of thinking embedded in language forms, and different normic values woven into the fabric of national discourses are ignored, then the presentation of the actual test items as well as the range of possible answers that can be given may favour students from one nation at the expense of students from another.

This is the problem of fair comparison. And in order to make a fair comparison, it may not just be a question of translating the words which are being used, that is, substituting one set (words, sentences, and language structures) for another, but transposing the example and the problem, so that it better reflects its new epistemic base. Underpinning the notion of an international test is the idea of a universal, i.e. culture-free, form of knowledge, which can be adapted so that superficial differences between nations are eliminated. However, it is never enough to say that a test simply

D. SCOTT

tests the capacities and knowledge constructs of a group (in this case a trans-national group) of students. What a trans-national test does is make a number of reductionist assumptions about the knowledge bases being tested which result in imperfect caricatures of all the national knowledge bases under consideration.

COMPARATIVE EMERGENT PROPERTIES

PISA results are expressed as comparative national tables rather than scores achieved by participants. The focus is on position rather than score, even though significant improvements made by one nation between two time points may be masked by improvements made by other nations. If you add to this the idea that there is some uncertainty or unreliability about the scores (i.e. marker error, poor performance by testees, cultural bias effects, epistemic differences, inability to transform internal knowledge into performative knowledge, etc.), it is hard to believe that such league tables can and do provide a nation with very much useful information. However, what we have here is a display mechanism (located initially at the transitive level, but also penetrating and thus taking on a capacity to operate at the intransitive level). This display mechanism clearly has scientific aspirations (cf. Habermas, 1971), adding further to the need to introduce critical and evaluative elements into any accounts made, whether they refer to individuals, groups within nations, or nations themselves.

Michel Foucault (1979, p. 191) suggested that the examination transformed the individual into an object for a branch of knowledge:

The case is no longer, as in casuistry or jurisprudence, a set of circumstances, defining an act and capable of modifying the application of a rule; it is the individual as he (sic.) may be described, judged, measured, compared with others, in his very individuality; and it is also the individual who has to be trained or corrected, classified, normalized, excluded, etc.

For the first time an individual could be scientifically and objectively categorized and characterized through a modality of power where difference becomes the most relevant factor.

Furthermore, the instrument (PISA) is a performative device, in so far as its intention is not just to describe the skills/dispositions of children but to promote and thus contribute to national policy-making. Certain forms of performative knowledge become the norm. The instrument for measuring knowledge and skill levels of children becomes an instrument for determining what those knowledge levels and skills should be, and how they should be learnt. The mechanism that underpins this series of actions is an example of synchronic emergent powers materialism (cf. Bhaskar, 2010), and as a result, it operates as a standardising device in relation to these matters (i.e. it creates a norm) and should not be understood as a device for making fair, reasonable and accurate judgements about the capacities of cohorts of students in different countries. There is a final point to be made, and this is that a nation's place in these league tables becomes part of the folkloric account a nation gives of and to itself. Since this account is an important

part of a nation's identity, then success in an international test such as PISA becomes even more important.

REFERENCES

- Bhaskar, R. (1989). *Reclaiming reality*. London: Verso.
- Bhaskar, R., Frank, C., Hoyer, K.-G., Naess, P. & Parker, J. (2010). *Interdisciplinarity and climate change*. London: Routledge.
- Brown, A., Fleetwood, S. & Roberts, J. (2002). *Critical realism and Marxism*. London and New York: Routledge.
- Curren, R. (2006). Connected learning and the foundations of psychometrics: A rejoinder. *Journal of Philosophy of Education*, 40(1), 17-29.
- Foucault, M. (1979). *Discipline and punish: the birth of the prison*. New York: Vintage.
- Stobart, G. (2008). *Testing times: the uses and abuses of assessment*. London: Routledge.
- Habermas, J. (1971). *Knowledge and human interests*. Boston: Beacon Press.
- Messick, S. (1989). Validity. In R. Linn (Ed.) *Educational measurement* (3rd Edition). American Council on Education, Washington, D.C.
- Organisation for Economic Cooperation and Development (OECD) (2000). *Manual for the PISA 2000 Database*. Paris: OECD.
- Organisation for Economic Cooperation and Development (OECD) (2001). *Knowledge and Skills for Life: First Results from PISA*. Paris: OECD
- Organisation for Economic Cooperation and Development (OECD) (2006). *Knowledge and Skills for Life: PISA*. Paris: OECD
- Organisation for Economic Cooperation and Development (OECD) (2009). *Knowledge and Skills for Life: PISA*. Paris: OECD.
- Torrance, H. (1981). The origins and development of mental testing in England and the United States. *British Journal of the Sociology of Education*, 2(1), 45-59
- William, D. (2007). Balancing dilemmas: traditional theories and new applications. In A. Haynes, & L. McDowell (Eds.), *Balancing Dilemmas in assessment and learning in contemporary education*. London: Taylor and Francis.
- Wood, R. & Power, C. (1987). Aspects of the competence-performance distinction: Educational, psychological and measurement issues. *Journal of Curriculum Studies*, 19(5), 409-24.

*David Scott
Institute of Education
University of London*

DONATELLA PALOMBA AND ANSELMO R. PAOLONE

COMPETENCIES VS. INTERCULTURALTY. STUDENT EXCHANGES IN THE AGE OF PISA¹

THE FRAMEWORK

The research on which this paper is based was not originally intended to examine directly the PISA *problématique* – rather, in studying a different topic, it has, so to speak, met PISA on its way, and in the ensuing interaction of the themes it opened up a perspective which seemed worth a closer analysis.

The focus of research, carried out by SICESE (the Italian Society for Comparative Education) on behalf of the *Fondazione Intercultura*, was the complex question of teachers' role and attitudes toward students' exchanges at secondary school level, with special reference to long term individual exchanges – that is, those exchanges in which individual students spend a whole school year (usually the last but one year of secondary school) attending school in a different country.

These exchanges are to be considered as an important tool for intercultural education (*formazione interculturale*). Even if much of the current reflection on intercultural education is focused on issues linked to minority and immigrant groups, other different aspects of intercultural relations and educational paths connected with them should not be overlooked. Namely, the specific aspect of intercultural relations represented by youth exchanges, especially long term individual ones, has a peculiar interest. The student, moving for one whole school year to a different system, experiences personally the condition of “otherness”; moreover, in a situation where the most obvious socio-economic inequalities are usually absent, the problems and dynamics which come to the forefront are more clearly the ones specifically linked to the differences between cultures as such.

These are the features which make unique the exchanges – as a personal experience and as a field of study as well. It is therefore a theme of peculiar interest in intercultural research; and it may be added that it is also a theme of great interest for a comparative research in which both institutional and cultural aspects are considered.

However, in many cases the question remains open of the *status* and the formal recognition in one's own educational system of the knowledge and the skills acquired attending a school in a different country. This is a crucial point, which varies significantly in the different countries, according to the different rules they adopt on the matter; but recognition is also strongly connected, in the classroom practice, with the attitudes adopted by the teachers.

In the framework of an overall engagement for the “internationalisation” of the school system, in Italy in the last years there has been an effort to give a new impulse

to these experiences. Recent research has however confirmed that the teaching body tend to be suspicious about the exchanges, especially long term ones; and this is due mainly (though not only) to reasons related with the curriculum, which in Italy is traditionally based on disciplinary knowledge.

In our analysis of teacher's attitudes, one of the hypotheses we formulated was that the diffusion of PISA surveys, officially supported by the Ministry, could contribute to make these attitudes more positive, even in schools that do not participate directly in the surveys. As far as there is a spreading of a cross-national agreement on the definition of learning achievements, of knowledge and skills to be acquired, of competencies to be mastered, and most of all of evaluation tools, with their "wash-back" effect – this could result in a significant facilitation of the recognition, by any individual school or system, of learning acquired in other schools and countries. From this point of view, one possible hypothesis is that the impact of PISA could have an indirect influence in promoting student mobility and exchanges.

Some of the indications coming from the first findings of the research seem to support this hypothesis – but at the same time they highlight the core ambiguity of the situation arising from such a dynamic. Actually, at exactly the same moment when there is promotion and a support of students' mobility – traditionally intended as a vehicle for intercultural experiences and formation - the conditions for the support of the mobility itself are defined according to the general criteria and approaches established by organisations operating at trans-national level (namely, in the case of PISA, with an emphasis on competencies as the desired learning outcomes, and therefore as the object of the assessment); approaches which do not necessarily respond to the different cultural traditions.

The case of the Italian school system is significant in this respect. Traditionally quite far from the notion of "competency", it is now, under the pressure of international discourse about school efficiency and international competition, trying to move, even if slowly and with many resistances, towards the acceptance of the new, competency-based, international "discipline". Not surprisingly, the "discipline" is adopted with more conviction by the most successful PISA schools, deemed the most "advanced" of the whole system, which also, usually, show a positive attitude towards the exchanges.

In the "Competencies vs. Interculturality" encounter, therefore, the definition of transnational criteria for defining school achievement, which tends to promote the recognition of learning acquired during the exchanges, and consequently support mobility, might have a negative impact on that same cultural diversity that should represent the core meaning of the mobility experience, bringing even classroom practice under a form of transnational governance which bears heavily on the approach to teaching and to the definition of its goals and ends.

THE UNEXPECTED IMPACT OF PISA ON STUDENT MOBILITY IN SOME ITALIAN SECONDARY SCHOOLS, IN THE AGE OF AUTONOMY

This section tries to show how the impact of PISA unexpectedly extends to fields such as that of student mobility producing, inside an evolving school system such

as the Italian, some interesting effects. The analysis is the result of field research, trying to understand the attitude of Italian secondary school teachers about individual long term exchanges. This attitude has been studied by observing how teachers act towards returnees, the assessment of competencies which the latter have acquired abroad, being a central moment of this acting². Such assessment is problematic because on the one hand, the Italian system officially recognizes these periods in the foreign countries, on the other hand it does not give clear rules on how to assess the knowledge and the competencies acquired abroad. Therefore, teachers are left in a dilemma: they are asked to assess, but they are not told how to do it. The two *case studies* in the final part of the paper, show one of the possible solutions found on the field: in the void left by inadequate legislation on this topic, some teachers have drawn inspiration from the evaluation of competencies in the PISA tests.

As it was already mentioned, the moment of assessment of what returnees have learned abroad is crucial for understanding the teachers' attitude about long term individual exchanges. To the eyes of the ethnographer, assessment is an activity in which the attitude of teachers is revealed. Firstly, because by appraising (or by trying to appraise -when this is difficult because of the inadequacy of the existing rules), teachers show whether they are benevolent or malevolent about exchanges, able or unable to consistently appraise them, creative or stiff in overcoming difficulties, willing to consider positively the whole of the human experience which the returnee did abroad or sticking to a vision of the curriculum mainly based on disciplinary learning, etc.

Secondly because, considered the necessity to assess and considered the lack of rules on assessment we have already mentioned, within the new regime of autonomy of the Italian school (as we shall see more in detail later), individual schools (i.e. the teachers, which are the actual agents within each school) are free to draw from other sources to understand, *inter alia*, how to deal with exchanges and returnees. Such "alternative" sources, the research has shown, are in good measure "international" (for instance, other exchange programmes such as Leonardo, Comenius, or individual international exchange agreements between schools etc.)

These "international" sources broaden the mentality of teachers and make them more articulate in understanding the problem herein discussed. Among these sources we can find the PISA, which provides teachers with an assessment model which can change their attitude towards student exchanges, not only by helping them to assess competencies, but also by contributing to build, in their mentality, a more tolerant, constructive and rich perspective on what is being learned abroad, by showing them that these contents can be at least in part assessed thanks to a system which is in fact a raising international standard. Teachers then feel that a prestigious institution could be of some help where the national rules are incomplete and unclear. The international dimension of such a standard makes teachers also consider that today students are no longer confined within isolated national school systems, but they compete in a global arena of knowledge and competencies, and in this scenery long term individual exchanges become a crucial experience for them.

Let us clarify how this positive influence changes teachers' mentality, by showing how, according to the research, it eases up a number of negative attitudes that we shall enumerate herein.

Firstly, the research has shown that in terms of contents and curricula of the foreign schools visited by returnees, the Italian school staff fears a sort of "dispossessing of teacher's professional role."

Italian teachers mostly have the tendency to perceive as desirable the exchanges in which students attend a foreign school whose syllabus and curriculum are similar or compatible with those of the school of origin.

Consequently, they perceive being robbed of their professionalism when the external organizations in charge of student exchanges decide in which school students should be sent (which is the general rule) disregarding curricular homogeneity and the advice of the school staff. This happens because the pedagogical tradition of such organizations aims at creating a cultural rupture in pupils' experience. By overcoming the acculturation shock, consequent to the attendance of a different school in a different country, the exchange student is expected to learn how to cope with a different culture at many levels: in the classroom but also in most aspects of the everyday life. The condition of *liminality* (Turner, 1974, 1982; Paolone, 2006) that he experiences is believed to be the spring of a multi lateral personal growth.

This pedagogical vision seem to clash with some of the traditional aims of the Italian school system, which was born under the auspices of national unification (Palomba, 2009; Morandini, 2001; Fornaca, 1994; Canestri, 1983; Talamo, 1960) and until recently has been a centralised one. At the local level, observable by the ethnographer, the negative reactions of the studied teachers (Paolone, 2010a, p. 21) consequent to this clash, seem to be at least twofold. On the one hand there is a preoccupation for the time the returnee has "wasted" in terms of what he has missed by not studying the Italian curriculum (e.g.: in the perspective of the final State exams). This preoccupation is particularly strong in staff members teaching subjects which are little taught in foreign schools (ancient Greek, Latin, philosophy, etc). Many of these staff members believe that the choice of a foreign school with a curriculum similar to that of the school of origin could ease such problems (Paolone, 2010a, pp. 63-64, p. 73).

On the other hand some staff members seem to be worried by the transformed mentality of the returnees. In fact, some returnees have actual problems of reinsertion not only in terms of school performance, but also in terms of peer group relationships. All this contributes to the scepticism of some staff members including, in some cases, the principals, who generally declare to be favourable to exchanges (Paolone, 2010a, p. 55).

But when these beliefs and scepticisms have been studied more in depth, it has emerged that actually behind them, more than facts, there is a form of ideological uneasiness: teachers are puzzled, *inter alia*, by the overall growth of the returnees because they cannot frame it into the (inadequate) formal schemes of the Italian school system, especially in terms of their assessment. Many aspects of this growth, at least those which are meaningful for the school, are in fact competencies

and they thus differ from what the Italian school is traditionally better equipped to assess: knowledge.

This is where PISA comes in and helps. PISA supplies a model of competency assessment which helps teachers solve part of the puzzle and which broadens their vision of what schooling is today, in terms of its growing international dimension.

This induced “internationally aware” mentality also alleviates the demand for curricular continuity and homogeneity. Under the influence of PISA teachers are induced to understand that the Italian curricular model, traditionally based on disciplinary knowledge and only recently influenced by an understanding, however limited, of competencies (the so called “disciplinary competencies”) is not the only possible one, and that what is being learned abroad, on the contrary, is often more in line with the competency minded spirit of international schooling, incarnated to their eyes by the PISA.

THE INADEQUATE ITALIAN LEGISLATION ON STUDENT EXCHANGES AND ON THE ASSESSMENT OF THEIR EXPERIENCE ABROAD

Let us now outline the framework of the Italian legislation on student exchanges and on the assessment of their experience abroad, in order to show, by the description of two case study, how the PISA influence fits in.

In Italy, study periods in foreign countries are organically disciplined by the Ministerial Circular 17/3/1997 no. 181, on “international student mobility”, concerning pupils of the Italian secondary schools.

In short, the Circular says that: *a)* individual periods of study can be done following agreements between Italian and foreign schools, or on the basis of the individual pupils’ initiatives; *b)* individual periods of study in the foreign countries by Italian secondary school pupils are recognized in terms of the pupils’ readmission in the schools of origin, and are assessed on the basis of their coherence with the didactic objectives provided by the Italian syllabi; *c)* to such purpose, for a preliminary appraisal of the foreign syllabus, the Italian competent teachers’ board (in the school where the returnee is to be readmitted) directly acquires from the foreign school that the pupil will attend, information on the syllabus that the pupil will study and on the system of assessment used in the foreign school; *d)* the period of study in foreign countries can not last more than one academic year and in any case has to finish before the beginning of the following school year; *e)* at the end of the period of study abroad, and before the beginning of the new school year, the Italian competent teachers’ board, having seen how the pupil has been assessed in the foreign school and having seen the results of an optional integrative test, deliberates on the readmission of the pupil.

It is interesting to notice that no direction is given on how the tests should be made, and that a basic problem seems to be ignored by this Circular: the assessment done in foreign schools is often very different from that done in the Italian schools, which usually test and assess disciplinary knowledge, and have not received clear instructions on how to adequately assess competencies, which are ubiquitous and fundamental in other school systems (the Italian law, apparently

unaware of the complexities of the task, in another document says that competencies should be assessed with just numeric marks) (Law 30/10/2008, no.169, art.3).

Before completing the description of this legislative framework, it is important to remember that since 1999 the Italian school system has been undergoing a revolution in terms of local autonomy. For over a century the State school system has been a centralised one, but following a general “local autonomy” trend in Italian society, provided for in the Constitution of 1948 and constantly progressing since the end of WW II, individual schools have recently been awarded the freedom of making decisions on a wide range of matters.

Subsequently, the topic of student exchanges has been tied to the *school autonomy* regulations, that attribute to individual schools the competency to handle the school career of the pupils and to discipline, through the legislation in force, “the registrations, the attendance, the certifications, the documentation, the assessment, the recognition of the studies done in Italy and in the foreign countries in order to allow the pupils’ progress in their career of studies, the assessment of the formative credits and formative debts, the participation to local and international projects, the making of international student exchanges” (DPR 275/1999, art. 14, c.2).

Among the consequences of this, there is a very important trend: individual schools react idiosyncratically to the legislation void (and to the consequent contradiction in which teachers are caught) that was mentioned at the beginning of this paper, and therefore the only possible approach for surveying the new solutions which schools are finding consists in doing ethnographic fieldwork in individual schools.

The Italian law neglects instructions on how to assess the competencies that students have acquired abroad, but then says that this assessment should be translated into the new Italian assessment system based on the complex and controversial concepts of “formative credits” and “oscillation band”. The Ministerial Circular n. 236 of October 8, 1999 offers specific explanations in terms of State exams. Considered that, in general, in the foreign schools attended by the Italian pupils the syllabi and criteria of assessment are different from those in Italy, the M.C. underlines that “for reasons of equity and parity of treatment, it is necessary to adjust the aforementioned matter to the forthcoming school credit, which has been introduced by the new law on State exams.”

To do so, the following instructions are given:

- a) The teachers’ board checks the returnees on the subjects that should have been studied in the academic year missed in Italy and which have not been studied in the foreign school.
- b) On the basis of the results of the aforesaid tests, the teachers’ board formulates a global assessment, which also considers the assessment expressed by the foreign school on the subjects which are common to the two systems³. Such assessment determines the insertion of the pupils in one of the “oscillation bands” of the school credit, provided by the law in force.

- c) The pupil that in the preceding academic year (the one not attended in Italy) has a formative debt will be awarded the lowest mark in the “oscillation band”. In case he/she overcomes the formative debt, in the year when the pupil is readmitted in the Italian school, the teachers’ board can integrate, in the final poll, the lowest score, within the limits of the “oscillation band” to which the assigned score belongs.

According to the dispositions of the Ministerial Circular 236/1999, and in the light of the legislation on State exams introduced by the law n. 425/1997, the assessment mentioned in point B is done by the teachers’ boards keeping in mind the following three elements:

- 1) Knowledge: the learning of other disciplines, uses, customs and socio-cultural themes.
- 2) Competencies (to know how to do): the acquisition of the language of the foreign country and/or the expansion of the language already studied.
- 3) Ability (to know how to be): the growth of the personality and of the cultural sensibility.

But again, no hints are given on how, in practice, to perform such competency assessment, and in other laws it is required to rank the results with plain numerical marks (Law 30/10/2008, no.169, art.3) which seems to exclude any articulate form of competency assessment.

Then, the problem of assessing the subjects taught abroad, which are not present in the Italian syllabus, is not organically disciplined by the Italian law. On the other hand, the law suggests that extracurricular acquisitions, such as the growth of personality and the cultural sensibility (which have not been named, as other official documents such as the EC declaration on key competencies do, in terms of “intercultural” or “cross cultural” competencies) should be assessed, but it does not define how this should be done in practice and, most of all, the official Italian school report card, do not provide an institutional title under which to inscribe such extracurricular assessment.

THE METHODOLOGY. SOME CONCEPTUAL IMPLICATIONS: A GROUND WHERE VARIOUS DIMENSIONS INTERTWINE

Having considered such premises, a field research has been done in several Italian secondary schools involved in year-long individual student exchange programmes, in order to understand how teachers dealt with this lack of directions. As already mentioned, the circumstance that today’s Italian schools are in a regimen of autonomy, which allows them, *inter alia*, to draw on international sources to build their own individual culture, introduces at least two methodological implications: On the one hand, the research had to do with the relationship between the “local” and the “global” which seems, besides, to be of renewed interest (Arnové, 2003 p. 16 and 2009; Paolone, 2010b, pp. 11-15, 107-116) in contemporary comparative education.

On the other hand, the influence of the global on the local is linked to the idiosyncratic aspects of locality and, to be studied, requires a multi local ethnographic approach.

Therefore the research consisted of a multi local qualitative fieldwork, based on participant observation, semi-structured interviews, recorded “open discussions”⁴, and the study of available documents, such as the schools’ POF (Italian acronym for: *plan of the formative offer*) a document (usually on the web site of the school) where the syllabus is described and the main organisational aspects of the school are displayed.

As already mentioned, at the core of this study there was the practice of assessment of the returnees, by their Italian teachers. It is important to spend a few words about some of the conceptual implications of such core object in terms of comparative research.

In the preliminary research it was found that, conceptually and methodologically, the teachers' attitude (caught by the ethnographer by observing teachers' “creative” assessment practice) about the returnees could be contextualised in a hybrid scene in which various different dimensions intertwined. Subsequently, I have tried, at least in first approximation, to analyse this plexus. I have then found that what the teacher tries to assess in the returnee is the result of a meeting among different cultures, not only in an institutional/scholarly meaning, but also in a more general anthropological meaning (as already mentioned, the M.C. 236/1999 combined with law n. 425/1997, foresee the assessment of “the growth of the personality and of the cultural sensibility” which is the consequence not only of the school experience, but of the overall impact of the foreign culture on the returnee).

On the one hand, when the teacher has to assess the returnee, there is a meeting between two different national school systems: that of origin and that in which the student has spent his/her period of exchange. Here the comparison among educational systems is not the sort of abstract exercise which R. Cowen has criticised, speaking of certain “old” comparative education⁵. Here the comparison among educational systems is concretely embodied in the relationship between teacher and returnee, and the latter brings in himself the fruits of the various experiences he has lived abroad, but must be reinserted in the Italian school system, and thus assessed. The ground of the comparison therefore is the person of the returnee and his/her problematic relationship with his/her school of origin, and therefore also and above all, with the teachers. The Italian school, through the person of the teacher, has to assess what the returnee has brought back from the foreign countries. And there is here an implicit, but urgent and concrete activity of comparison among aspects of the foreign school system involved, and of the Italian system.

On the other hand, there is another *genre* of comparison: in fact the returnee is an Italian student that has lived an acculturation in the foreign countries, as already mentioned, he/she has lived a condition of *liminality* that presumably has made him develop a new personal perspective and learn new things on his life in general, and presumably has allowed him/her to develop a “third code” (Favret-Saada, 1977 and 1981; Paolone, 2006) a series of intercultural competencies that allow him/her to

understand two different cultures: the native and the one he/she has met abroad. The teacher is confronted to this comparison when he/she has to assess the returnee in terms of “the growth of the personality and of the cultural sensibility”.

Therefore, the study of the attitude of teachers should not be limited to the part related to the appreciation and assessment of what has been learned in the foreign schools which, even if it is different from the knowledge acquired in Italy, nevertheless belong to the field of formal education. The attitude of teachers should also be studied regarding how they react towards that personal growth of the returnee (which is one of the pedagogic strongholds of the long term periods of study abroad), growth that the Italian school seems to be unable to assess as it lacks the appropriate tools.

In this sense we assist to a comparison among acquisitions belonging to the sphere of the informal education (the personal growth, which is not only a consequence of the experience done in the foreign school, but also of the effort to integrate the culture of the foreign countries altogether) and the problem of how and whether to assess them in the terms of the formal education (the system of evaluation of the Italian school).

The teacher stands in this meeting ground of different realities, among different dimensions (which at times are conflicting), which makes his/her role towards the returnees extremely complex also on the conceptual plan, in the meaning of a suitable understanding of the pedagogic and anthropological categories which are involved.

This has methodological repercussions. The methodology through which to appreciate the attitude of the teachers toward the returnees should be not only an empirical one, but should be also able to function comparatively, comparatively among different school cultures, comparatively also between the sphere of the formal and the informal education.

Due to its multi-local varieties and its ability to study the intertwining of various dimensions (such as the global influence on the local, the relationship between informal and formal education, etc.) a postmodern ethnography aware of the potential of textual strategies has proved to be a viable approach to the study of this plexus (Paolone 2010a, pp. 35-77, 2010b, pp.107-116 and 2008, pp. 117-152).

The use of such methodology helped to find, in two of the schools studied –the two case studies reported in this paper- that in order to assess the competencies acquired by returnees, teachers draw inspiration, in various (and at times complex) ways, from the PISA tests.

But this happens within a “localist” framework: teachers tend to use concepts and tools they contribute to build in the “culture” (in terms of school autonomy, which we have mentioned above) of the individual school they are working in. The two schools studied here have approached PISA in two completely different ways and have integrated aspects of PISA in their own culture, *translating* and *transforming* (Cowen 2010) these elements, according to their local tradition, previous experiences and actual needs.

The ethnographic approach was chosen also because it allowed to apprehend and represent the individual schools as *cosmos*, as holistic institutions, with their internal structures, rules, idiosyncrasies⁶. In a system such as the Italian, evolving

towards autonomy, each school is developing its own unique symbolic and value system, which is also determined by the territorial environment. In these terms, the studied schools belong to two different scenarios: one is in the suburbs of Rome, in the heart of the peninsula; the other is in the north eastern region of Friuli, at the crossroads between Italy, Austria and Slovenia. In terms of PISA, Friuli is the Italian region with the highest marks (the school in question is one of the highest ranking in Italy), while Lazio, the region of Rome, is in a grey zone of mediocrity. In terms of autonomy Friuli has a long experience (being a special statute region). The school in Rome has been under the rule of the Ministry of Education in the long decades of centralised rule and only recently, with the enthusiastic spirit of the neophyte, has experimented with autonomy. The school in Friuli is a cosmopolitan scientific high school, sharing all the fashionable European and international slogans (from Knowledge Based Economy to Lifelong Learning) and programmes (from Leonardo to CLIL- *Content and Language Integrated Learning*). Students regularly participate to international contests such as the mathematics and physics Olympic Games. The school in Rome is a *Liceo Classico, Linguistico e Scientifico* which pursues excellence mostly in the traditional humanistic terms established by Giovanni Gentile almost a century ago. The most important subjects are the ancient languages, philosophy, history, Italian literature. The spearhead of internationalisation is the language department, which is entitled to award some international language certifications. It is also the area of the school where the concept of competencies is better known thanks, among others, to the introduction some years ago of the *European Portfolio of Languages*.

THE TWO SCHOOLS: DIFFERENCES IN THE WAY PISA HAS INFLUENCED THE ATTITUDE OF TEACHERS TOWARDS STUDENT MOBILITY

Due to these characteristics, the approach to PISA of the school in Friuli has been much more obvious and linear. The school regularly takes part in PISA surveys and has partially changed its POF in order to have better performances in the tests. In fact, pupils are systematically trained in this type of tests, and from such practices of teachers and pupils, the school has drawn the know-how that is then used for the evaluation of the returnees' competencies. PISA influence contributes to create amongst teachers the conviction that competencies acquired abroad are as important as the traditional disciplinary knowledge that the returnee might have not acquired by missing one year of Italian high school. This is why the assessment of returnees is based on competencies and tested through adequate instruments.

Another point is that teachers of this school say they are preparing pupils not for the final State exams, but for succeeding in higher education. This means that cross-curricular competencies and problem solving are as important in the school's curriculum as traditional disciplinary knowledge.

As already mentioned, local students participate in many international competitions. In the Physics Olympics, they have had a student that made it twice to the world selections and eventually ranked third.

Students take European language certifications, with an average 98% success rate, possibly the highest in Italy.

It must be said though, that regarding this school, PISA tests are just one of the beneficial international influences that contribute to create a favourable teacher attitude towards student exchanges and study periods abroad. The school makes a wide use of CLIL (Content and Language Integrated Learning) of which it has been a pioneer in Italy and, in summer, sends teachers abroad, to study foreign languages, English above all. Soon, eight staff members will go on a study period, including teachers of Mathematics, Sciences, Physical Education, and Italian. The teachers participate in CLIL as volunteers. This activity is usually developed by a language teacher and one of another discipline working together; for instance, a teacher of English works one or two “pedagogical units” with a Science colleague and they design the assessment tools. This type of experience constitutes a background that positively influences the attitude of the teaching staff towards the exchanges with the foreign countries, allowing teachers to acquire an “international sensibility”.

Regarding the second school, the path through which it met PISA, as a source of inspiration for making up for the lack of rules on how to assess returnees’ competencies, is much more complex and interesting. The school in Rome does not participate in PISA and therefore the approaching of the two was rather indirect. It all came through one of the rare “retraceable” influences that PISA has on the Italian institutional school assessment system.

As already mentioned, at the legislation level, the concept of competency is still not fully realised in the Italian secondary schools. It can only be found in junior high schools (*scuole medie*), but limitedly to the final exams (*licenza media*) where, since 2008, competencies have to be certified. Due to the void of clear institutional directives on competency assessment, the “strong” international model from which to draw inspiration is that of the PISA tests.

At this level, PISA leaks into the Italian System for at least three reasons:

- a) The prestige that OECD and PISA have in Italy (they are perceived as the source of one of the most outstanding international benchmarks), even if only very recently official policy has started to take it into account for specific actions.
- b) The fact that the Italian law, conscious of the lack of directions on the subject, allows the freedom to experiment with competency certification. The Italian law explicitly awards this freedom to junior high schools for the final State exams (Ministerial Circular Letter n. 51/2009, 5th clause).
- c) The PISA model is imported into this plexus via the INVALSI⁷ national test, which is part of the final junior high school exams. This test, one that should certify competencies, is based on models of international assessment, among which the PISA is expressly quoted (INVALSI, reference framework for the Italian language, p. 10; reference framework for Mathematics, p. 4).

The Liceo of Rome, object of my research, met the PISA model when it tried to design an experimental integrated system of “input-output” competency assessment for its pupils. This system has to measure the competencies of pupils when entering the Liceo, and when leaving at the moment of the final exams. It also states which are the competencies needed to access every course year. This is the moment when the system concerns returnees: when they come back from abroad, their level of competencies is tested to understand if they miss something they need in order to be reintegrated and, in that case, they do an update, a ‘crash course’, before the beginning of the new academic year.

This system, whose unifying element is the concept of competency, has been designed also in order to facilitate the passage from junior high school to Liceo, through a method of assessment/orientation connected to the junior high-school final exams (in which, as we have already mentioned, the PISA influence is tangible). Such a system is used in a local area network of junior high schools (allowed by the Presidential Decree n.275/1999, art. 7) from which the Liceo recruits its own students. The tests are conceived to facilitate the orientation of those newly licensed from junior high school, in their passage to Liceo. As said, this Liceo has various specialties (classical, scientific, linguistic) and the tests are needed for directing the freshmen towards the studies which better suit them.

In this case, the tests for competencies assessment are a plus (not an official item), and only for reasons of convenience they are administered together with final junior high school State exams. The elaboration of such tests, done by the staff of the involved schools, among which is the Liceo I have studied, is influenced by the PISA model (influence coming from the familiarity with the INVALSI test administered in the junior high school final exam) and is the source of a know-how on competency assessment, which within the described system, is also used for other purposes, such as the assessment of returnees’ competencies.

BY A WAY OF CONCLUSION

The familiarity with PISA (and other international programmes, based on competency assessment) makes the teachers less mistrustful towards what returnees have studied and learned abroad. Not only does it provide teachers with new instruments to assess competencies, but it makes them aware of the current development of a sort of transversal, international pedagogy of competencies, which legitimates student exchanges. At the end of the day –teachers tend to think – what returnees have learned abroad is not inconsistent with what is being taught at home.

Within the “intercultural exchanges”, these effects can be seen, in conclusion, as facilitating an international dialogue, but also as a cultural homologation, progressively impoverishing exchanges of one of their original pedagogical meanings, based on the educating power of the acculturation (and of the “acculturation shock”) to a foreign school system and to its environing culture.

NOTES

- ¹ The first part of this chapter – The framework – has been written by Donatella Palomba, the following parts by Anselmo R. Paolone
- ² Other aspects of this attitude will be discussed later, in the third paragraph
- ³ The Circular does not mention the subjects which exist only in the foreign schools.
- ⁴ In a sort of informal round table, the interviewees were invited to discuss among them the relevant topics. Their conversations were recorded, transcribed and analysed.
- ⁵ R. Cowen unpublished lecture held in the University of Rome “Tor Vergata”, 11–5–2009.
- ⁶ This perspective has been inspired by ethnographic researches such as: C. Lacey, *Hightown Grammar, the School as a Social System*, Manchester, Manchester University Press, 1970; the methodological approach has been influenced by: M. Osborn, “New Methodologies for Comparative Research? Establishing ‘constants and context’ in educational experience” in: *Oxford Review of Education*, vol. 30, n. 2, 2004; R. Webb, Vulliamy, G., S. Hamalainen, A. Sarja, E. Kimonen, R & Nevalainen, “A Comparative Analysis of Primary Teacher Professionalism in England and Finland” in: *Comparative Education*, vol. 40, n. 1, 2004; M. Osborn, “Constants and Contexts in Pupil Experience of Learning and Schooling: Comparing Learners in England, France and Denmark” in: *Comparative Education*, vol. 37, n. 3, 2001; G. Troman & B. Jeffrey, “Qualitative Data Analysis in Cross-Cultural Projects” in: *Comparative Education*, vol. 43, n. 4, 2007
- ⁷ INVALSI is the Italian acronym for ISTITUTO NAZIONALE PER LA VALUTAZIONE DEL SISTEMA EDUCATIVO DI ISTRUZIONE E DI FORMAZIONE (*National Institute for the Assessment of the School System.*)

REFERENCES

- Arnove, R.F. (2003). Introduction: Reframing comparative education. The dialectic of the global and the local. In R.F. Arnove & C.A.Torres (Eds.) *Comparative Education. The dialectics of the local and the global*. Lanham, Maryland: Rowman and Littlefield (second edition).
- Arnove, R.F. (2009). World-systems analysis and comparative education in the age of globalization. In R. Cowen & A.M. Kazamias (Eds.), *International Handbook of Comparative Education*. Heidelberg and London: New York, Springer.
- Canestri, G. (1983). *Centovent'anni di storia della scuola 1861/1983*. Torino: Loescher.
- Cowen, R. (2010). Transfer, translation and transformation. In A.R. Paolone (Ed.), *Education between boundaries*. Padova: Imprimerie .
- Favret-Saada, J. (1977). *Les mots, la mort, les sorts*. Paris: Gallimard.
- Favret-Saada, J. (1981). *Corps pour corps*. Paris: Gallimard.
- Fornaca, R. (1994). *Storia della scuola moderna e contemporanea*. Roma, Anicia.
- Lacey, C. (1970). *Hightown grammar, the school as a social system*, Manchester: Manchester University Press.
- Morandini, M.C. (2001). Da Boncompagni a Casati: la costruzione del sistema scolastico nazionale (1848–1861). In L. Pazzaglia & R. Sani (Eds.) *Scuola e società nell'Italia unita*. Brescia: La Scuola.
- Osborn, M. (2004). New Methodologies for Comparative Research? Establishing ‘constants and context’ in educational experience, *Oxford Review of Education*, 30 (2), 265–285.
- Osborn, M. (2001). Constants and contexts in pupil experience of learning and schooling: Comparing learners in England, France and Denmark *Comparative Education*, 37(3), 267–278.
- Palomba, D. (2009). Education and state formation in Italy. In R. Cowen & A.M. Kazamias (Eds.), *International Handbook of Comparative Education*. Heidelberg and London, New York: Springer.
- Paolone, A.R. (2006). Etnografi dell'educazione, soggettività, “terzo codice”. In M. Benadusi (Ed.), *Dislocare l'antropologia. Connessioni disciplinari e nuovi spazi epistemologici*. Rimini: Guaraldi.

D. PALOMBA AND A.R. PAOLONE

- Paolone, A.R. (2008). *Educazione comparata e etnografia tra globalizzazione e postmodernità*. Roma, Monolite.
- Paolone A.R. (2010a). I returnees di “lungo corso”. L’atteggiamento degli insegnanti. In D. Palomba, A.R. Paolone, C. Roverselli, O. Niceforo & C. Cappa. *Internazionalizzazione della scuola e mobilità studentesca. Il ruolo degli insegnanti*. Colle Val D’Elsa: Fondazione Intercultura.
- Paolone A.R. (Ed.) (2010b). *Education Between Boundaries*. Padova, Imprimitur.
- Talamo, G. (1960). *La scuola dalla legge Casati all’inchiesta del 1864*. Milano, Giuffrè.
- Troman, G., Jeffrey, B. (2007). Qualitative data analysis in cross-cultural projects *Comparative Education*, 43(4), 511–525.
- Turner, V. (1974). *Dramas, fields, metaphors*. Ithaca, NY: Cornell University Press.
- Turner, V. (1982). *The ritual process*. Ithaca, NY: Cornell University Press.
- Webb, R., Vulliamy, G., Hamalainen, S., Sarja, A., Kimonen, E. & Nevalainen, R. (2004). A comparative analysis of primary teacher professionalism in England and Finland. *Comparative Education*, 40(1), 83–107.

Donatella Palomba
Università di Roma “Tor Vergata”

Anselmo R. Paolone
Università di Roma “Tor Vergata”
Università di Udine

SECTION III

THE ASSESSMENT OF PISA, SCHOOL EFFECTIVENESS AND THE SOCIO-CULTURAL DIMENSION

KATHARINA MAAG MERKI

THE INTRODUCTION OF STATE-WIDE EXIT EXAMINATIONS

Empirical Effects on Math and English Teaching in German Academically Oriented Secondary Schools

INTRODUCTION

International comparative achievement studies like PISA have various functions. First and foremost, they serve as macro-level monitoring instruments, providing information to policymakers and other stakeholders on how students in their country compare, in terms of academic achievement, to students in other countries. Their methodological design entails that they are conducted at regular three-year intervals, which allows them to reflect developments at the macro-level—for example, changes in average achievement levels, in the homogeneity of achievement, or changes in group size at individual competence levels. They do not, however, serve a diagnostic function regarding achievement levels or their evolution over time at the individual (student) level, class level, or school level. Given their distance from the micro-level activities of teachers, which are one of the key conditions for teacher professionalization (Day, 1999; Day & Sachs, 2004; Hopkins, 2005), international comparative achievement studies have a limited impact on teaching-quality. The results of external achievement tests conducted under the US “No Child Left Behind Act” of 2001 do, however, suggest “that policies can penetrate classroom boundaries” (Hamilton, Stecher, Russell, Marsh, & Miles, 2008, p. 38). Thus, macro-level strategies do appear to play an important role in teaching practice under specific conditions, although the findings on this subject tend to suggest more negative than positive effects.

While changes in teaching-quality accompanying participation in international comparative studies cannot be investigated in the framework of the PISA studies, these changes can be analyzed by focusing on another macro-level monitoring instrument: the state-wide Abitur exit examination. The debate over Germany’s mediocre PISA results led to the introduction of state-wide Abitur exit examinations in all of the states (Germany’s Bundesländer, totaling 16 since unification) that had not already instituted them after the end of World War II. The question is to what extent these state-wide exit examinations are effective macro-level monitoring instruments for teaching-quality in the final year of upper secondary education (Gymnasien, academically oriented secondary school) and whether they produce positive or negative changes. Unfortunately, there are virtually no empirically

sound findings on this subject for German-speaking countries. There is also an acute lack of longitudinal studies examining the transition from a school, or class-based exit examination system to a state-wide one. The present study aims to fill this research gap.

2. CLARIFYING THE CONCEPTS

The German Abitur exit examination is an output-based testing procedure administered at the end of Gymnasia, an academically oriented upper secondary school. Independent of whether the examinations are developed and administered in a centralized (at the state level) or decentralized manner (at the school or class level), the Abitur exit examinations are important for students since they are generally required for enrollment in university studies. In this sense, the Abitur exit examinations have a selective function. The main difference between class-based and state-wide examination systems is related to the question: who develops the examination tests? While it used to be that the teacher of each individual subject and course designed the examination tests for his or her class — which meant that students in the same subjects in the same states had to complete different tests in each of the different classes — now there are standardized, externally developed exams in one subject for all the schools, and courses in the entire state. Grading is still done in a class-based manner, that is, by the individual teacher of the class. However, to standardize grading, unified grading guidelines have been developed and given to the teachers.

International comparative analyses have shown that state-wide exit examination systems are far from identical from one country to the next (Klein, Kühn, Van Ackeren, & Block, 2009). The only aspect they all share in common is that the examinations are administered at a single point in time across an entire country or state, under the supervision of the course instructor, who is also the first examiner. The systems differ in a number of respects, for example, in the weight assigned to exit examination scores in the final grade. This varies significantly between countries, and differences are also found in the degree of selectivity in examination procedures. According to Klein, Kühn & van Ackeren (2009), the majority of German states prefer the organizational form of state-wide exit exams, which represents a low to moderate level of standardization for Germany, compared with the 15 other OECD countries. In contrast to the state-wide high school exit examinations required for graduation in some US states, the concepts used in the majority of OECD countries weight students' grades on the centralized exit examinations in terms of their final secondary school grades. In the US, final grades are based solely on final exam performance, without taking any class grades into account, although some US states have recently started moving toward end-of-course exit exams (Zabala, Minnici, McMurrer, & Briggs, 2008). In addition, the option to choose one's individual exam subjects does not exist in the US to the same degree as in Germany. The selection potential of state-wide exit examinations as introduced in Germany is thus estimated as relatively low compared to the US. Furthermore, in Germany, state-wide exit examinations were not introduced as a

high-stakes practice for teachers and schools, and are therefore not used as a basis for determining wage increases, layoffs, or school closures, as has been done in several US states.

3. REVIEW OF THE EMPIRICAL RESEARCH

The introduction of state-wide Abitur exit examinations at the end of Gymnasia school is based on the assumption that these examinations lead to greater standardization of evaluation procedures across classes and across schools, to higher teaching-quality, and to improved student achievement (Hamilton et al., 2008; Maag Merki, 2010). This argument cites the low level of standardization among classes, schools, and states that has been identified in the German educational system (Baumert & Watermann, 2000; Köller, Baumert, Cortina, Trautwein, & Watermann, 2004; Köller, Baumert, & Schnabel, 1999; Neumann, Nagy, Trautwein, & Lüdtke, 2009). The question arises, however, whether this assumption is justified, and whether the introduction of state-wide exit examinations has a positive effect on instructional quality.

In the Anglo-American countries, several studies have examined the effects of state-wide high school exit examinations on the classroom behavior of teachers and students. These studies identified some isolated positive effects, but primarily negative effects (e.g., Abrams, 2007; Amrein & Berliner, 2002; Stecher, 2002), which indicate that instructional quality declines in the context of rigid monitoring systems with strong control mechanisms and severe negative consequences for students, teachers, and school administrators (high-stakes testing). This reduction in instructional quality is manifested, for example, in a narrowing of the curriculum to focus on test content or an adaptation of teaching or examination methods to the test format (e.g., Au, 2007; Hamilton et al., 2008; Herman, 2004; Swanson & Stevenson, 2002). Previous experience with low-stakes testing, on the other hand, shows that the potential for productive changes in instructional practice is far greater in non-punitive systems than in high-stakes systems. Low-stakes systems appear to produce a much lower degree of undesirable, unintended effects than sanctions-oriented systems (Brozo & Hargis, 2003; Clarke et al., 2003; Pedulla et al., 2003).

In German-speaking countries, empirical findings addressing the effects of state-wide exit examinations on instructional practice are sparse. In a recent study, my co-author and I produced initial findings (Maag Merki & Holmeier, 2008) based on data collected since the first year of introduction of state-wide Abitur exit examinations in Bremen, that instructors of courses with state-wide examinations tend to reduce the range of themes dealt with in class and are less responsive to students' interests and current everyday issues than in courses with decentralized examinations. Only a small group of teachers (less than 12%) showed evidence of weighting courses with state-wide exit examinations differently to courses with school or class-based exit examinations in terms of lesson preparation and efforts to foster student performance. In terms of instructional quality, we found that the introduction of state-wide Abitur exit examinations at the end of upper secondary school only in basic-level courses leads to increased perceived cognitive activation

and support by teachers. This effect does not seem to appear when state-wide exit examinations are administered in all three written exam subjects, as was the case in Hessen. Studies analyzing the change from a class-based to a state-wide examination system comparing two years in a specific subject (Maag Merki, Holmeier, Jäger, & Oerke, 2010) have also demonstrated that effects vary from one subject to the next. While positive effects on instructional quality were found in Math and particularly in English, none were found in Biology or German. Furthermore, if the testing system remains stable (as in Hessen), one finds only minor changes in teaching-quality. Additional analyses (Maag Merki, Klieme, & Holmeier, 2008) point to systematic differences among schools with regard to instructional quality in basic and advanced-level courses, whereby an increased focus on basic-level courses with state-wide exit examinations is not necessarily accompanied by decreased instructional quality in advanced-level courses with class-based exit examinations.

These analyses provided the basis for studying the change of system from a class-based to a state-wide testing procedure by comparing the year before introduction of the new system to the year after. The question remains, however, to what extent these short-term effects remain stable over time. This question can be addressed in the framework of a longitudinal study carried out in Germany from 2007 to 2009. In 2008, state-wide Abitur exit examinations were introduced in Bremen in some advanced-level courses (e.g., English and Math) at the end of Gymnasia school. By comparing these results with the data from 2007 under a class-based system, we can study changes in instructional quality accompanying the change from a class-based (2007) to a state-wide (2008) system. When taking the results for 2009 into account as well, we can also study the longer-term trends in instructional quality from 2007 to 2009 and thus examine the stability of the results over time.

The plausibility of the results in advanced courses is assessed by comparing them to two different groups of courses: a) to changes in instructional quality in the basic courses in which state-wide exit examinations have been used since 2007 and b) to changes in instructional quality in the advanced and basic courses in a different German state (Hessen), in which state-wide exit examinations were introduced in all subjects in 2007.

4. QUESTIONS AND HYPOTHESES

This article seeks to address the following questions:

1. To what extent does instructional quality change in the final year of Gymnasia in advanced courses in Bremen, where exit examinations were administered class-based in 2007 and state-wide in 2008?
2. To what extent is this change attributable to the introduction of state-wide exit examinations?

In line with the empirical findings from Anglo-American countries (Hamilton et al., 2008), and based on the comparison of the first two years of the study (Maag

Merki et al., 2010), Hypothesis 1 assumes some longer-term effects of the introduction of state-wide exit examinations on instructional quality in advanced courses. Moreover, we have to expect differential effects in terms of subject (Baumert & Watermann, 2000; Maag Merki et al., 2010) (Hypotheses 2). Since low-stake testing was introduced in Bremen and Hessen, one can also assume that we will tend to see positive changes in terms of improved instructional quality (Brozo & Hargis, 2003; Clarke et al., 2003; Pedulla et al., 2003) (Hypothesis 3). These changes will probably become most evident with a change of testing system (from 2007 to 2009), and less when the testing system remains stable, as was the case from 2008 to 2009 (Maag Merki et al., 2010) (Hypothesis 4).

5. METHODS

5.1 Survey and analysis procedures

The questions are addressed in the framework of a three-year longitudinal study carried out in the states of Bremen and Hessen in Germany. In these two states, the Gymnasia introduced state-wide exit examinations for the first time in 2007. In Hessen, state-wide exit examinations were introduced in all subjects and all courses (both advanced and basic) in 2007. In Bremen, state-wide exit examinations were introduced in 2007 only in the basic courses, and in 2008 in advanced German, Math, Natural Sciences, and advanced Foreign Language courses, while class-based testing continued in the rest of the advanced courses there.

In order to compare the effects of introducing state-wide exit examinations across different German states, the two systems should ideally be identical, with the only difference being the two different points in time at which they were implemented, as described above. The two systems are not entirely identical; however, as comparative analyses by Klein, Kühn, Van Ackeren & Block (2009) have shown, substantial overlaps do exist between the two systems, suggesting that these two states can reasonably be used for comparison. Both states have chosen organizational forms that are very similar in terms of how the main topics are determined, how test tasks are developed, and how the examinations are administered and graded. Furthermore, the two states show a low level of standardization in comparison with the other OECD countries (Klein et al., 2009) or even the United States. Thus, bad exam results cannot result in severe consequences for schools or teachers, which means that for teachers and schools, both procedures under examination here can be referred to as “low-stake” testing procedures.

Differences between these two systems exist probably in the subject-specific demands, the range of key topics, the levels of expectations, and the criteria used for grading and evaluation. However, there has been no systematic comparative analysis of this issue to date. It should also be kept in mind that any differential changes in instructional practice may also be the result of other factors like different learning cultures, rules, or structures in the two states that have nothing to do with the exit examination system.

To address these issues, we have conducted descriptive and inference statistical analyses and multilevel analyses (HLM 6.06). The focus here is on the subjects of Math and English. It was in these subjects – which are among the most popular subjects in Germany – that the change of exam system took place in advanced-level courses. Furthermore, significant positive effects on instructional quality had already been identified when comparing the first two years of data (Maag Merki et al., 2010). These facts made an analysis of the stability of the findings in these subjects particularly interesting.

In the framework of this study, following on from the work of Jürges, Schneider & Büchel (2005), we computed difference-in-differences estimates. With this method, we estimated a possible effect of introducing state-wide exit examinations in the Gymnasias, comparing the two states of Bremen and Hessen, the two course types (basic and advanced), and three consecutive years. We assumed that the differences between the two states and course types would be smaller when state-wide examinations were administered in both states and both course types, as was the case in 2008 and 2009, but that the differences would be greater when state-wide examinations were administered in one state (Hessen) and one course type (basic courses) and class-based examinations were administered in the other (Bremen respectively advanced courses), as was the case in 2007. As a result, we expected different annual effects to emerge.

To address these issues, we collected various indicators to measure cognitively stimulating and supportive teaching (Klieme, 2006) through standardized questionnaires for students. All students evaluated how they perceived teaching-quality in their three written exam subjects, enabling subject-specific evaluations to be carried out.

Introductory sentence: To what extent do the following statements apply to instruction in your various exam subjects?

Answer scale: 1 = does not apply at all ... 4 = applies completely

- *Elaboration (4 Items):* “During class, we frequently have the opportunity to link the skills acquired in that subject to skills acquired in other subjects.” *Cronbach’s Alpha:* advanced Math courses: .65; basic Math courses: .68; advanced English courses: .62; basic English courses: .62; *Intraclass-Correlation:* advanced Math courses: .027; basic Math courses: .015; advanced English courses: .095¹; *Source:* Leutwyler & Maag Merki (2005)
- *Teacher capacity to motivate students (5 Items):* “Our teacher is sometimes really inspiring to us students.” *Cronbach’s Alpha:* advanced Math courses: .79; basic Math courses: .78; advanced English courses: .82; basic English courses: .80; *Intraclass-Correlation:* advanced Math courses: .061; basic Math courses: .062; advanced English courses: .089; *Source:* Leutwyler & Maag Merki (2005)
- *Autonomy support (4 Items):* “During class, I have the opportunity to explore new issues autonomously.” *Cronbach’s Alpha:* advanced Math courses: .65; basic Math courses: .65; advanced English courses: .70; basic English courses: .63; *Intraclass-Correlation:* advanced Math courses: .020; basic Math courses: .014; advanced English courses: .068; *Source:* Prenzel et al., (1996)

- *Competence support (5 Items)*: “During class, the teacher regularly lets me know how I’m progressing.” *Cronbach’s Alpha*: advanced Math courses: .75; basic Math courses: .77; advanced English courses: .79; basic English courses: .73; *Intraclass-Correlation*: advanced Math courses: .029; basic Math courses: .037; advanced English courses: .042; *Source*: Prenzel, Kristen, Dengler, Ettl & Beer (1996)

The reliability of the scales on the individual level is moderate to high. Based on the scale values, missing values were imputed using multiple imputations in SPSS 18 (Graham, 2009; Lüdtke, Robitzsch, Trautwein, & Köller, 2007).² Here, ten data sets were produced in which plausible estimated values were included for the missing values that vary slightly among the data sets.

In a first step of the analyses procedure, the descriptive and inferential statistical analyses were carried out within the two states. This means that based on the Bremen and Hessen data, differences between the years were calculated. Multivariate regression analyses with corresponding dummy variables were performed. The descriptive statistics reported in the following combine the values of the individual data sets using the formula developed by Rubin (1987).

The second step was to conduct multilevel analyses of the four subjects. Multilevel analyses estimate the annual effects as a function of the specific school attended. This allows us to take into account that the students’ evaluations of their advanced courses are not necessarily independent of each other, but that they might vary by school attended since the different schools also represent different learning environments.

The estimates are based on a two-level model, using the teaching indicator as estimated from the students’ perspective as the dependent variable. The independent variables used at Level 1 are the dummy-variables “year08” (1=2008) and the variable “year09” (1=2009) and at Level 2 the variable “state” (0=Hessen, 1=Bremen). Both the fixed effects and the random effects are included in the regression equation, and the independent variables are entered non-centered into the analyses. The respective equation is:

$$\text{Teaching indicator} = G_{00} + G_{01}*(state) + G_{10}*(year08) + G_{20}*(year09) + G_{11}*(state)*(year08) + G_{21}*(state)*(year09) + U_0 + U_1*(year08) + U_2*(year09) + R$$

5.2 Data Collection and Sample

The indicators were collected by written questionnaires in 37 Gymnasia schools preparing students for the Abitur exit examination. The survey was conducted in early February, prior to written exit examinations, and was administered at the same schools over the three years. In Bremen all 19 Gymnasia with upper levels were analyzed. In Hessen, 18 schools were selected for analysis according to specific criteria (region, city/rural, school size, school focus), in order to attain the most representative sample possible for a single state.

Within the schools, students in four final-year classes were surveyed: they were selected from one advanced English and one advanced Math course, and from one

basic English and one basic Math course. The students were asked to assess instruction in their three written exit exam subjects, independent of which course they were surveyed in as part of this study. This focus was important in order to be able to use the students' evaluations for more in-depth analyses (e.g., to study the relationship between teaching-quality and student performance on the written exit exam). However, it should also be kept in mind that at larger schools, not all of the students who were taking, for example, advanced Math as an exam subject, were actually in the same advanced Math course. For this reason, the schools and not the courses, are used as Level 2-variable in the multilevel analyses, and as a result, the Level 2 effects tend to be underestimated. Previous analyses (Maag Merki & Holmeier, 2008) do indicate, however, that different learning environments can be identified through the formation of clusters at the school level.

Table 1. Sample

	<i>Total</i>				<i>Bremen</i>				<i>Hessen</i>			
	<i>total</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>total</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>total</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>
<i>Advanced courses</i>												
Math	1961	600	649	712	897	253	306	338	1064	347	343	374
English	2573	845	868	860	1236	408	422	406	1337	437	446	454
<i>Basic courses</i>												
Math	1923	567	639	717	479	152	170	157	1444	415	469	560
English	685	186	232	267	493	130	155	208	192	56	77	59

Students' rates of response to the questionnaires in Hessen were relatively high at 68% (2007), 71% (2008), and 74% (2009) and stable. In Bremen, the response rate in the first year was somewhat lower, at 52%, due to organizational problems in four schools, but response rates in 2008 and 2009 were comparable to Hessen (68% and 71%, respectively). Of the total 6,331 students that participated in all three years, 1,961 students chose Math and 2,573 students chose English as their advanced course for the exit exam, 1,923 students chose Math and 685 students chose English as their basic course (see Table 1). The relatively small group size in the basic English courses in Hessen should also be kept in mind. However, this result is not due to a low response rate. In contrast, very few students in Hessen took these courses as examination subject.

Table 2. Sample Multilevel Analyses

<i>course</i>	<i>reduction pupils</i>	<i>final sample HLM</i>	
Math advanced courses	3.6%	34 schools	1891 pupils
Math basic courses	8.6%	28 schools	1757 pupils
English advanced courses	0.5%	36 schools	2559 pupils

Since the courses were not selected in all of the schools and in all of the three years, for the multilevel analyses, the sample was reduced. Additionally, only those schools were integrated into the analyses which had at least five students every year in the different courses (cf. [table 2](#)). Due to the small sample in the English basic courses multilevel analyses couldn't be computed.

6. RESULTS

6.1 Individual-level Analyses

[Table 2](#) presents the results of the analyses for the subject of Math on the individual level. Overall, taking into account the three year period, no negative effects on instructional quality were identified. In some areas, however, positive, but small effects appeared.

In the *advanced Math courses* in Bremen, significant annual effects in the dimension of “elaboration” from 2007 to 2008 (standardized coefficient $\beta = .17$, $p < .001$) appeared with the positive change from a class-based to a state-wide testing system. Yet no further changes were identifiable between 2008 and 2009 – both years with state-wide exit examinations. Over the three year period (2007 – 2009), the effect remained almost stable (standardized coefficient $\beta = .15$, $p < .001$). In Hessen, where exit examinations were organized state-wide for the entire period, no significant changes were identified between the three years in this dimension. In the advanced courses in Bremen, one additional significant effect was found: In a three year period from 2007 to 2009, there was a significant increase in perceived “competence support” (standardized coefficient $\beta = .09$, $p < .05$).

In Hessen, the two dimensions “teacher capacity to motivate students” and “competence support” each showed significant effects. With regard to the first dimension, from 2007 to 2008 a negative effect appeared (standardized coefficient $\beta = -.09$, $p < .05$). However, from 2008 to 2009 the perceived “teacher capacity to motivate students” increased again significantly (standardized coefficient = $.16$, $p < .001$). Consequently, we found a positive change by trend from 2007 to 2009 (standardized coefficient $\beta = .06$, $p < .10$). Considering the second dimension “competence support”, we identified a positive change, too. From 2007 and 2008, respectively, to 2009, the perceived competence support increased significantly (standardized coefficient $\beta = .09$, $p < .01$).

In the *basic courses* with a stable testing system, two significant effects occurred in Hessen. From 2007 to 2008, we saw a significant increase in the two dimensions “teacher capacity to motivate students” (standardized coefficient $\beta = .07$, $p < .05$) and “autonomy support” (standardized coefficient $\beta = .09$, $p < .05$). However, in both dimensions these effects didn't remain stable. Consequently, we didn't find any significant effects in a longer term from 2007 to 2009.

In Bremen, however, which also had a stable testing system in the basic courses, changes in two dimensions were identified, both of which indicate an improvement in instructional quality in a three year period. This is the case in the dimension “competence support” with a significant effect between 2007 and 2009 (standardized

coefficient $\beta = .13$, $p < .05$) and 2008 and 2009 (standardized coefficient $\beta = .12$, $p < .05$), respectively. Additionally, the perceived “teacher capacity to motivate students” increased from 2007 to 2009 by trend (standardized coefficient $\beta = .11$, $p < .10$).

Table 2. Instructional quality in Math courses

		Bremen			Hessen		
		N	M	SD	N	M	SD
<i>Advanced Math courses</i>							
elaboration	2007 cb	253	1.88	0.65	347	2.06	0.69
	2008 sw	306	2.13	0.67	343	1.98	0.59
	2009 sw	338	2.10	0.70	374	2.04	0.61
teacher capacity to motivate students	2007 cb	253	2.57	0.69	347	2.74	0.60
	2008 sw	306	2.51	0.66	343	2.61	0.70
	2009 sw	338	2.60	0.71	374	2.83	0.67
autonomy support	2007 cb	253	2.51	0.63	347	2.48	0.62
	2008 sw	306	2.50	0.64	343	2.49	0.65
	2009 sw	338	2.52	0.63	374	2.55	0.59
competence support	2007 cb	253	2.47	0.67	347	2.55	0.64
	2008 sw	306	2.49	0.66	343	2.56	0.63
	2009 sw	338	2.59	0.66	374	2.69	0.61
<i>Basic Math courses</i>							
elaboration	2007 sw	152	1.79	0.68	415	1.78	0.62
	2008 sw	170	1.83	0.65	469	1.85	0.63
	2009 sw	157	1.76	0.61	560	1.79	0.61
teacher capacity to motivate students	2007 sw	152	2.38	0.74	415	2.56	0.67
	2008 sw	170	2.41	0.70	469	2.65	0.65
	2009 sw	157	2.54	0.69	560	2.54	0.64
autonomy support	2007 sw	152	2.22	0.66	415	2.26	0.62
	2008 sw	170	2.27	0.61	469	2.37	0.62
	2009 sw	157	2.35	0.63	560	2.32	0.59
competence support	2007 sw	152	2.52	0.76	415	2.52	0.63
	2008 sw	170	2.53	0.70	469	2.58	0.68
	2009 sw	157	2.71	0.67	560	2.50	0.67

Response scale (Likert scale): 1 = do not agree at all ... 4 = agree completely;

M = combined mean; SD = combined standard deviation (naive pooling); SE = standard error of the mean; N = numbers; cb = class-based examination system, sw = state-wide examination system; Level of significance: * $p < .05$, ** $p < .01$, *** $p < .001$, n.s. = not significant;

With reference to instructional quality in the *English courses* (see Table 3), significant changes could only be identified in the advanced courses, all of which showed positive developments in instructional quality.

THE INTRODUCTION OF STATE-WIDE EXIT EXAMINATIONS

Table 3. Instructional quality in English courses

		<i>Bremen</i>			<i>Hessen</i>		
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
<i>English Advanced courses</i>							
elaboration	2007 cb	408	2.44	0.67	437	2.70	0.61
	2008 sw	422	2.69	0.63	446	2.76	0.59
	2009 sw	406	2.70	0.61	454	2.73	0.62
teacher capacity to motivate students	2007 cb	408	2.29	0.72	437	2.50	0.72
	2008 sw	422	2.53	0.70	446	2.59	0.68
	2009 sw	406	2.56	0.68	454	2.65	0.69
autonomy support	2007 cb	408	2.31	0.67	437	2.44	0.59
	2008 sw	422	2.44	0.63	446	2.56	0.61
	2009 sw	406	2.46	0.63	454	2.50	0.63
competence support	2007 cb	408	2.37	0.70	437	2.53	0.67
	2008 sw	422	2.47	0.69	446	2.64	0.68
	2009 sw	406	2.50	0.64	454	2.59	0.67
<i>English Basic courses</i>							
elaboration	2007 sw	130	2.51	0.68	56	2.49	0.54
	2008 sw	155	2.42	0.67	77	2.58	0.60
	2009 sw	208	2.52	0.61	59	2.56	0.61
teacher capacity to motivate students	2007 sw	130	2.48	0.66	56	2.65	0.68
	2008 sw	155	2.52	0.64	77	2.65	0.62
	2009 sw	208	2.56	0.68	59	2.62	0.69
autonomy support	2007 sw	130	2.39	0.60	56	2.55	0.59
	2008 sw	155	2.43	0.59	77	2.52	0.60
	2009 sw	208	2.45	0.57	59	2.44	0.59
competence support	2007 sw	130	2.61	0.61	56	2.68	0.59
	2008 sw	155	2.77	0.67	77	2.79	0.62
	2009 sw	208	2.69	0.62	59	2.71	0.56

Response scale (Likert scale): 1 = do not agree at all ... 4 = agree completely;

M = combined mean; *SD* = combined standard deviation (naive pooling); *SE* = standard error of the mean; *N* = numbers; *cb* = class-based examination system, *sw* = state-wide examination system; Level of significance: * $p < .05$, ** $p < .01$, *** $p < .001$, n.s. = not significant;

In Bremen, positive annual changes in the *advanced courses* can be seen in all four dimensions. However, the effects are rather weak, especially in the last two dimensions “autonomy support” and “competence support”. It is apparent that instructional quality improved mainly from 2007 to 2008 with the significant change from a class-based to a state-wide examination system, without any further significant changes in instructional quality from 2008 to 2009, when the testing system remained stable:

- “elaboration”: 2007-2008: standardized coefficient $\beta = .13$ ($p < .001$); 2007-2009: standardized coefficient $\beta = .14$ ($p < .001$)
- “teacher capacity to motivate students”: 2007-2008: standardized coefficient $\beta = .12$ ($p < .001$); 2007-2009: standardized coefficient $\beta = .13$ ($p < .001$)
- “autonomy support”: 2007-2008: standardized coefficient $\beta = .07$ ($p < .05$); 2007-2009: standardized coefficient $\beta = .08$ ($p < .01$)
- “competence support”: 2007-2008: standardized coefficient $\beta = .05$ ($p < .05$); 2007-2009: standardized coefficient $\beta = .06$ ($p < .05$)

Comparing the three years in the advanced courses in *Hessen*, where the testing system remained stable, a small positive change in instructional quality in a three-years-term were only found in the dimension of “teacher capacity to motivate students” (standardized coefficient $\beta = .07$, $p < .01$). A positive change by trend was already identifiable from 2007 to 2008 (standardized coefficient $\beta = .04$, $p < .10$).

A small positive change from 2007 to 2008 also appeared in the dimensions “autonomy support” (standardized coefficient $\beta = .06$, $p < .05$) and “competence support” (standardized coefficient $\beta = .05$, $p < .05$), but the differences between 2007 and 2009 were not significant anymore due to a small, non-significant reduction from 2008 to 2009.

In the *basic courses* in Bremen with a stable testing system, the instructional quality increased slightly only in the dimension “competence support” from 2007 to 2008 (standardized coefficient $\beta = .05$, $p < .05$) without any longer effect on 2009, however. In the basic courses in *Hessen* we didn’t find any significant year effects.

6.2 Multilevel analyses

Taking the multilevel structure of the data into account, multilevel analyses were then computed. With regard to the *advanced Math courses*, significant effects related to the main questions only appeared in the dimension of “elaboration” (see [Table 4](#)). Here, the interaction effects between the dummy variable “year 2008” and “state” are significant (nonstandardized coefficient $\beta = 0.384$, $p < .05$). This result indicates that the dimension “elaboration” showed a stronger positive change between 2007 and 2008 in the schools in Bremen with the change from a class-based to a state-wide examination system than in the schools in *Hessen* with a stable state-wide examination system. Considering the longer-term change between 2007 and 2009, the interaction effect was smaller and only significant by trend (nonstandardized coefficient $\beta = 0.238$, $p < .10$), which can be explained by the small non significant decrease from 2008 to 2009 in the dimension “elaboration”. Additionally, we identified a main effect on level 2 which means that the level of elaboration in math advanced courses is higher in *Hessen* than in Bremen (nonstandardized coefficient $\beta = -0.195$, $p < .05$). Year-specific analyses (t-test, independent samples), however, showed a significant mean difference only in 2007 ($t = 3.104$, $df = 839$, $p < .002$) and 2008 ($t = -3.058$, $df = 803$, $p < .002$), but not anymore in 2009 ($t = -1.181$, $df = 236$, n.s.). With regard to the other instructional variables examined here, no systematic changes or interaction effects were found.

THE INTRODUCTION OF STATE-WIDE EXIT EXAMINATIONS

Similar results were seen for the *advanced English courses*. Here, too, a significant interaction effect was found. However, the interaction effect from 2007 to 2008 was only significant by trend (nonstandardized coefficient $\beta = 0.180$, $p < .10$), whereas the longer-term effect from 2007 to 2009 was significant on a 5%-level of significance (nonstandardized coefficient $\beta = 0.209$, $p < .05$). Based on these findings, we can assume that instructional quality in the advanced English courses improved from 2007 to 2009 more strongly in Bremen with a change in the testing system than in Hessen with a stable testing system. Again, we saw a main effect on level 2 due to a higher level of elaboration in the English advanced courses in Hessen than in Bremen (nonstandardized coefficient $\beta = -0.272$, $p < .05$). Contrary to the results in Math advanced courses, however, year-specific analyses (t-test, independent samples) show only in 2007 a significant mean difference between the two states ($t = 5.828$, $df = 1797$, $p < .001$), but not in 2008 ($t = 1.444$, $df = 151$, n.s.) and in 2009 ($t = 0.619$, $df = 459$, n.s.).

Table 4. Multilevel analyses (HLM) for advanced courses

	Math "Elaboration" (ICC: 0.027)	English "Elaboration" (ICC: 0.095)
Intercept (1 to 4)	2.08*** (0.061)	2.736*** (0.066)
<i>Level 1</i>		
Dummy: year 2008 (1=2008)	-0.106, n.s. (0.090)	0.039, n.s. (0.066)
Dummy: year 2009 (1=2009)	-0.054, n.s. (0.091)	0.013, n.s. (0.063)
<i>Level 2</i>		
state (0=Hessen, 1=Bremen)	-0.195* (0.084)	-0.272* (0.090)
year 08 *state	0.384** (0.126)	0.180 ⁺ (0.106)
year 09 *state	0.238 ⁺ (0.136)	0.209* (0.191)
<i>random effects</i>		
u ₀ (Level 2)	0.041***	0.059***
u ₁ (slope year 2008)	0.096***	0.069***
u ₂ (slope year 2009)	0.121***	0.062***
r	0.383	0.337

nonstandardized coefficients (stand. error)

Level of significance: * $p < .05$, ** $p < .01$, *** $p < .001$, + $p < .10$, n.s. = non-significant;

In *Math basic courses*, we didn't identify any significant main or cross-level interaction effects which mean that the changes in Hessen and Bremen, both with a stable state-wide testing system, were similar.

6. DISCUSSION

This article has examined the question of whether the introduction of state-wide Abitur exit examinations in English and Math advanced courses in Bremen was accompanied by an improvement in instructional quality in those courses (Question 1). In line with our expectations (Hypotheses 1 and 3), the results indicate positive

effects on instructional quality, and in fact none of the analyses showed negative longer-term effects. Additionally, there are differential effects in terms of subject (Hypothesis 2). The main effects are seen in advanced English courses, where students assessed instructional quality much lower in 2007 under a class-based exit examination system than in 2008 and 2009 under a state-wide exit examination system. In advanced Math courses we find a similar result only for the dimensions “elaboration” and “teacher capacity to motivate students”. While the effects of introducing state-wide Abitur exit examinations are already confirmed by early analyses from a short-term perspective (Maag Merki et al., 2010), the present study shows that the positive effects remain stable over time, since instructional quality changed mainly from 2007 to 2008 but not from 2008 to 2009 (with one exception in the dimension “competence support” where the significant change is only seen in the second year of implementation).

The question arises as to what extent these effects can be identified as the result of introducing state-wide Abitur exit examinations (Question 2). In Bremen, where state-wide exit examinations in the basic courses were introduced in all subjects in 2007, the comparison between *advanced English and basic English* courses already provides some information for the interpretation: it shows that instructional quality changed from one year to the next particularly in the advanced courses but not in the basic courses with a stable testing system. This could suggest that the changes in instructional quality in the advanced English courses were systematically linked to the change of testing system.

However, it could also be that these differential effects are systematically linked not to the introduction of state-wide exit examinations but to the differing levels of requirements in basic and advanced courses.

Evidence along this line can be found in comparison with Hessen, where state-wide exit examinations have been used in both basic and advanced courses since 2007. In the advanced English courses, accounting for the multilevel structure of the data, the results show differential developments between the two states particularly in the area of cognitive activation in the classroom (“elaboration”). The results in this dimension indicated systematic, positive developments in the advanced courses in Bremen but not in Hessen. In the other dimensions, there was no evidence of differential developments. Thus, at least for the degree of perceived cognitive activation in the English advanced courses, it can be assumed that the changes in Bremen indicate a systematic relationship with the introduction of state-wide exit examinations.

Considering the English basic courses – in which state-wide examinations have been administered in both Bremen and Hessen since 2007 – the comparison between Bremen and Hessen (without being able to account for the multilevel structure) shows in the basic courses only marginal annual differences between the two states. Therefore, we can assume that the annual changes in the basic courses were quite similar in the two states.

In the *subject of Math*, state-specific effects that would indicate a positive impact of introducing state-wide exit examinations were found in only one case: in the dimension of “elaboration.” These result suggests that in advanced Math

THE INTRODUCTION OF STATE-WIDE EXIT EXAMINATIONS

courses in Bremen, as with advanced English courses, cognitive activation showed a more positive development with the change from a class-based to a state-wide testing system than in Hessen, where the testing system remained stable. In the other dimensions, no differential effects were identified. These results can be confirmed even after controlling for students' sex and general cognitive ability as possible influencing variables (Maag Merki, in press).

In the basic Math courses, two positive changes were identified under a stable testing system: students in Bremen perceived stronger support from their teachers in the areas of "teacher capacity to motivate students" and "competence support" than students in Hessen. However, taking the multilevel structure into account, these differential effects could not be confirmed, and therefore the results suggest that developments in the basic courses Math were comparable in Bremen and Hessen.

These findings support the hypothesis that the introduction of state-wide exit examinations – taking into account differences between subjects in the advanced courses – co-varies with positive improvements in instructional quality, particularly with the degree of cognitive activation in the classroom. The results provide no evidence of negative developments. However, the effects are rather small. With regard to the explanation of differential effects, further analyses are needed, focusing on specific subjects and teaching methods, including the analysis of test tasks, in order to determine possible reasons for this.

Overall, the results appear to suggest that state-wide Abitur exit examinations like those introduced in Germany are accompanied by more productive outcomes for instructional quality than the high-stake testing procedures used in the US. The low-stake procedures allow teachers more room to employ functional approaches that can be tailored to students' needs (Abrams, 2007; Clarke et al., 2003). It must be taken into account, however, that the dimensions under consideration here do not present a complete picture of instructional quality, and that the results reported are based on analyses covering a relatively short period of time. Furthermore, with regard to instructional quality, it needs to be tested whether the observed effects actually do result in improved student performance.

ACKNOWLEDGEMENTS

This study was funded by the German Research Foundation (DFG). We thank the Bremen and Hessian Ministries of Education and all the teachers and students of the schools that have participated in our study.

NOTES

- ¹ Due to the small sample in the English basic courses no multilevel analyses were conducted (see chapter 5.2).
- ² The reason to impute only scale values is related to the fact that with the given number of cases on $n > 6,000$, the software could not handle the large number of individual items despite maximal hardware requirements. The multilevel structure of the data set was taken into account in that for the multiple imputations dummy variables were entered for the units at the higher level (Graham, 2009).

REFERENCES

- Abrams, L. M. (2007). Implications of high-stakes testing for the use of formative classroom assessment. In J. H. McMillan (Ed.), *Formative classroom assessment. Theory into practice* (pp. 79–98). New York, London: Teacher College, Columbia University.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18), Retrieved April 4 2009 from <http://epaa.asu.edu/epaa/v2010n2018/>.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267.
- Baumert, J., & Watermann, R. (2000). Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Eds.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (pp. 317–372). Opladen: Leske+Budrich.
- Brozo, W. G., & Hargis, C. (2003). Using low-stakes reading assessment. *Educational Leadership*, 61(3), 60–64.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L. M., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut Hill, MA: Boston College, National Board on Educational Testing and Public Policy.
- Day, C. (1999). *Continuing professional development*. London: Falmer Press.
- Day, C., & Sachs, J. (Eds.). (2004). *International Handbook on the Continuing Professional Development of Teachers*. Maidenhead: Open University Press.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
- Hamilton, L. S., Stecher, B. M., Russell, J. L., Marsh, J. A., & Miles, J. (2008). Accountability and Teaching Practices: School-level actions and teacher responses. In B. Fuller, M. K. Henne & E. Hannum (Eds.), *Strong stakes, weak schools: The benefits and dilemmas of centralized accountability* (pp. 31–66). Bingley: Emerald Group Publishing Limited.
- Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest KFT 4–12+ R (für 4. bis 12. Klassen, Revision)*. Göttingen: Beltz-Test GmbH.
- Herman, J. L. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore (Eds.), *redesigning accountability systems for education* (pp. 141–166). New York/London: Teachers College Press.
- Hopkins, D. (Ed.). (2005). *The practice and theory of school improvement. International handbook of Educational Change*. Dordrecht: Springer.
- Jürges, H., Schneider, K., & Büchel, F. (2005). The effect of central exit examinations on student achievement: Quasi-experimental evidence from TIMSS Germany. *Journal of European Economic Association*, 3(5), 1134–1155.
- Klein, E. D., Kühn, S. M., Van Ackeren, I., & Block, R. (2009). Wie zentral sind zentrale Prüfungen? – Abschlussprüfungen am Ende der Sekundarstufe II im nationalen und internationalen Vergleich *Zeitschrift für Pädagogik*, 55(4), 596–621.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52(6), 765–773.
- Köller, O., Baumert, J., Cortina, K. S., Trautwein, U., & Watermann, R. (2004). Öffnung von Bildungswegen in der Sekundarstufe II und die Wahrung von Standards. *Zeitschrift für Pädagogik*, 50(5), 679–700.

THE INTRODUCTION OF STATE-WIDE EXIT EXAMINATIONS

- Köller, O., Baumert, J., & Schnabel, K. (1999). Wege zur Hochschulreife: Offenheit des Systems und Sicherung vergleichender Standards. *Zeitschrift für Erziehungswissenschaft*, 2(3), 385–422.
- Leutwyler, B., & Maag Merki, K. (2005). *Mittelschulerhebung 2004. Indikatoren zu Kontextmerkmalen gymnasialer Bildung. Perspektive der Schülerinnen und Schüler: Schul- und Unterrichtserfahrungen. Skalen- und Itemdokumentation*. Zürich: Forschungsbereich Schulqualität & Schulentwicklung, Pädagogisches Institut, Universität Zürich.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58, 103–117.
- Maag Merki, K. (2010). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Eds.), *Handbuch Neue Steuerung im Schulsystem* (pp. 145–170). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maag Merki, K. (in press). *Zentralabitur. Die längsschnittliche Analyse der Wirkungen der Einführung zentraler Abiturprüfungen in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maag Merki, K., & Holmeier, M. (2008). Die Implementation zentraler Abiturprüfungen. Erste Ergebnisse zu den Effekten der Einführung auf das schulische Handeln der Lehrpersonen. In E.-M. Lankes (Ed.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (pp. 233–244). Münster: Waxmann.
- Maag Merki, K., Holmeier, M., Jäger, D. J., & Oerke, B. (2010). Die Effekte der Einführung zentraler Abiturprüfungen auf die Unterrichtsgestaltung in Leistungskursen in der gymnasialen Oberstufe. *Unterrichtswissenschaft*, 34(2), 173–192.
- Maag Merki, K., Klieme, E., & Holmeier, M. (2008). Unterrichtsgestaltung unter den Bedingungen zentraler Abiturprüfungen. Eine Analyse auf Schulebene mittels Latent Class Analysen. *Zeitschrift für Pädagogik*, 54(6), 791–809.
- Neumann, M., Nagy, G., Trautwein, U., & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen. Leistungs- und Bewertungsunterschiede zwischen Hamburger und Baden-Württemberger Abiturienten und die Rolle zentraler Abiturprüfungen. *Zeitschrift für Erziehungswissenschaft*, 12(4), 691–714.
- Pedulla, J., Abrams, L. M., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: Boston College, National Board on Educational Testing and Public Policy.
- Prenzel, M., Kristen, A., Dengler, P., Ettl, R., & Beer, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik. Beiheft* 13, 108–127.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher & S. P. Klein (Eds.), *Making Sense of Test-Based Accountability in Education* (pp. 79–100). Santa Monica: Rand.
- Swanson, C. B., & Stevenson, D. L. (2002). Standards-based-reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessment. *Educational Evaluation and Policy Analysis*, 24(1), 1–27.
- Zabala, D., Minnici, A., McMurrer, J., & Briggs, L. (2008). *State High School Exit Exams: A Move toward End-of-Course Exams*. Washington: Center of Education Policy.

*Prof. Dr. Katharina Maag Merki
Institute of Education, University of Zurich*

GERRY MAC RUAIRC

THE PISA GIRLS AND TICKING THE BOXES

An Examination of Students' Perspectives on Pisa Testing

INTRODUCTION

The rationale underpinning this chapter draws on four interrelated domains of scholarship in educational research. The first relates to the well documented accounts of the disproportionate patterns of underachievement with respect to equality of outcomes among working-class populations as evidenced by data from national assessments (Mac Ruairc, 2009) and larger scale assessment such as PISA, PIRLS and TIMMS. What is questioned here is not the construct of large scale assessment itself, but rather the need to engage fully with the key issue of social class in terms the content, design, structure and process of conducting assessments of this genre, particularly when one considers the persistent and reproduced patterns of asymmetrical attainment between social class groups. In this regard the second focus of this chapter draws on the more recent empirical use of the concept of social class in educational research. This cultural turn in class theory provides a set of conceptual models that facilitates an examination of the 'doing' of class and the investigation into particular accumulations of class based outcomes that have specific relevance to the field of education. The 'big science' work of large scale assessments of educational attainment which has produced a significant amount of comparative data in recent years is also critiqued because of the tendency within this scholarship to avoid engaging with how the process of testing operates at the local micro level. By marginalising the individual experience of the students who agree to undergo this assessment, the students themselves become subordinate to the highly instrumentalist functioning of this form of testing. What has emerged is a situation which resonates strongly with some of the more extreme manifestations of the neo liberal, new managerial ideology. Finally, in establishing how individuals react to test and engage in the process of testing, it is important to take on board the perspectives of individual students themselves. Developments within the field of the sociology of childhood in particular have contributed greatly to a view of children that challenges the tradition 'apprenticeship for adulthood' perspective on childhood. What is considered here is an exploration of 'how children and young people see the world, their values and priorities and the ways in which they feel themselves marginalised' (Roche, 1999, p. 477). By examining the views of students, this case-study will provide an insight into how the PISA assessment (2009) was experienced by a group of working-class girls in a disadvantaged inner city school in a large urban area in the Republic of Ireland. It is suggested here that the socio-economic background of the

students mediates this experience of testing resulting in a number of negative consequences for the student group in the study. It highlights the need for a more proactive approach to student consolation and a more nuanced model of assessment to take account of social class difference, precisely because these differences continue to constitute the prime source of differential attainment patterns. By doing this, it may be possible for test instruments to identify attributes and strengths within working -class populations rather than the current tendency to repeatedly point out failure, thereby contributing to the already well-established trajectory of deficit scholarship with respect to this social group.

SOCIAL ECONOMIC CLASS (SEC) AND PATTERNS OF ACHIEVEMENT

There is a significant body of literature, both nationally and internationally, that points to the high level of correlation between educational attainments, longevity in the education system, and socio-economic background (Cosgrave et al., 2000; Cosgrave et al., 2004; Osborne & Leith, 2000; Sheil, 2006; Higher Education Authority 2007 and 2009). An example of these patterns of persistent underachievement can be seen in the National Reading Assessment conducted periodically in the Republic of Ireland. The 1998 and 2009 assessments of reading in schools concluded that there had been no change in the performance of low achieving pupils in English reading attainment since 1988 and that the majority of this student group were located in areas of socio-economic deprivation (Cosgrove et al., 2000). These patterns continue despite a significant degree of targeted investment in this area. The overall consensus in the national data strongly indicates that quantitative measurement of attainment levels in schools in working-class areas is showing a persistent resistance to change.

From an international perspective, Ireland has consistently performed well on PISA assessments. The figures for literacy place Ireland 5th among OECD countries and 6th among the 56 participating countries. Performance in mathematics and science is not at the same level but nonetheless remains close to the overall average. These broad statistics however mask some important patterns particularly when viewed from the perspective of socio-economic class. There are 12.2% of students performing at or below level one in literacy, a figure explained almost entirely by SES grouping (Sheil, 2006). In the cases of performance in mathematics the students in the bottom 1/3 have a mean score that is 62 points lower than the top and 32 points lower than middle 1/3 with the result that this group 'did not demonstrate even the most basic mathematical proficiencies associated with PISA mathematics' (Close, 2006, p. 72). The facts and figures identifying the SES issues are now available in abundance. Is it necessary to continue to confirm what is already confirmed? An alternative approach might be to explore, in a much more systematic way, how the structure and organisation of the school system and the modes and models of teaching, learning and assessment contribute to the persistence of this problem. It is common in aspects of some current approaches to research to 'control for' factors such as socio-economic class for the purpose of statistical analysis (Wrigley, 2008) relegating it to part of the

overall ‘noise’ or ‘outside background factors’ that needs to be stripped away in order to reveal the true impact of school factors (Angus, 1993, p. 361). To challenge this limiting, reductive view of the impact of fundamental categories such as socio-economic class, it is necessary to take into account the complexity of the school population and focus directly, rather than ‘control for’, the manner in which educationally marginalized groups participate and engage in the process and practice of schooling.

THE CULTURAL TURN IN SOCIAL CLASS THEORY

Recent scholarship in relation to social class has reframed the concept “as a dynamic mobile aspect of identity that continues to permeate daily interactions despite its marginalisation in prevailing contemporary discourses” (Reay, 1998a, p. 259). This cultural turn (Devine and Savage, 2005) provides a more enriched understanding of the complexities of class in contemporary society (Reay, 1998a; Charlesworth, 2000; Woodin, 2005) by problematising the relationship between material circumstance and cultural discursive resources. The concept of class has been extended from economic capital alone to cover human capital, social capital (Putnam, 2000) and cultural and symbolic capital (Bourdieu, 1977, 1998, 2000). The now broader focus examines how cultural and symbolic resources function to position people in a similar way to economic resources. It allows for an examination of the real lives and everyday activities of individuals as they live and make sense of the material, social and cultural worlds. Identity reaches the individual’s psyche and therein one can see the effect of class where “denial, dis-identification, defensiveness, pride and shame are familiar and often competing responses to living class on a day to day basis” (Skeggs, 1997, p. 75). Social differences are fundamentally connected to the way an individual is brought into being (Hey, 2005). They become part of the “embodied habitus” of an individual (Bourdieu & Wacquant, 1992, p. 86). This essentially Bourdieuan perspective on class has a notable resonance in the field of education. Recent research in relation to the workings of socio economic class in education has identified a number of patterns of practice with respect to the mobilisation and conversion of forms of capital which produce clearly identifiable benefits for the middle classes (Reay, 2000; Laureau, 2003). The politics of school choice in the Republic of Ireland during the recent economic boom, mirrors patterns identified internationally and has led to a significant growth in the private school sector at second level in particular (Lynch & Moran, 2006; Mac Ruairc, 2010). This in turn has produced a clearly delineated classification of chosen and unchosen schools (Mathews, 2010) especially in urban areas. An examination of the patterns of transfer to third level education (HEA 2009) between different social groups in the Republic of Ireland confirms the socio-economic basis of the inequality of educational outcomes and the socio-economic apartheid currently mediated by the education system.

In an era where the knowledge society seems to be the only trajectory for economic development, it is now essential to re-engage with the more radical

theorists in education who have provided a range of conceptual tools and frameworks for the critical analysis of education (Apple, 1996, 2001, 2002; Ball, 2006, 2008; Devine, 2000, 2003; Giroux, 1983; Giroux & McLaren, 1994; Lynch & O’Riordan, 1996; Lynch & Lodge, 2002). Against the conservative claim that schools transmit objective knowledge, we now need to refocus on areas of hidden curricula as well as ideologies that identify specific interests that underlie different knowledge forms and the measurement of the acquisition of these forms of knowledge. There is a need to examine how schools and education systems individualise failure and legitimise inequalities within a structure where failure is attributed to inborn facilities or where “cultural deficits relayed by the family... come to have the force of inborn facilities” (Bernstein, 1996, p. 13). Such an approach makes central the need to analyse how human experiences are produced, contested and legitimated within the dynamics of everyday classroom life (Giroux & McLaren, 1989). It is concerned with the ways in which “oppression is structured and legitimated in the taken-for-granted norms, habits and rules of institutions” (Barton, 1996, p. 10). Testing and the measuring of the class differences in relation to educational outcomes are part of this overall problematic.

WHERE THE SHOE PINCHES¹ EXPLORING STUDENT PERSPECTIVES

The analysis that has occurred within the sociology of childhood has been a significant development in the examination of the role of children in society. This perspective has highlighted how children have been marginalized within systems because of their subordinated position in society and the theoretical conceptualisations of childhood and socialisation (Qvortrup et al., 1994). The lack of consideration of children as political figures has resulted in the subordination of their opinions and interests to adult dominance (Hendrik, 2000; Pinkerton, 2004). Traditionally, childhood was viewed as a transitional phase which is complete when children enter adulthood (Holloway & Valentine, 2000). Typically the reservations of adults about children’s ability to take part in consultative processes revolve around perceptions of their lack of competency (Devine 2003), their perceived disinterest or in order to shield them from the trials of the adult world (Matthews et al., 2000). In terms of research in education, much has changed in this area and there is an increasing body of work that address the significance of schools as key sites for the construction and experience of childhood, as well as the impact of schooling on the construction of children’s subjectivities in line with dominant norms (Devine, 2003; Kampmann 2005). Schools and children’s experience of everyday life within them are deeply embedded in the power structures in society. Accessing children’s voices on this experience gives unique insights not only into the experiences of children as a distinct social group but also their active participation in the processes of production and reproduction in society at large.

STUDY DESIGN

This case-study was conducted in a disadvantaged girls' secondary school in a traditional working-class inner city area in the Republic of Ireland. This site was chosen because of the author's professional contact with the school while working on a range of projects over the past number of years. The study comprised a visit to the school on the day following the administration of the 2009 PISA test. It was decided not to visit the school during the testing process in order to avoid the added upset and distraction of having an observer in the testing centre in the school. Instead the test coordinator and the school principal agreed to monitor the students' reactions and progress on the test during the testing process. Individual interviews were conducted with both during the school visit. The students' participation in the research comprised focus group interviews with three groups of students.

Group one: 6 in Junior Certificate Exam year (Grade 3)
Group two: 4 post Junior Cert (Grade 4)
Group three: 5 post Junior Cert (Grade 4)

Participation in the group discussion was voluntary. Of the 25 students who took the PISA 2009 Assessment 15 agreed to take part in the study. The focus group discussions explored the students' perspectives on the test content, test structure and test administration. For the purposes of the discussions in relation to the test items in literacy, mathematics and science, students drew on their recollections of test items from previous day. In the case of the Student Questionnaire which contained 52 questions based on student were given a copy of the test booklet and were asked to re-examine questionnaire as part of the discussion. The interviews were transcribed verbatim and the transcripts were not edited for grammatical omissions or errors. In the interest of clarity explanatory comments are inserted in square brackets in order to clarify colloquial and other local terminology

FINDINGS

This review explores the reality of administering the test to a group of working-class students in an urban inner city school. Three clearly identifiable themes emerge from the thematic analysis of the interview and focus group transcripts. These are reported below under the following headings:

- The volume and intensity of the testing process
- Ticking the boxes: The variety of strategies used by the children in completing the test items with implications for the validity of some of the responses to test items.
- Too many personal questions: The personal and consequently problematic nature of many of the questions on the Student Questionnaire.

The volume of content and intensity of the testing process

This was considered to be a highly problematic aspect of the testing process. All participants in the research concur with the fact that the test was too long, there were not enough breaks and not enough support was available for students with special educational needs. These supports would be available for other forms of testing in the school including state examinations. The lack of an assigned reader for children with reading difficulties was noted as the most significant oversight in this regard.

It is very intense. Too intense for the children. Too intense for me. I had barely time to get a glass of water. It doesn't need to be like that. It does affect how they do it – they don't work well under that kind of pressure. It was just too long. They lost concentration. Some of them just gave up. You could see them wilting and slipping off (**PISA School Co-ordinator**).

There was a two hour questionnaire and a break after it of about five minutes. For people with special needs that's not on. We have a girl with a physical disability, in a wheelchair and we hardly had time to get her out of the room. And there was nowhere that it said that you could have somebody in to support them, to turn the page or readers if they needed them like our own state exams. (**Principal**)

Looking at how intense it was, you don't have to go to school to figure out that this is wrong. It looks wrong. I know we were doing well in PISA at one point but clearly we'll be doing very bad on this one. What monster is this feeding? I would have to question it. It would have been great to video it and to watch the body language and they pace that some of them were filling it up at. I would have to question how the children approach it – it's so intense. I just thought it was pretty harsh – for us here. They're 15 years old, you can't go in and just do this to them – like they are some sort of a rat in a lab (**Principal**).

The comments from the students concur with this in relation to the volume of testing and they also identified problems with specific question types that were considered by the students to be unfamiliar in terms of format and focus.

Amount of reading. There was just too much reading.

Content of reading. Some of them were really old fashioned. About a boy who wanted to kill lions and I didn't want to read it like, I just wasn't interested.

Difficulty with continuous text items. Some of them were very long so you don't read it. You go 'ah here' and then look at the question and go back for the answers. If they're too long yeh [you] don't read it.

Problems with the metacognition questions. I just ticked anyone, I didn't know what they were or what they were looking for so I just ticked anyone. I didn't get them. I didn't know what they were about.

These really confused me. I didn't know what they were looking for there.

I read some of them and tried to think logically what they were looking for. And sometimes I just ticked a few boxes if I couldn't get it

The comments also reveal that there are deeper more fundamental consequences for the group particularly relating to the negative impact of the testing experience on the students' sense of themselves in the school context. It is clear from the comments included below that there was a degree of forbearance needed among the students to get through the test. There was also a sense that the students were victims in this process for a number of reasons. Because individual students do not receive the results of the tests, the process is neither a formative nor summative form of assessment so there is little potential for learning in this exercise. The perspectives articulated by the students indicate that the format and model of testing are viewed very negatively by the group. The principal commented on this by stating: 'They were only doing it for us, because Miss Madison was doing it with them. Otherwise we would have a hope getting them to do it.' The school therefore is implicated in the process and to an arguable degree the negative feelings become conflated with negative feeling towards school which are often present in abundance without resort to PISA testing. For this and many other schools working with students on the margins every effort is made to ensure that school experience is as positive as possible; 'we try to coax them here – that what these need coaxing' (Principal). In the case of this school, the PISA assessment did not fit into this overall school mission.

The time was too long and the break was too short. Not enough breaks and not long enough breaks. There were 60 odd pages in it.

They need to cut it down a bit and not have so many pages. When we were like told about it I thought it was just going to be a few pages – a few tick boxes and that would be it. But it's like a book, like that thick.

I thought it would be more like the Junior Cert. [Junior Certificate – a state examination held after three years in second level school and completed by the grade 4 students in the sample the previous June]. The book was about 12 pages I was shocked when I saw it. The first thing I did was look at the last page when I got it and thought 'Oh my God'. It was like all of the exams in the Junior Cert put together.

I was into it [doing the test] at the start but it was too much – it was hard. For the first while I done alright but then it was too hard and too long and I got sick of just sitting there.

G.M. RUAIRC

I was trying to do me best at the start but by the time I got to page 40 odd I just thought this is never going to end. Then I kinda stopped, and didn't think as much about the answers or just picked one, it depended like.

I got about $\frac{3}{4}$ of the way through it and then looked and thought there's loads left, and then I started to give up and at the end I just ticked.

The teachers were standing there waiting for me to finish and making a show of me. It was too long and all the reading and thinking about the questions and all, it was just too much. So I just looked at what was left and ticked as fast as I could. There were a few of us who did the same.

Ticking the boxes

One of the consequences of the difficulties experienced by the students in competing the test relates to the quality of answers generally which in some cases were selected on the basis of a straightforward guess.

I just ticked the boxes going along I didn't read them all.

The reading one was the worst. I just got fed up and went here and just went tick, tick, tick.

This strategy had specific consequences for the difference in the quality/reliability of the answers at the beginning of the test and those answers given at the end of the test. Whatever the chance was of getting an answer that was the result of a full consideration by the student in the early part of the test, for many students this would not have been possible at the end. This data also provides additional support for the negative emotional impact of the test experience on the students.

You concentrate at the start and then you get sick of it. I flicked through to see what was left and then gave up.

I stuck with it for a while and then when I looked through the pages at what was left. I just thought 'oh crap' and started to tick everything.

I was into at the start but it was too much – it was hard. For the first while I done alright but then it was too hard and too long and I got sick of just sitting there.

I was trying to do me best at the start but by the time I got to page 40 odd I just thought this is never going to end. Then I kinda stopped, and didn't think as much about the answers or just picked one, it depended like.

Too Many Personal Questions: Student Questionnaire:

Following completion of the PISA assessment test booklet on Literacy, Mathematics and Science, students were asked to complete a 52 page Student Questionnaire with a total of 79 items. The purpose of this document was to

collect data in relation to a number of background variables including family and home circumstances, children's own reading activities, and classroom / school climate among others (c.f. DES, 2009 for list of questions). While the volume of content of this document was also problematic given the fact that students had already completed a lengthy test booklet, this was outweighed by the concern expressed by the students in relation to the lack of anonymity and the personal and sometimes upsetting nature of some of the questions in this booklet. This concern is attributable to two main causes, firstly the requirement for students to write their names on the booklet and secondly the personal nature of the questions. In relation to the former, assurances were included in the test booklet in relation to the confidentiality of the answers:

Your answers will be combined with others to make totals and averages in which no individual can be identified. All your answers will be kept confidential (Department of Education and Science, 2009)

The PISA coordinator also assured the children that the questionnaires were confidential.

The kids said to me at the end, "This is supposed to be confidential but why are our names on it". They asked me would I be looking at them. I said I wouldn't and they saw me sealing the package to return the answered tests – but still they were not happy at all with their names on it. (PISA School Coordinator).

These assurances were not adequate. It is clear from the comments below that the children had reflected on this matter. Some of the questions were more problematic than others in this regard. Many of the students felt vulnerable when responding to items concerning attitudes and perspective in relation to school life, classroom climate, the quality of teaching and their individual home circumstances.

They had yer [your] name typed onto the yoke. I didn't like answering them.

You know what I thought? We shouldn't have had to give our names and then I would have no problem answering them questions. I would have had no problem answering all them nosey questions because they wouldn't have known who it was.

There not giving back our results so why do they need our names? We would have answered freely if they weren't on it.

I liked the bit where you could criticise your teacher but then yer [your] name was on it so I wanted to put stuff down but didn't.

I was ticking them all good because my name was on it.

No, I answered them honestly. They were checking how you learn in class

I was afraid to disagree in case the teacher saw it.

The impact of the personal nature of many of the questions, categorised below under two broad headings, provides the greatest imperative for dealing with this matter in a different way in subsequent testing situations..

Just too Personal: All of the students in the sample felt that the question was much too personal. Many were upset having to answer them and felt that they were in some way betraying their parents' confidentiality by answering questions on their behalf:

They were very personal, I thought it was just about us but it was about our Mas and Das [mothers and fathers] and all that. It was a bit personal, a bit nosey like. Does your Ma have a PhD? And yer [your] name was on it and we were told it would be confidential.

I was bothered by the question about who lives at home with you. It made me feel uncomfortable. It's our business.

I thought a lot of the questions were very personal. They were questions that me mam and dad should have been asked not me. I kind of felt awkward answering questions about them [i.e. About her mother and father].

They [parents] should have to say if they wanted to tell what they do. It's their business to tell that if they want - not my business to tell it for them.

I felt bad writing down me ma's business. It's hers not mine.

Class Consciousness: The students felt that a judgement about their background and family circumstances was implicit in many of the questions. The impact of this for some was a negative sense of themselves and a notion of class inferiority arising from the fact that many of the group clearly understood that the presence of the many of the school supports mentioned in the questionnaire was an indicator of good home support for school, interested/ 'proper' parenting and the right kind of lifestyle. The absence of these, which would have been the case for most of the children, represented an inferior way of life with a consequent negative impact on the outcome of education.

More nosey questions on what you talk to your parents about, we don't sit around and talk much and we don't talk about politics, maybe the soaps sometimes. There just checking to see if yer well informed and if your ma and da help you to be well informed. And if they do - they're good parents. They're saying if your parents don't talk to you that you are going turn out bad.

If you have all these things then you're smarter. Like the people living in the richer houses are smarter and better and you get a better chance and the ones that don't well they don't have a much of a chance. If you have an ensuite bathroom then you probably can afford private education. But that's not always true.

I think that they are saying that if you have all them DVDs and expensive stuff and not books and the schools things that your Ma doesn't really care

about school or what yeh [you] do in school. There spending money on things that don't help you at school, but that's what they want to do.

I didn't really want to write down what it was [mother's occupation]. I did write it down in the end like but I felt that I should lie on this one to make my mam's job sound better. She's a secretary but I wanted to write like a manager or something.

The lack of resonance between the lives of the students in this sample and some of the questions in the questionnaire is well articulated in the quotes below. It is here that the 'us and them' rationale underpinning the thrust of many of the questions on the questionnaire is captured most succinctly.

And the lists of jobs are all smart jobs. Like my ma stays at home and minds kids and like there wasn't space for that.

The jobs down there are ones that you'd have to go to college to do. Not like a builder or a painter like normal jobs. They weren't jobs like we have here. Even the different types [kitchen hand] are not normal jobs.

CONCLUSION

The issue of socio-economic class which is the key construct underpinning the issues raised in this chapter is complex, resilient and persistent. Exploring the reproduction of patterns of failure and alienation in schools requires the critical deconstruction of the fabric of school life. This is a difficult task and one which has the potential to disrupt the distribution of well established patterns of power, prestige and outcomes within the field of education. Consequently it is a focus of enquiry which is often marginalised in the discourse framing educational debate (Apple, 1996; Ball, 2006, 2008; Giroux & MacLaren, 1989). The findings of this single case study school and the small number of students who participated in the study pale into insignificance when considered in the light of the scope of large scale assessments such as PISA. It could be argued that the comparative data that emerges from these assessments is bound to have some kind of negative impact that cannot be controlled for, such is the scale of the study. However, this is not an adequate response when one considers the negative impact of aspects of the test on students. It is very likely that the findings reported here are part of the unintended consequences of the PISA assessment process. What follows is the imperative to locate who is responsible for these consequences unintended and all as they may be. It is asserted here that those working at a national and international level on the development of assessment tests need to examine the workings of the test at the individual student level and makes amendments where necessary. What would happen if disadvantaged schools refused to subject their students to this regime of testing? What if school leaders act as gate keepers of initiatives that do not adequately take account of the social make-up of the student group? This tendency for schools in lower socio economic catchment areas not to participate in these types of assessments is already identified as being an issue particularly in countries

G.M. RUAIRC

with lower participation levels (Prais, 2003). If we must continue to measure and engage in surveillance of this kind then it would be better to ensure that this is carried out in the most empowering way possible. The findings here clearly indicate that the lack of consideration for students and particularly for those in lower socio economic groups at all stages in the assessment process is avoidable. Many of the problems identified by the students can easily be rectified. The broader question of the repeated measurement of class inequality requires a much more powerful paradigm shift but is vital if the current patterns of inequality are to ever change.

It is, I suggest, appropriate to conclude with two quotes because that say almost all that needs to be said at this point in relation to this matter by capturing the essence of what is happening for some students who participate in the PISA assessment as it is currently structured:

They're 15 years old, you can't go in and just do this to them – like they are some sort of a rat in a lab (**Principal**)

They expect you to sit down for two and half hours and fill in all this and then not ask you how you feel about it! (**Grade 5 student**)

ACKNOWLEDGEMENTS

I would like to express my gratitude to Miguel A. Pereyra for the opportunity to share this story at the conference in Santa Cruz La Palma and to thank the many contributors to the conference for their support, encouragement and empathy in relation to the views expressed by the PISA girls. I would like to thank in particular Clara Morgan, Vlatka Domovic and Robert Cowen.

NOTE

- ¹ Based on the German proverb: 'No one knows better where the shoe pinches than s/he who wears it'

REFERENCES

- Angus, A. (1993). The sociology of school effectiveness. *British Journal of Sociology of Education*, 14(3); 353–367.
- Apple, M.W. (1996). *Culture, Politics and Education*. Buckingham: Open University Press.
- Apple, M.W. (2001). 'Will standards save public education' *Educational Policy*, 15(5), 724–729.
- Apple, M.W. (2002). 'Does education have independent power? Bernstein and the question of relative autonomy'. *British Journal of Sociology of Education*, 23(4), 607–616.
- Ball, S. J. (2008). *The education debate*. Bristol: Policy Press.
- Ball, S. J. (2006). *Education policy and social class*. London: Routledge
- Barton, L. (1996). Sociology and disability: some emerging insights. In L. Barton (Ed.), *Disability and Society: Emerging Issues and Insights* (pp. 3–17). London & New York: Longman.
- Bernstein, B. (1996). *Pedagogy, symbolic control and identity: Theory, research and critique*. London: Taylor Francis.
- Bourdieu, P. (1977). *Outline of theory and practice*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1998). *Practical reason: On the theory of action*. Cambridge: Polity Press.

- Bourdieu, P. (2000). *Pascal meditations*. Cambridge: Polity Press
- Bourdieu, P. & Wacquant, L. (1992). *An Invitation to reflexive sociology*. Cambridge: Polity Press.
- Charlesworth, S. J. (2000). *A phenomenology of working class experience*. Cambridge: Cambridge University Press.
- Close, S. (2006). The junior cycle curriculum and the PISA mathematics framework. *Irish Journal of Education*, 37, 53–78.
- Cosgrove, J., Kellaghan, T., Forde, P. & Morgan, M. (2000). *The 1998 national assessment of English reading*. Dublin: Educational Research Centre.
- Cosgrove, J., Shiel, G., Sofroniou, N., Saztruski, S. & Shortt, F. (2004). *Education for life: the achievements of 15 year olds in Ireland in the second cycle of PISA*. Dublin: Educational Research Center.
- Department of Education and Science (2005). *Literacy and numeracy in disadvantage schools: challenges for teachers and learners*. Dublin: Government Publications Office.
- Devine, D. (2000). Constructions of childhood in school: power, policy and practice in Irish Education. *International Studies in Sociology of Education*, 10 (1), 23–41.
- Devine, D. (2003). *Children, power and schooling: how childhood is structured in the primary school*. London: Trentham Books.
- Devine, F. & Savage, M. (2005). The cultural turn, sociology and class analysis in Devine et al., *Rethinking class: culture, identities and lifestyles* (pp.1– 23), London: Palgrave Macmillan.
- Devine, F., Savage, M., Scott, J. & Crompton, R. (2005). *Rethinking class: culture, identities and lifestyles*. London: Palgrave Macmillan.
- Giroux, H. A. (1983). *Theory and resistance in education: A pedagogy for the opposition*. London: Heinemann.
- Giroux, H.A. (2001). Cultural studies as performative politics. *Cultural Studies/Critical Methodologies*, 1(1), 5–23.
- Giroux, H. A. & McLaren, P. (1994). *Between borders*, New York: Routledge.
- Giroux, H.A. & McLaren, P. (1989). *Critical pedagogy, the state and cultural struggle*. New York: State University of New York Press.
- Hendrick, H. (2000). The child as a social actor in historical sources: problems of identification and interpretation. In L. Christensen et al., (Eds.), *Research with children: perspective and practices* (pp. 36–61). London: Falmer Press.
- Hey, V. (2005). The contrasting social logics of sociality and survival: cultures of classed be/longing in late modernity. *Sociology*, 39 (5), 855–872.
- Higher Education Authority (2009). *Higher education: Key facts and figures*. Dublin: HEA.
- Higher Education Authority (2007). *National office for equity in higher education: Annual report 2006, outline plans 2007*. Dublin: HEA.
- Holloway, S.L. & Valentine, G. (Eds.) (2000). *Children's geographies. playing, living, learning*. London and New York: Routledge.
- Kampmann, J. (2005). *Understanding and theorizing modern childhood in Denmark: Tendencies and challenges*. Copenhagen: Copenhagen Business School Press.
- Kellaghan, T., Weir, S., Óhualacháin, S., & Morgan, M. (1995). *Educational disadvantage in Ireland*. Dublin: Department of Education/Combat Poverty Agency/Educational Research Centre.
- Lareau, A. (2003). *Unequal childhoods*. Berkeley LA: University of California Press.
- Lynch, K. & A. Lodge (2002). *Equality and power in schools*. London: Routledge Falmer.
- Lynch, K. & O’Riordan, C. (1996). *Social class, inequality and higher education: Barriers to equality of access and participation among school leavers*. Dublin: Equality Studies Centre, UCD.
- Lynch, K. & Moran, M. (2006). Markets, schools and the convertibility of economic capital, the complex dynamics of class choice. *British Journal of Sociology of Education*, 21, 221–235.
- Mathews, D. (2010). *School self-evaluation and school improvement*, unpublished doctoral thesis. Dublin: National University of Ireland, Maynooth,
- Mac Ruairc, G. (2004). Schools, social class and children’s perception of language variation. In *Primary Voices*, Deegan, Devine & Lodge (Eds.), Dublin: IPA.

G.M. RUAIRC

- Mac Ruairc, G. (2009). Dip, Dip, Sky Blue - Children's experience of standardised testing: A Socio-cultural perspective. *Irish Educational Studies*, 28(1), 47–66.
- Mac Ruairc, G. (2010). This way please, the OECD and school leadership improvement. *Journal of Educational Administration and History*, 42(3), 223–246.
- McLaren, P. & Farahmandpur, R. (2001). Educational policy and the socialist imagination: revolutionary citizenship as a pedagogy of resistance. *Educational Policy*, 15(3), 343–378.
- McLaren, P. (2001). Che Guevara, Paulo Freire and the politics of hope: reclaiming critical pedagogy. *Cultural Studies/Critical Methodologies*, 1(1), 108–131.
- Osborne, R. & Leith, H. (2000). *Evaluation of the targeted initiatives on widening access for young people from socio-economically disadvantaged backgrounds*. Dublin: HEA.
- Pinkerton, J. (2004). Children's participation in the policy process: some thought on policy evaluation based on the Irish National Children's Strategy. *Children and Society*, 18, 119–130.
- Prais, S. J. (2003). Cautions on OECD's recent educational survey. *Oxford Review of Education*, 29 (2), 139–163.
- Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
- Qvortrup, J., Bardy, M., Sgritta, G. & Wintersberger, H. (Eds.) (1994) *Childhood Matters: Social Theory, Practice and Politics*. Aldershot: Avebury.
- Reay, D. (1998). Rethinking social class: qualitative perspectives on class and gender. *Sociology*, 32 (2), 259–277
- Reay, D. (2000). A useful extension of Bourdieu's conceptual framework: emotional capital as a way of understanding mother's involvement in their children's education. *The Sociological Review*, 48(4), 568–585.
- Roche, J. (1999). Children: rights, participation and citizenship. *Childhood*, 6(4), 475–493.
- Shiel, G. (2006). The PISA assessment of reading literacy. *Irish Journal of Education*, 37, 79–100.
- Skeggs B. (1997) *Formations of Class and Gender*. London, Sage.
- Woodin, T. (2005). Muddying the waters: changes in class and identity in a working class cultural organisation. *Sociology*, 39(5), 1001–1018.
- Wrigley, T. (2008). School improvement in a neo-liberal world. *Journal of Educational Administration and History*, 40(2), 129–148.

Gerry Mac Ruairc
School of Education
University College Dublin
Ireland

MARIE DURU-BELLAT

FROM THE APPEALING POWER OF PISA DATA TO THE DELUSIONS OF BENCHMARKING

Does that Challenge Any Evaluation of Educational Systems?

INTRODUCTION

The aim of this paper is not to add any additional criticism regarding PISA data to the already existing mass of technical comments by journalists, policy-makers and, often in a very relevant manner, by researchers (e.g. see Goldstein, 2004, and the contributors of this book). Rather, we will discuss the important issue of what role PISA data can and cannot play in the more global objective of assessing the quality of education systems. In doing so, we will suggest that some of the criticism and resistance aimed at PISA data actually have to do with their misuse in the narrow process of benchmarking, as well as, more widely and less overtly, the evaluation of education systems, which is more debatable.

This perspective does matter because the development and success of PISA is nested in a global context, with a strong stress on education, as well as correlated public funding and concern for accountability. In the world's wealthiest countries, the dual role – both economic and social – of education and training is constantly reaffirmed. That is especially true for Europe, where in 2000, the Lisbon strategy was put forward, aimed at making Europe “the most competitive and dynamic knowledge-based economy in the world.” It was followed by the concrete “Lisbon goals”, listing a set of objectives with precise quantitative benchmarks (European Commission, 2007), such as increasing participation in pre-school education, lowering the number of drop-outs, reducing the percentage of low-achieving 15-year-olds in reading literacy, increasing access to tertiary education, developing the participation of adults in life-long learning, and improving the entire population's level of educational attainment across the board. EU country figures are published in regular reports; among this wide variety of figures, PISA data have reached a specific audience and are given an important place in “Education at a Glance” and various OECD publications.

WHY ARE PISA DATA SO APPEALING DESPITE THEIR LIMITATIONS?

The main reason for the international success of PISA is straightforward: if education is considered to be an important good, both in itself and as a strategic tool in international competition, PISA data provide what is considered to be an objective

assessment of what education systems “produce.” For assessing the amount of “human capital” available in a country, one can observe what percentage of individuals has attained a given level of education. For instance, the OECD (OECD, 2009) regularly publishes figures concerning the percentage of adults having achieved an upper-secondary or tertiary education, as well as enrolment and graduation rates by age. However, especially when one is interested in cross-country comparisons, limiting oneself to a generation’s rates of attaining a variety of degrees is not enough, since there is no universally valid correspondence between degrees and the knowledge and skills they certify. Therefore, assessing precisely what students have learned on a comparable basis is a much more valuable tool. Moreover, this is also a legitimate concern for governments and policy-makers, since important public funding is allocated to teaching. Would it not be shocking if policy-makers did not manifest any interest concerning educational output?

In that way, PISA data are truly appealing since they give a concrete picture of 15-year-old students’ performance in subjects or exercises that are supposedly relevant for daily life. The fact that what is assessed is supposed to be “life skills” and not pure conformity to academic knowledge is especially attractive, since the ability to cope with life is the final objective of schooling, rather than mastering pure academic contents. No matter what legitimate criticisms may be raised regarding PISA data, no one would argue that it is not valuable to gather information about what young people are able to perform, at what level of mastery and with what disparity. For example, even crude information concerning the percentage of 15-year-olds facing important difficulties in understanding texts they encounter in daily life is quite valuable in itself.

Similarly, whatever the limitations of the PISA data, simply showing that both mean performance and disparities among students vary across countries with similar levels of development and public funding is quite interesting. In addition, it is interesting to note that the dispersion of performance itself varies largely across European countries: in some countries, such as Spain or Finland, student performance is much more homogenous than in others, such as Belgium or Germany. While this may appear trivial, it means that failing (or not) at school, as measured at age 15, is not a fatality but rather relates to clearly social (and therefore politically relevant) factors. This is also true in terms of social inequalities: while income inequalities are similar in some Northern and Eastern European countries, the social inequalities in student performance assessed by recent PISA data are much larger in the latest ones (Duru-Bellat & Bydanova, 2009). This suggests that some systems manage to compensate better than others for the inequality that exists in the society in which they are embedded, or at least that some social conditions, including some educational system characteristics, do matter.

PISA data has brought to light other stimulating even, again, if they are far from perfect. One such example is the empirical test they provide concerning some current ideas prevailing in the education world, like the possible (or even probable) trade-off between excellence of education and its distribution, or what economists classically call a trade-off between “efficiency and equity”: since the pursuit of equity requires focusing on the weakest students, the best students suffer and the

overall achievement level is lower. Actually, PISA data demonstrate that this is not becomes cut and dry. They show that countries with the highest share of very bright students are often also those in which the gap between the weakest and the brightest is narrowest. Such is the case, for example, in Scandinavian countries, while very few countries have both an above-average mean student performance and a large between-student variation (e.g. Austria). In a similar manner, research based on PISA data shows that countries with the largest social inequalities also tend to have a larger proportion of weak students (e.g. see Duru-Bellat & Suchaut, 2005; Haahr, 2005). Conversely, countries with a high degree of social equality have, on average, larger proportions of high performing students. So, all in all, there should not be any competition between policies aimed at improving average efficiency and those aimed at reducing social inequality.

Actually, concerning equity, some structural characteristics of the system such as its degree of openness and differentiation may matter more. For instance, country comparisons based on PISA data show that early tracking is generally associated with both increased inequality and a lower mean level of achievement (Hanushek & Wössmann, 2006). Any curricular differentiation tends to increase social inequality because privileged students are more likely to receive the type of education that contributes to higher performance (Perry, 2008). Generally speaking, research also shows that in those countries with low disparities between students, little variation in student performance is seen between schools, meaning that a certain degree of quality (defined here as the performance achieved) is guaranteed to every student regardless of his or her school. The magnitude of academic dispersion is also linked to the amount of social inequality. In countries where schools do not succeed in ensuring equal performance by every student, the unequal social advantages they possess has a stronger impact on their performance, resulting in more social inequality in educational outcomes. In other words, homogeneity in educational quality guarantees educational equality. PISA data also suggests (with caution, since school composition is not accurately assessed in the national samples used) that student intake matters, and results in better mean performance when a school's socio-economic composition is more favorable. As Perry says (2008, p. 83), "...in summary, results from the PISA survey suggest that educational inequality can best be tackled by making schools more similar to each other in terms of curriculum, resources, and students".

Without presenting an exhaustive list of all the pieces of information found in PISA data, one can maintain that even if they are imperfect and questionable, they remain helpful in highlighting differences in educational outcomes across countries. By doing so, they enable debates that would be otherwise limited to only one country, within one institutional framework, which is often the normal and seemingly obvious way to think about educational issues, at the risk of some fatalism. Would the theory of social reproduction have been as deterministic if its authors, Bourdieu and Passeron had heard about countries without important social inequalities at school?

M. DURU-BELLAT

MISUSES AND LIMITATIONS OF PISA, OR MISUSED AND LIMITATIONS OF BENCHMARKING?

Here we will focus on what appears to be the main downsides of PISA data, limiting ourselves to the way in which they are used in cross-country comparisons rather than dealing with their technical limitations resulting from the choice of specific items, imperfections in the student sample, problems linked with translation and possible cultural differences in the way questions are understood by students, etc.

Some pitfalls of the ranking obsession...

From the outset, student assessment has been linked to international comparisons. This is understandable since assessment cannot exist without comparisons. The first international studies (e.g. the First International Mathematics Study launched in 1964 and, more broadly, the first International Association for Evaluation of Educational Achievement, IEA, studies) clearly intended to compare countries. From the 1980s onward, the concern for accountability and standards became more and more compelling, along with the development of statistical indicators in every domain of economic and social life. In 2000, the European Commission launched its 16 quality indicators, intended to be the milestones of the Open Method of Coordination. They were to allow regular benchmarking, resulting in the identification of countries performing well in particular areas so that their expertise and good practices could be shared with others, due to some peer pressure. Since comparisons are the ultimate motivators for all these surveys and indicators, it is all the more necessary to scrutinize the way in which these comparisons are made.

The first point to be stressed is that in comparative education as in other social sciences, results are contingent on the sample used. That is obvious as far as ranking is concerned, but the degree of variety in the sample matters as well. This means that one cannot estimate the influence of a characteristic universally spread throughout a country sample. Moreover, if in one analysis we combine factors that vary little across countries with others that vary sharply, the weight of the second will be more obvious than that of the first. If we include both rich and very poor countries in the sample of countries studied, country wealth may show up as influential with regard to high scores, which would not be the case with a more homogeneous sample of countries. Likewise, and unsurprisingly, contextual factors associated significantly with student performance can vary for rich and poor countries, since many factors, such as the teachers' level of training and the number of books per student, only come into play below a minimum threshold.

A second point that strongly limits the stake of the comparison is the fact that (only) approximately 10% of the total variance in PISA student performance is explained by membership to a given country (inter-country variance) (Haahr, 2005). Only this 10% can then be related to differences between countries and are in turn explainable by national contextual characteristics and schooling organization factors. The remaining 90% of the variance concerns individual countries themselves (intra-country variance). This means that structural factors varying from one

country to the next are of little importance in terms of the skills mastered by students at age 15 compared to family inequalities, which impact much more on inequalities between students. It may also be that PISA countries share all the characteristics that affect efficiency (see the previous point). So, the 10% figure cannot be considered valid at the world scale, and it is likely that if countries varying to an even greater degree in terms of educational systems, wealth, or overall level of education were taken into account, structural factors would increase in importance. Indeed, certain studies show that schooling organization factors weigh more on student performance in poor countries than in rich ones (Heyneman & Loxley, 1983). This leads us to underscore once again two fundamental points of methodology for international comparisons: 1) the only thing we ever evaluate is what varies; 2) the relationships identified depend on what country sample is being used (i.e. sample variety).

Another important point regarding PISA data often neglected by the media (and often even by policy-makers) concerns the unavoidable lack of precision of figures computed using a sample. If one takes into account the statistical array of variation linked to the sample observed, very limited differences exist between most similarly developed countries. However, every ranking tends to magnify small inequalities.

What is measured?

Another and perhaps more important problem when comparing countries concerns the nature of what is assessed in the countries' scores. Like any test, PISA data necessarily focus on certain skills. This may be considered to be an asset, since it requires operationalizing the objectives of the education system, which are often expressed only in a very vague form. However, while these global objectives may be consensual, this is much less often the case for so-called "life skills," especially when they are defined at an international level, since these skills may differ from one country to another.

More broadly, comparisons rely on a selection of indicators, which amounts to giving priority to certain objectives (most often, the easiest to assess); prioritizing students' mean academic scores in the ranking may lead to neglecting students' attitudes. However, no countries consider the mastery of certain skills to be the only and ultimate aim of education. Actually, PISA data also include some measurements of students' opinions about school life and some attitudes considered to be by-products of schooling (such as the pleasure of working with others). In these respects, countries prove to be very different; actually, there is no correlation between academic scores and the degree of student satisfaction regarding school or "positive" attitudes towards others (Duru-Bellat et al., 2008). Consequently, there exists some room for choices; for instance, some countries may "prefer" to have students who are satisfied with their school life and have learned to cooperate with their peers, but have less brilliant academic performance.

Similarly, the obsession to rank students based on mean scores may lead to neglecting the dispersion of those scores. However, the question may be raised as

M. DURU-BELLAT

to whether the ranking should be based on the academic level of average students, the top 10%, or the weakest students. In some countries, it may be judged preferable to have both a moderate and homogeneous mean level of performance, while other countries may prefer the reverse situation with a high mean level and higher disparities.

Moreover, for a variety of reasons, one cannot limit oneself to measurements taken at age 15. First, in some countries, some selection may occur earlier, the result being that both the mean level and inequality among the “survivors” are misleading and not directly comparable to other countries. In addition, across countries, students may have unequal opportunities to advance in their studies, meaning that their academic level at age 15 may be more or less important for their future. In a broader perspective, one may raise a question regarding what economists call internal versus external efficiency; that is, when assessing and comparing educational systems, should we prioritize those students who perform best or those who are best prepared for entering the job market and experiencing adult life? We will revisit this issue later.

These important decisions, like most, should involve extensive discussion since none of them are a priori made via a consensus. The core problem with benchmarking is that benchmarks are set using the most readily available data. PISA data are all the more open to criticism because they use a specific definition of what is considered to be valuable in an education system. And this is generally done as if it was obvious, while actually no indicators, even some traditional economic indicators such as GDP, are unquestionable... One may maintain that a list of indicators and the correlated benchmarks do not amount to an evaluation of the system from which political lessons may be drawn. However, we should not to throw the baby (i.e. the evaluation) out with the bath water (i.e. the benchmarks).

EVALUATION: A PROCESS MUCH MORE COMPLEX THAN BENCHMARKING

Benchmarking is a means to an end; it is meant to suggest some direction for educational policies. In social sciences, comparisons have always been considered to be a substitute for experimentation, and comparative education relies on this idea. Who would maintain that it is not legitimate to borrow some ideas from our neighbors, even if it often risky and difficult?

Borrowing some elements to boost efficiency?

Even if it is better to have figures rather than impressions or ideas (especially with regards to educational matters, for which everyone has ideas!), the interpretation of comparative figures is never straightforward. Indeed, interpretation requires making hypotheses about the causes of variations observed across countries. When observing that some countries have both high achieving students and some educational features, it is appealing to make a leap from correlation to causality, and “explain” the former using the latter. However, this is clearly a risky undertaking when using PISA data, for several reasons.

First, PISA data are cross-sectional. For 15-year-olds, those differences are a joint and cumulative result of the curricula in use and the educational structures and types of teacher-student relations that have prevailed throughout schooling, as well as of all the learning experiences from birth to the present. In other words, they not only reflect what students have acquired during schooling but what they have learned from their family, peers, the media, daily life, etc. And the capacities so assessed may be more or less dependent on what school has offered. Actually, cross-sectional IEA surveys such as PISA do not permit the evaluation of the real efficiency of educational systems, which would require longitudinal data.

Another point is that efficiency in education probably relies on a combination of factors rather than on just one. Too often the OECD and certain countries attempt to draw lessons from correlations observed between one precise characteristic of the education system and student performance. In that case, the risk lies in over-interpreting the correlation which may appear if specific countries are included in a sample and vanish once they are removed. More importantly, the correlation may be false because it may be produced by other characteristics statistically linked to the former one. More generally, a classic stumbling block in international comparison is to attribute too much importance to an element isolated from its context. In general, it is risky to rely upon correlations that match a given performance and a given educational system characteristic, since what is probably at work is a whole set of traits that constitute each country's societal coherence. Ideally, comparison should take into account these overall structures. Certainly this runs counter to the more or less explicit aim of those comparisons, i.e. to identify supra-national "laws"; in that case only, relationships between a given system's mode of organization and a given student "product" would be transferable.

This also runs counter to policy-makers' concern regarding the transferability of successful experiments implemented in other countries. This transferability is often imperfect because what works in one place, with certain kinds of students and teachers, does not always prove successful in another context. Again, this is because what is probably at work is a whole set of traits. Ideally, any comparison should entail all of these overall structures. Moreover, successful experiment should be adapted to local constraints and culture; for instance, a successful strategy for dealing with immigrants in one country, where they are only a tiny minority and come from specific countries, may fail completely in a country where their number or origin are very different. This is also why the European "open method of coordination" should be considered with some caution, as well as so-called "evidence-based" policies.

This is an old debate in comparative education. It opposes what some researchers (Fuller & Clarke, 1994) call a "policy mechanism" – identifying particular school inputs that would raise student achievement no matter what the context – with "classroom culturalists" who maintain that input effects are always conditioned by the social rules of classrooms and settings. Without entering into this debate here, what should be highlighted is that isolated pieces of information are not sufficient for mechanically assessing systems. Indicators are valuable information, but evaluation is a different matter, requiring the combination of

M. DURU-BELLAT

indicators and most of all, the more qualitative interpretation of their meaning (Bottani, 2008).

Evaluation requires constant trade-off and value-loaded choices

While indicators provide valuable and necessarily partial pieces of information, assessing the quality of an education system (in order to draw some political lessons) requires implementing a much wider perspective, and not a value-free one. Thus, limiting oneself to PISA data is clearly impossible.

At every step, apparently technical considerations are mixed with more normative ones. Consider, for example, the way in which one country defines excellence in education. Actually, many things must be specified. Should student performance be defined: 1) when it is being produced (e.g. reading skills in primary education), i.e. when individuals are still in school; 2) later on in their schooling career, based on secondary education requirements; 3) when they have left school and are adults, based on “life skills”? Another question: excellence in what? According to which of the many aspects that define students is that excellence assessed? Should we define a few fundamental subjects or take the average of a variety of subjects? Should we give more weight (or no weight at all) to some of them? Another aspect is that education aims at imparting formal knowledge on the one hand, and values and behavioral norms on the other hand. While the former are relatively easy to measure and compare, this is less true for the latter. None of these questions are purely technical.

Moreover, it proves impossible (or at least debatable) to define excellence without considering equity. Whose excellence is it? Should we take into account the average level or the level achieved by the top or bottom 10%? Should the focus be placed upon differences across individuals or groups, or on the proportion of the population that is below some identified minimum threshold, according to a Rawlsian perspective? As far as school is concerned, the content of this minimum package may be identified with a short-term perspective and an academic nature (e.g. every individual has to reach a set minimum reading level by the end of primary school). However, a medium-term perspective can also be adopted, defined as what is necessary in adulthood (for a “good life,” as Rawls would suggest; for researchers adopting a Rawlsian philosophy, see Benadusi, 2007). This would lead us to make the case for threshold indicators such as the minimum level of formal academic achievement for every student (in PISA, for instance, the percentage of a country’s population below a defined threshold). This would be an equity criterion in democratic societies, warranting equality of citizenship or basic human rights.

Last but not least, when assessing equity at the system level, it is impossible not to take into account both the output dimension and the way by which these outputs have been produced. Should we consider the volume of public resources allocated to education or the volume of resources students actually receive, these resources being not only money but also time, content, pedagogical assistance, etc.? Another set of questions concerns whether excellence is better defined when taking into account the distribution of those resources between students (e.g. their degree of

concentration). So defined, excellence would be a mix of average performance and equity: one would not maintain an excellent system if it received very abundant resources but allocated them to a very restricted part of its population. Similarly, one would not say a system is excellent if it generates large inequalities in student performance, or if it generates small inequalities but at a very low level of achievement. Either way, these choices obviously embody value considerations.

All this implies that when assessing excellence, it is important not to limit oneself to measurements of student achievement but rather to include measurements of system characteristics such as coverage, financing (public/private), and structures (early/comprehensive tracking, types of student groups, etc.). The poorer a country is and the more limited its resources are, the more important it is to focus on system characteristics; however, this remains a general rule since resources are always scarce and excellence a never fully achieved goal!

Whatever the country, it is a value-laden choice to decide to what extent the system should focus on equity consideration. The question is how many years of schooling is the objective for achieving universal coverage? Here the equity consideration prevails. However, at this stage, efficiency once again matters. A primary concern is how much basic education should be secured by every child in the country. Since resources are always limited, this requires taking the efficiency consideration into account. Giving very little to everybody for the sake of equity would make no sense if the amount given was so small as to not be effective. Equity cannot be the lone argument; it depends on whether it is achieved at a low or high level of performance. Here again, value judgments are unavoidable.

In contrast, the limits and risks attached to benchmarking are clear: the logic of benchmarking leads one to assess a system from a purely instrumentalist point of view, not on the basis of its principles or values, but solely on the basis of its outcomes, measured using available indicators. From that perspective, the only concern is efficiency, so that “anything goes”... All systems are then judged comparable, and it appears plausible to borrow one feature to one country and another from another and build the ideal system, as if playing with a set of building blocks. Efficiency is the justification for every choice, all of which are technical and defined by experts.

CONCLUSION

Anyway, in our opinion, legitimate debates and criticisms that focus on benchmarking and the inclusion of PISA data should not lead us to renounce processes that evaluate education systems based on their output. Educational choices and policies have to be justified, and the results they produce – knowledge, skills, attitudes, etc. – are crucial ingredients. One could even go so far as to be critical of any system that does not give great importance to the changes education produces in students, since to educate students is to make them change in some way, which justifies the input/output perspective! Even if some lively debates do exist regarding this issue (e.g. Olson versus Slavin, 2004), what alternative methods exist for justifying educational policies when so many decisions are to be taken?

However, one can easily understand the variety of resistance to such external evaluation. As regards teachers, they may legitimately think that precise evaluation of the results (i.e. output) of their actions provides a stronger framework for their work. Conversely, some voices maintain that external evaluation is what makes more autonomy possible, allowing teachers to experiment with practices of their choice as long as the output goals are met. Others think that with the development of benchmarking and results-based evaluation, the way the system is regulated is changing, with more power being given to experts and scientists, while policy-makers and actors have no choice other than to execute what the science demonstrates. However, mobilizing education partners with global (and ultimate) objectives may be judged more important than polarizing them with intermediate and often short-sighted ones.

A last point must be raised, which is relevant both for teachers and researchers in comparative education. Any measurement or evaluation of output is based on what exists at the present time, no matter how sophisticated the methods of analysis are. For example, one pedagogical practice may appear to be the best among all those that exist. That said, perhaps something still better could be designed and tweaked! Furthermore, it is not simply enough to observe that an action does have some impact. Indeed, it is necessary to be reasonably sure that this action is better than all others involving the same amount of resources. Verifying this requires conducting a variety of experiments and, of course, not renouncing pedagogical imagination. To do so would be disastrous for comparative education. However, even if limiting oneself to using PISA data would be a complete failure and would constitute an a-sociological perspective, what would neglecting student output do to comparative education?

REFERENCES

- Benadusi, L. (2007). Education equality indicators in the nations of the European Union. In R. Teese, S. Lamb & M. Duru-Bellat (Eds.), *Internatinal studies in educational inequality* (pp. 155-190). Dordrecht: Springer, vol.1.
- Bottani, N. (2008). Le niveau d'huile, le moteur et la voiture : les enjeux d'une évaluation de la qualité de l'enseignement par les indicateurs. *Education et formations*, 78, 9-23.
- Duru-Bellat, M. (2004). *Social inequality at school and educational policies*. Paris: UNESCO.
- Duru-Bellat, M., & Suchaut, B. (2005). Organization and context, efficiency and equity of educational system. *European Educational Research Journal* 4(3), 181-194.
- Duru-Bellat, M., Mons, N. & Bydanova, E., (2008), Cohésion scolaire et politiques éducatives. *Revue Française de Pédagogie*, 164, 37-54.
- Duru-Bellat, M. & Bydanova, E. (2009), L'éducation nationale: l'illusion égalitaire. *Revue du CIEP*. (www.ciep.fr/ries/colloque-2009).
- European Commission, (2007). *Progress towards the Lisbon objectives in Education and Training*, SEC, 1284.
- Fuller, B. & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom, tools, rules and pedagogy. *Review of Educational Research*, 64(1), 119-155.
- Fuchs, T. & Woessmann, L. (2004). What accounts for international differences in student performance? A Re-examination using PISA data, *CESifo Working paper* No. 1235 (www.CESifo.de).

FROM THE APPEALING POWER OF PISA DATA TO THE DELUSIONS OF BENCHMARKING

- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education*, 11, 319-330.
- Haahr, J.H. (2005). *Explaining student performance. Evidence from the international PISA, TIMSS and PIRLS surveys*. Danish Technological Institute (www.danishtechnology.dk).
- Hanushek, E.A. & Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 11, 63-76.
- Heyneman, S. & Loxley, W. (1983). The effects of primary school quality on academic achievement across twenty-nine high and low income countries. *American Journal of Sociology*, 88(6), 1162-1194.
- Marks, G. N., Cresswell, J. & Ainley, J. (2006). Explaining socioeconomic inequalities in student achievement: The role of home and school factors. *Educational research and Evaluation* 12(2),105-128.
- OECD (2004). *What makes school systems perform?* Paris: OECD.
- OECD (2007). *No more failures. Ten steps to equity in education*. Paris: OECD.
- OECD (2009). *Highlights from Education at a Glance*. Paris: OECD.
- Olson, D.R. (2004). The triumph of hope over experience in the search for “What Works”, *Educational Researcher*, 33(1), 24-26.
- Perry, L.B. (2008). Using PISA to examine educational inequality, *Orbis Scholae*, 2(2), 77-86.
- Slavin, R. R. (2004). Education research can and must address “What Works” questions, *Educational Researcher*, 33(1), 27-28

Marie Duru-Bellat
CNRS-OSC (Paris)

JAVIER SALINAS JIMÉNEZ AND DANIEL SANTÍN GONZÁLEZ

ARE YOU ON THE EDUCATIONAL PRODUCTION FRONTIER? SOME ECONOMIC INSIGHTS ON EFFICIENCY FROM PISA

INTRODUCTION

In most countries, the public sector considers basic education as a preferential good which is essential for getting the development of a worthy life. Since the Second World War until now, many countries have introduced free, compulsory, and sometimes decentralized educative systems, which have contributed to their economic development.

Mincer (1974) emphasized the positive role of years of education on economic profitability, with accumulated human capital explaining wage differences among individuals. The relationship between education and income can be explained by the Human Capital theory, which was initially suggested by Becker (1964). This theory points out that education is not just a consumption good but also an investment asset; i.e. education increases productivity and this greater productivity would be rewarded in the job market with a higher wage. This premise was one of the bases that gave birth to the Economics of Education as a discipline. Nowadays there is a broad consensus about the fact that the educational level of a country is a key factor for its economic growth, its productivity and, in short, for its welfare. [Barro, 1991; Mankiw et al., 1992].

When assessing the educational level of a country, qualitative aspects are extremely important. Some current research corroborates that what really matters is not just the years of education, but also their quality¹. It is obvious that a year of education received in one country is not the same than that received in another one since the expected performance would be greater in countries with higher quality education. The academic results obtained by schools as measured in standardized tests is a common way of measuring quality across different schools and countries. A greater overall quality in the educational system of a country has been shown to have a strong and stable association with greater economic growth rates. This result should have significant implications both in the economic and educative structure of a country. From the Public Sector's perspective, investigating the variables that have a stronger effect on students' academic performance is a key factor in order to improve the educational system in general and individuals' educational attainment in particular.

In general, there is a large consensus about what measures reflect a higher quality of education. For instance: a decrease in general school failure, increasing

years of schooling or, most importantly, improving the quality of the received knowledge. However, the way in which these goals can be reached is not so clear. It is commonly accepted that the extent to which educational systems can be improved depend to a great extent on budgetary restrictions and the political discussion is often reduced to the more money-better results statement. *Caeteris paribus* it seems clear that budgetary increases should lead to an improvement in education quality. However money seems to be only a necessary condition, but not a sufficient one, in order to achieve significant advances in a country's quality of education. A successful educative system should deal with problems related to educational quality from a multidimensional point of view, and should pay attention to specific student's characteristics, family level variables and to school level ones; and should of course ensure an efficient management of educational resources.

In this chapter we present an overview of the problems related to the assessment of efficiency in education and we describe how the PISA data have been used for carrying out these kind of studies. In the second section the educative process is described from an economic perspective, studying the relationship between educational inputs and outputs and assuming the possible existence of inefficient behaviors. In the third section we will discuss the usefulness of productive frontiers for evaluating the educative institutions in a comparative way (yardstick competition) and for identifying the best schools that will serve as reference (benchmarking) with regard to inefficient ones. The fourth section is devoted to describe the main variables that have been used in empirical studies that use PISA data and to summarize the conclusions that can be derived from those studies. While describing the main variables and findings reported in previous work is the main goal of the paper, we will also point out a few recommendations on what additional information should be included in future waves of the PISA reports in order to improve the quality of the empirical analyses that could be conducted using PISA data. Lastly, the final section is devoted to concluding remarks.

INPUTS, OUTPUTS AND EDUCATIVE EFFICIENCY

In Economics of Education, educational production functions relating the outputs (i.e. results reached), to the inputs, (i.e. resources consumed), are usually adopted². One of the most widely used models can be outlined as follows (Levin, 1974; Hanushek, 1979):

$$A_{is} = f(B_{is}, S_{is}, P_{is}, I_{is}) \quad (1)$$

where i denotes the student and s represents the school. This model assumes that there is a productive technology f which establishes that the vector of results, A_{is} (commonly measured as the result from one or several objective knowledge tests), is determined by four kinds of educative inputs conditioning the degree of knowledge acquisition. The first vector of variables, B_{is} , relates to the student's socioeconomic family background, such as: parental educational attainment, access to cultural goods at home, and the attention that parents pay to their kids. The

ARE YOU ON THE EDUCATIONAL PRODUCTION FRONTIER?

second group, S_{is} , consists on variables related to the school resources, e.g. educational expenses per student, quality of the resources and school facilities, etc. The third group, P_{is} , collects the so-called peer group effect; the underlying idea is that the knowledge that a student can assimilate depends on his or her mates' features and on the different interactions that are generated among students, both inside and outside the classroom. Lastly, I_{is} refers to student's abilities, which could refer both to innate capacity and to other non-observable student's characteristics. Due to the lack of information about this group of variables, researchers usually assume that they are normally distributed, and therefore include them as part of the model's error term.

The previous educational production function can take into account that there are inefficient behaviors from both the schools and the students; this possibility is considered in Equation 2:

$$A_{is} = f(B_{is}, S_{is}, P_{is}, I_{is}) - u_{is} \quad (2)$$

where u_{is} is the level of efficiency. The idea of inefficiency suggests a waste of the existing resources when carrying out a productive activity. The students' inputs, such as those regarding individual, familiar and school variables, can present right values and, however, the academic results can be low if they are compared to those of similar students in the same conditions. This is possible because of the existence of inefficient behaviors. Inefficiency appears when one or several involved agents, such as students, students' parents, teachers, school principals, authorities, etc., do not properly fulfill its role when the inputs are to be turned into educational outputs. This leads therefore to a lower amount of outputs (or to the use of a greater amount of resources to get a given level of output) compared to what it could be obtained.

Figure 1 illustrates the inefficiency concept in a simple framework where only two outputs are considered.

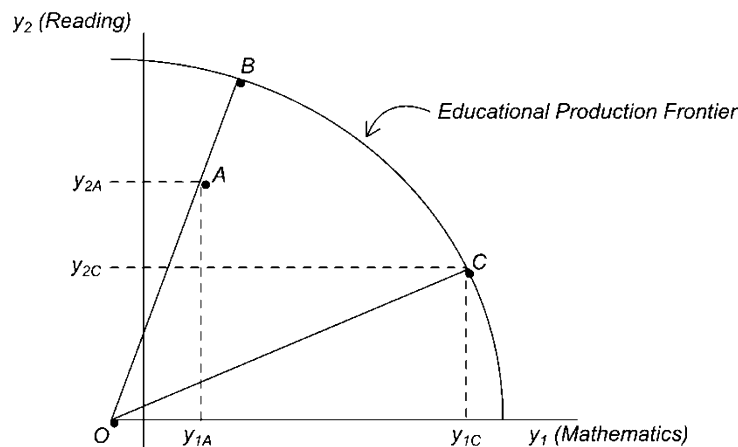


Figure 1. Educative Efficiency Concept.

This figure shows the case of three different units, e.g. three schools, A, B and C, that use the same amount of inputs, x , for getting two outputs like math and reading. The units in the production frontier, B and C, are efficient since given the inputs they produce the highest possible amount of both outputs. However, unit A is inefficient because it lays below the production frontier and it could therefore improve its results.

Reasons for school inefficiency are diverse and are related to the legal framework, the pedagogical system and school management and the process of students' selection. Regarding the students, though it is clear that variables included in Equation 1 affect the performance of students, there are other features that may condition the educative process. For instance, immigrant students face additional difficulties due to the integration process. Another example would be the type and the relationship between the parents, as an indicator of the affective difficulties that may influence in the learning process.

THE MEASUREMENT OF EFFICIENCY IN EDUCATION: TECHNIQUES AND CONCEPTS

The analysis made in the previous section assumes that we know the technology that determines the production frontier. However in practice it is difficult to know the exact technological relationship between educative inputs and outputs (f technology in Equation 1), which is problematic in terms of the empirical measurement of efficiency. Efficiency measurement requires hence an intermediate step. Before calculating the efficiency scores of a set of units, we need to know the production frontier. In practice, the production frontier must be inferred from empirical data and, therefore, 'true efficiency' scores can not be calculated but rather efficiency scores coming from 'best observed performance'. The empirical measures of efficiency are then determined by measuring the distance of the different units to the production frontier, which is built from the best practices observed in the sample.

Two main approaches can be used to estimate the production frontier:

- Parametric Approaches: An a priori well defined functional form (e.g. Cobb-Douglas, Douglas, quadratic, translog, etc) is set for the production function. Its parameters are estimated in such a way that all productive units lie on or below the frontier. In general, stochastic frontiers in which the econometric error term is separated into two independently distributed components, i.e. random noise and inefficiency, are usually adopted.

- Non-parametric Approaches: Non-parametric approaches do not assume that the underlying technology adopts a specific functional form. Subject to certain assumptions regarding the structure of the production technology (e.g. free disposability, convexity, etc.), the units are enveloped as tightly as possible. This is done by solving a mathematical programming model. In general, the expression 'data envelopment analysis (DEA)' is used to refer to those nonparametric approaches. This approach has had a very important theoretical development which was directly related to the numerous empirical studies undertaken since the late seventies, when this technique was firstly formulated.

ARE YOU ON THE EDUCATIONAL PRODUCTION FRONTIER?

A question arises when an efficiency analysis of educational institutions is to be made: which methodology does offer more reliable results for the educative efficiency measure? The answer to this question is not obvious since both the parametric and non-parametric approaches show strength and weakness which have to be taken into account before using them in the educative context. [Table 1](#) shows a summary of the main advantages and disadvantages of each method regarding different dimensions of the efficiency analysis.

Table 1. A comparison of efficiency measurement techniques in education

Analyzed Dimensions	Parametric	Non-Parametric
<i>Statistics assumptions</i>	<i>High</i>	<i>Low</i>
<i>Flexibility</i>	<i>Medium</i>	<i>High</i>
<i>Statistical significance</i>	<i>Yes</i>	<i>No</i>
<i>Elasticity calculation</i>	<i>Yes</i>	<i>No</i>
<i>Projection, Generalization</i>	<i>Medium-High</i>	<i>No</i>
<i>Multi-input, Multi-output</i>	<i>Yes</i>	<i>Yes</i>
<i>Cost of the analysis</i>	<i>Medium</i>	<i>Medium</i>
<i>Type of frontier</i>	<i>Stochastic</i>	<i>Stochastic (bootstrap)</i>
<i>Easy inclusion of categorical variables</i>	<i>Yes</i>	<i>No</i>
<i>Number of applications in education</i>	<i>High</i>	<i>Very high</i>

Source: Santín (2006) and own elaboration.

Both the non-parametric DEA and the parametric methods have been used widely in the Economics of Education literature in the last two decades, especially the first one. The traditional advantages of the parametric methodologies and DEA are well known. Regarding the first two dimensions of [Table 1](#), DEA does not assume any functional form about the educational technology which relates productive factors to results and it is able of tracing the productive frontier with some little restrictive assumptions, like convexity and monotonicity. On the other hand, the parametric methods need to elaborate key assumptions about several aspects such as the functional form of the technology, the error distribution function or the inefficiency distribution. On the other side of the balance, the main weaknesses of the DEA methods are reflected in dimensions three, four and five of [Table 1](#). Thus, DEA is very sensitive to the presences of extreme data, it does not allow neither for the calculation of the output-input elasticities nor for making predictions. In addition, the use of categorical variables, though possible, may reduce the sample size and lead to an unworkable analysis. Moreover, total flexibility in DEA has been criticized since important factors may be all but ignored in the analysis; some inputs and output measures may not be considered when assessing the relative efficiency of some schools. As a result, the relative efficiency of a school may not really reflect its performance on the inputs and outputs taken as a whole³.

Regarding other aspects such as the possibility to use many inputs and outputs, the computational effort, or the treatment of the frontiers as stochastic, both techniques share similar difficulties.

WHAT INFORMATION DOES PISA OFFER TO THE ECONOMISTS FOR
MEASURING THE EFFICIENCY IN THE EDUCATIVE FIELD?

In most cases, the measurement of efficiency in education has been carried out with aggregate data at the school level. The main reason for using aggregate data has been the lack of availability of micro databases at the student level. In this sense, the PISA report, and other studies like TIMSS (Trends in International Mathematics and Science Study) or PIRLS (Progress in International Reading Literacy Study)⁴, expands the range of research possibilities within this research agenda by offering data at the student level in a costless way. A major advantage of these databases is that they are designed to allow cross-country comparisons as well as comparisons for different groups of students [Summers y Wolfe, 1977; Hanushek et al., 1996; Santín, 2006; Waldo, 2007]. In addition, it allows us to consider the effects of different socioeconomic characteristics related to family characteristics, and schoolmates effects independently.

The educational outputs.

In general, it can be argued that the educational output has a multidimensional intangible character that makes difficult to define it theoretically and also to measure it empirically. A good education does not only imply the capacity to repeat information and solve problems. It involves also the ability to process the information, to construct analytical reasoning to develop critical thinking and last but not least to behave according to good manners, all of which are very difficult to quantify. As a consequence of this multidimensional character, several intermediate outputs have been considered as good candidates to approximate the final output. In general, most of the work in economics has used performance results in standardized tests as the best way to approximate quality. The PISA report includes data on students' results in standardized tests in reading, mathematics, and science, with each PISA wave focusing specially on one of these subjects. Both PISA 2000 and 2009 give special attention to reading, while PISA 2003 focuses on mathematics and PISA 2006 on science. Moreover the score obtained by each student is not offered as a unique mean value but as five plausible values. These five plausible values represent the level of abilities that each student has; they are basically random values which have previously been extracted from the results distribution function estimated for each student given their answers in each of the tests. Consequently, in order to obtain the mean values and standard deviations that appear in the OECD estimations, estimations using PISA data have to be replicated five times, each of them using one of the different plausible values. It is also noteworthy that the scores reported in the PISA report, are also associated with an illustrative pedagogical interpretation.

Besides academic outputs, other objective outputs from the PISA information can be used. First, the academic year (GRADE) variable would inform us about whether 15 years old students had to repeat any academic year or not; we would lack however, information about the specific year in which the student repeated the academic year. Secondly, and related to the equality of educational opportunities,

ARE YOU ON THE EDUCATIONAL PRODUCTION FRONTIER?

in some countries the ISCEDO (ISCED Orientation) and PROGN (Unique national study program code) variables allow us to know if 15 years old students have been guided towards vocational education (with a higher orientation to the job market) or towards general education (showing greater orientation to University). This information allows us to know if the students and their parents have chosen their orientation in an exogenous way, just influenced by their vital preferences, or, on the contrary, if the educational system tries to guide the best or the worst students towards a specific option, with the consequent equity problems that it could arise. In this sense it would be useful for future waves of the PISA report to indicate at which specific age each student chose her educational track.

The educative inputs

Student's socioeconomic background

The PISA report collects information about the student's socioeconomic background. It gathers information about variables related to parents, such as their studies (MISCED, FISCED, HISCED, PARED), their income and their labor status (BMMJ, BFMJ, BSMJ, HISEI, MSECATEG, FSECATEG, HSECATEG), and to wealth and possessions (CULTPOSS, HEDRES, HOMEPOS, WEALTH). This information is conveniently grouped in an index known as Economic, Social and Cultural Status (ESCS). This index is obtained from HISCED, HISEI y HOMEPOS and, in our opinion, it seems to properly proxy familiar status.

Moreover, there are control variables that offer additional information on the student and her familiar background. Two meaningful variables in PISA 2006 are INMIG and LANGN. They inform about the student's origin and the mother tongue used at home, although in many cases is not possible to identify both the specific country and language. Other relevant variables in PISA 2000, like FASTRUM (familiar structure), which is also present in PISA 2003, NSIB (number of siblings) and BRTHORD (birth order) are not collected in PISA 2006 anymore. We believe that this information should be included again since it can be useful for the purposes of identifying students with a potential high risk of school failure. Another important control variable is the student's month of birth because it may influence students' results at that age [Sprietsma, 2010].

Variables related to the school resources

One of the main problems in the different PISA waves is the lack of objective variables related to the school resources except for the RATCOMP (Ratio of computers to school size) and STRATIO (Ratio of students to teachers) variables. Regarding the ratio between students and teachers, this variable is usually introduced into the efficiency analyses as an input, assuming that smaller groups would lead to improvements in performance. However, in the literature on the Economics of Education does not provide a clear answer regarding what is the optimal class size in terms of facilitating the learning process. Similarly, previous

research is inconclusive regarding what is the maximum number of students that can receive education in a classroom without detrimental effects on learning. In fact, the effect of the number of students on learning is a highly controversial issue (Hanushek, 1986, 1997, 2003; Krueger, 1999; Hoxby, 2000, Rivkin et al., 2005).

More precise objective indicators are therefore required in order to finding out the influence of the school on the educational results. The three PISA waves offer information about the quality of the educative resources which are assigned to the instruction (SCMATEDU), but this variable is constructed on the basis of the school principals' perception and not on objective tests. Another interesting variable, that has similar problems, is SCMATBUI (educative infrastructure), which is present in the two first PISA waves (2000 and 2003). Apart from these variables, it would be very useful to report budgetary information of the schools. Knowing the total budget of each school, which part of the budget is allocated to teaching and non-teaching staff, and current and investment expenditure, would provide very valuable information.

Peer group effect

The existence of non-controllable inputs results in an attribution problem. Thus, the effects of variables that are out of the school's control should be discounted in order to appropriately evaluate its performance, taking into account only those inputs that the school can manage. If those non-controllable inputs are not discounted, the schools showing worse conditions regarding these factors could appear as being inefficient units even if they efficiently manage the inputs under their control. As a consequence, it does not make much sense to publish rankings of different schools on students' performance if this information is not related to what kind of students attend to each school. On the other hand, if schools can select in some way the kind of students, so this variable would not be completely exogenous, some cream skimming effect could happen.

In this sense, data from PISA allow to extract information related to the effect that the schoolmates have on other students. To this end, the average level of the ESCS variable or the percentage of immigrant students could be used. Another variable providing information on this aspect was school disciplinary climate (DISCLIMA in PISA 2000) or disciplinary climate in math lessons (DISCLIM in PISA 2003); this variable disappeared in PISA 2006 and, from our point of view, it should be included again.

SOME RESULTS ABOUT THE EDUCATIVE EFFICIENCY OBTAINED WITH PISA DATA

Since the publication of the PISA reports many different papers have been published measuring efficiency using both parametric and non-parametric methods. [Table 2](#) offers a compilation of some of these studies, summarizing the technique adopted for doing the efficiency analysis, the scope of the study, the variables used therein and the main obtained results.

As shown in [Table 2](#), most of the works analyzing educational efficiency with PISA data use the non-parametric approach (DEA and FDH) instead of the parametric one. This choice may be mainly due to the ignorance about the mathematical form of the education production function and to the consequent difficulty for choosing a parametric specification for the production frontier. In spite of this limitation, some empirical works [e.g. Perelman and Santín, 2008] use a parametric approximation with very flexible functional forms, such as the translog, which allows to exploit the advantages of the parametric approach without making very strong assumptions regarding the functional form.

Secondly, most empirical studies use aggregate data with, the country or the school being the preferred unit of analysis. Working with aggregate data have the advantage of offering less difficulties from a computational point of view and the results may also be more attractive for policy makers and institutions in charge of the design of educational policies. However, we think that as PISA database becomes more widely known, we shall see many more analyses at the student level, which will offer more and better information to policy makers.

Lastly, it is worth remarking that some of the works reported in [Table 2](#) do not use the plausible values of the different objective tests for carrying out the analyses. Instead, they sometimes use an average of the three tests. This wrong use of the data would highlight that a more detailed knowledge of the PISA data is essential for increasing the results reliability. Moreover, the use of the plausible values when non-parametric methods are adopted offers the additional advantage of generating confidence intervals for the measures obtained.

CONCLUSIONS

The evaluation of educational efficiency provides several benefits for the different agents working in the field of education. First, it helps to define and clarify educational goals and provide information on schools' performance, allowing one to know if the required efficiency is achieved, It also allows to know whether available resources are used in an optimal way. Secondly, it provides useful information for parents, taxpayers and policy makers. Thirdly, it can induce greater coordination among different agencies and administrations in charge of the education policy. And lastly, it facilitates yardstick competition between students and schools, learning about best practices and pedagogical or management methods. At the same time it allows to carry out benchmarking with respect to the best performers.

In this context, the PISA report constitutes a very valuable source of information for the analyses of educational efficiency. Although the survey design may not be the most appropriate for the purposes of efficiency measurement, or for performing international comparisons regarding the efficiency of different education systems', there is no doubt that it provides very useful information for evaluating educational policies. In this sense, some modifications and the incorporation of new variables in the survey could contribute to improve the analyses.

In addition, we would like to point out that when schools are used as the units of analysis, two problems that may determine the results and that should be further

analyzed, arise. First, when aggregated data at the school level are used, and given that this aggregation bases on 15 years old students, we evaluate schools, and even the country performance, taking into account the results of only this group of students. Secondly, the mean score in the different PISA tests is usually considered as an output, but two education institutions may have a similar average result but very different distributions. Hence, it would be useful to introduce into the analyses some information about the standard deviation or the percentage of students who do not reach a specific level. In this sense, several concepts such as the stochastic dominance or the level distribution could offer very valuable information about the schools performance, although it is not clear how these variables could join the efficiency analyses.

Finally, we would like to highlight the analytic advantages that would be derived from having a longitudinal database where students were surveyed in two different moments (for instance at ten and fifteen years old). This information would allow us to consider as an output the value added generated by any school. Moreover, a database of this type could offer the possibility of carrying out quasi-experimental design analyses on different education policies, which would in turn allow in mid-term to select the most successful ones for different groups of students.

NOTES

- ¹ See for example Hanushek and Kimko (2000) or Barro (2001)
- ² For example, Bishop and Woessmann (2004) and Woessmann (2003) used educational production functions for considering the effects of family environment and educative resources on the students' academic results.
- ³ Nevertheless, the impact of many of these traditional limitations has been reduced with recent methodological developments of the technique. See, for example, Wilson (1995), Pedraja et al. (1997) and Simar and Wilson (2000).
- ⁴ To obtain more information about the TIMSS and PIRLS studies, see the following website: <http://timss.bc.edu>

REFERENCES

- Afonso, A. & St. Aubyn, M. (2006). Non-parametric approaches to Education and Health efficiency in OECD countries. *Journal of Applied Economics*, 8 (2), 227–246.
- Afonso, A. & St. Aubyn, M. (2006). Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs, *Economic Modelling*, 23(3), pp. 476–491.
- Badescu, M. (2007). Measuring investment efficiency in Public education: some cross-country comparative results. Institute for the Protection and Security of the Citizen. *CRELL Research Paper* 5. 2007 EUR 22719 EN.
- Barro, R. J. (1991). Economic Growth in a cross section of countries, *Quarterly Journal of Economics*, 106 (2), 407–443.
- Becker, G. (1964). *Human capital: a theoretical and empirical analysis with special reference to education*. Columbia University Press: New York.
- Bishop, J.H. & Woessmann, L. (2004). Institutional effects in a simple model of educational production (with J.H. Bishop). *Education Economics*, 12(1).

ARE YOU ON THE EDUCATIONAL PRODUCTION FRONTIER?

- Bradley, S. & Taylor, J. (2002). The effect of the quasi-market on the efficiency-equity trade-off in the secondary school sector. *Bulletin of Economic Research*, 54, 295–314.
- De Jorge, J. & Santín, D. (2010). Determinantes de la eficiencia educativa en la Unión Europea. *Hacienda Pública Española – Revista de Economía Pública*, 193 (2/2010).
- Farrell, M.J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society (A)*, 3, 253–290.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14, 351–388.
- Hanushek, E. A. (1986). The economics of schooling. *Journal of Economic Literature*, 24(3), 1141–1171.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update”. *Educational Evaluation and Policy Analysis*. 19, 141–164.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113, 64–98.
- Hanushek, E. A., Rivkin, S. G. & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics*. 78(4), 611–627.
- Hanushek, E. A. & Kimko, D. D. (2000). Schooling, labor force quality and the growth of nations. *American Economic Review*, 90(5), 1184–1205.
- Hoxby, C. M. (2000). The effects of class size on student achievement: new evidence from population variation. *Quarterly Journal of Economics*, 115, 1239–1285.
- Krueger, A. B. (1999). Experimental estimates of educational production functions. *Quarterly Journal of Economics*, 114, 497–532.
- Levin, H. M. (1974). Measuring efficiency in educational production. *Public Finance Quarterly*, 2, 3–24.
- Lovell, C. A. K. & Muñiz, M. (2003). Eficiencia y productividad en el sector público. Temas dominantes en la literatura. *Papeles de Economía Española*, 95, 47–65.
- Mankiw, N., Romer, D. & Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(2), 407–437.
- Mincer, J. (1974). *Schooling, experience and earnings*. New York: Columbia University Press.
- Pedraja, F., Salinas, J. & Smith, P. (1997). On the Role of Weight Restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis*. 8, 215–230.
- Perelman, S. & Santín, D. (2008). Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. *Education Economics*, To link to this article: DOI: 10.1080/09645290802470475
- Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, 73(2), 417–458.
- Santín, D. (2006). La medición de la eficiencia en las escuelas: una revisión crítica. *Hacienda Pública Española – Revista de Economía Pública*, 177(2), 57–82.
- Simar, L. & Wilson, P.W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6), 779–802.
- Sprietsma, M. (2010). The effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using PISA 2003. *Education Economics*, 18(1), 1–32.
- Summers, A. A. & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review* 67(4), 639–652.
- Sutherland, D. & Price, R. (2007). Linkages between performance and institutions in the primary and secondary education sector. OECD Economics Department *Working Papers* No. 558
- Sutherland, D., Price, R., Joumard, I. & Nicq, C. (2007). Performance indicators for public spending efficiency in primary and secondary education. OECD Economics Department *Working Paper* No. 546
- Waldo, S. (2007). On the use of student data in efficiency analysis. Technical efficiency in Swedish upper secondary school. *Economics of Education Review*, 26, 173–185.

J.S. JIMÉNEZ AND D.S. GONZÁLEZ

- Wilson, P. (1995). Detecting influential observations in deterministic non-parametric frontiers models. *Journal of Business and Economic Statistics* 11, 319–323.
- Wilson, P. (2005). Efficiency in education production among PISA Countries, with emphasis on transitioning economies. *Working Paper*. Department of Economics. University of Texas.
- Woessmann, L. (2003). Schooling resources, educational institutions, and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, 65(2), 117–170.
- Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education, *Education Economics*, 9(3), 245–268.

Javier Salinas
Department of Applied Economics IV
Universidad Complutense de Madrid

Daniel Santín
Department of Applied Economics VI
Universidad Complutense de Madrid

Table 2. Empirical studies which analyze the educational efficiency with the PISA database

AUTHORS	SAMPLE	METODOLOGY	INPUTS, OUTPUTS, CONTROLS	MAIN FINDINGS
Wilson (2005)	5,528 observations, schools in 40 countries from PISA 2000	DEA	4 inputs: home educational resources, family wealth, full-time teachers and proportion of certified teachers. 3 outputs: test scores, grade level attainment and number of students in each school	Emphasis is placed on transitioning countries in the former Soviet Union. Results show that many schools operate under regions of the production frontier where returns to scale are decreasing.
Afonso and St. Aubyn (2005)	18 countries from PISA 2000	FDH and DEA	2 inputs: hours per year in school and teachers per student. 1 output: average from the three test scores.	Finland, Japan, Sweden and Korea are the most efficient countries.
Afonso and St. Aubyn (2006)	25 countries from PISA 2003	DEA + Tobit regression (controls)	2 inputs: hours per year in school and teachers per student. 1 outputs: average from the three test scores. Controls: Parent education and GDP per capita, PPP (USD), public to total expenditure ratio.	After introducing the controls, Portugal, Australia, Korea and Hungary are the most efficient countries.
Perelman and Santín (2008)	6,997 Spanish students from PISA 2003	Parametric distance functions (translog) with an instrumental variable approach for public-private voucher ownership.	8 inputs: Background: ESCS, Late at school School: SCMATEDU, SCMATBUI, SCHAUTON, Class-size Peer: DISCLIM, Mean school ESCS Controls: Gender, console, native, preschool, family, school type.	Once educational inputs and potential school choice endogeneity are taken into account, there are not differences between students' efficiency levels across public and private-voucher schools.
Sutherland, Price, Joumard and Nicq (2007)		DEA and stochastic frontier analysis (SFA)	SFA Background: ESCS, language background School: STRATIO, computer availability, cross products between school variables. DEA 2 inputs: socioeconomic background, expenditure per	Holding resources constant, PISA scores could be boosted by an average of 5% for OECD countries. In Mexico, efficiency is high but PISA scores are low. Decisions about the amount of resources devoted to education need to be made not only in terms of efficiency.

	28 countries from PISA 2003		student. 2 outputs: average from the four test scores and equity objective	
Sutherland and Price (2007)	30 countries from PISA 2003	DEA and SFA	SFA Background: ESCS, language background School: STRATIO, computer availability, school autonomy, ownership teacher qualifications, school size DEA 2 inputs: socioeconomic background, teachers per 100 students. 2 outputs: average from the four test scores and, equity objective (homogeneity of PISA scores).	Greater decision-making autonomy at the school-level tends to be associated with higher levels of efficiency. Small school size and residence-based selection tend to be associated with greater inefficiency.
Badescu (2007)	16 European countries from PISA 2003	FDH and DEA	Inputs: average teaching hours, teachers per 100 students, adult attainment. Output: reading score, equity objective.	The Czech Republic is a good education performer, Finland is the best performer in compulsory education and does not spend too many resources whereas Sweden is also a good performer but spend comparatively more.
De Jorge and Santín (2010)	87,974 European Union students from PISA 2003	DEA + regression + ANOVA (school as factor)	Inputs: ESCS, average ESCS by school, SCMATEDU, SCMATBUI. Output: plausible values in mathematics, reading and science. Controls: EU countries, repeater, immigration status, class-size, disciplinary climate, ownership, family structure, disciplinary climate in math lessons	First-generation immigrants or those who repeat some academic year are more inefficient. Finland, Sweden, Poland, Ireland and Belgium have the most efficient educational systems, while Greece, Slovakia and Spain are the worst. More than a half of inefficiency is explained by school differences in the Netherlands and Austria.

SECTION IV

PISA AND THE IMMIGRANT STUDENT QUESTION

AILEEN EDELE AND PETRA STANAT

PISA'S POTENTIAL FOR ANALYSES OF IMMIGRANT STUDENTS' EDUCATIONAL SUCCESS

The German Case

INTRODUCTION

Migration is a universal reality. The United Nations estimated that, in 2010, approximately every tenth person living in the more developed regions of the world was born in another country (United Nations Department of Economic and Social Affairs [UN DESA], 2009). Accordingly, the integration of immigrants and their descendants is of considerable concern worldwide. School systems play a central role in the integration process. They are in charge of developing knowledge and skills relevant for participation in the receiving countries' economic and socio-political systems, they grant school leaving certificates channelling career options, and they help to transmit norms and values relevant for social cohesion.

Immigrant students¹ lag behind their peers from native families in terms of achievement and school success in many countries. In Germany, for instance, students from immigrant families are much more likely to quit school without a school leaving certificate and much less likely to reach the general qualification for university admission ("*Abitur*") than students from native families (Statistisches Bundesamt, 2010). For a long time, the conditions of these disparities were poorly understood. Large-scale assessment studies on student achievement, such as PISA, have advanced our understanding of immigrant students' educational disadvantage considerably. The present article illustrates this for the German case. We start by describing what was known about the situation of immigrant students in the German school system before PISA. Subsequently, we discuss how the study improved our understanding of this situation. The final section of the article addresses the types of questions PISA cannot answer. As a monitoring device, studies like PISA are powerful tools for identifying strengths and weaknesses of school systems and possible target points for interventions. However, it is impossible to infer what measures should be taken to remedy the problems identified by the data.

THE STATE OF KNOWLEDGE BEFORE PISA

Before PISA and other international large-scale assessment studies such as TIMSS (Third International Mathematics and Science Study) or PIRLS (Progress in International Reading Literacy Study) were carried out, little was known about the outcomes of the German school system in general, and about the situation of

immigrant students in particular. Prior to these studies, analyses exploring the situation of immigrant students had to rely on data bases with limited potential in this regard. Until 2005, the official statistics (*Microcensus*) only recorded respondents' nationality, so that it was impossible to identify naturalized immigrants (Statistisches Bundesamt, 2006). Similarly, the German Socio-Economic Panel Study (SOEP), which was introduced in 1984, failed to collect comprehensive information on the migration background of its participants for a long time. Before 2001, immigrants born in Germany and possessing German citizenship could not be identified reliably in the data set (Lohmann, Spieß, Groh-Samberg & Schupp, 2008). Moreover, although the SOEP oversampled immigrants from five countries from its start in 1984 and, in 1994, added a sample of immigrants who had migrated to Germany after 1984, immigrants from other countries as well as recent migrants are not sufficiently represented in the data set. In addition, the SOEP focuses on households as sampling units and the sample sizes of children and adolescents living in these households within each cohort are too small for in-depth analyses

Even more importantly, none of the data sets available in Germany before large-scale assessment studies were carried out contained information on students' levels of achievement. In order to explore determinants of disparities in educational success between immigrant students and students without an immigrant background, however, such information is essential. Moreover, international comparisons of disparities in school success need to be based on achievement data. As school leaving certificates vary across countries, it is difficult to compare them internationally.

PISA provides representative data on students' achievement levels in reading, mathematics, and science for a large number of countries. The study also collects background information related to migration, such as students' age at the time they moved to the receiving country, the first language they learned as a child, and the languages they speak at home. In addition, the assessments include measures of students' school-related motivation, attitudes, and aspirations. The assessment design makes it possible to compare the magnitude of disparities between immigrant students and students from native families internationally and to explore the relative role of potential determinants of immigrant students' achievement within and – to a more limited extent – between countries.

In the following section we first describe what Germany has learned from PISA about the size and composition of immigrant student groups in secondary schools. After a short summary of the knowledge we have gained about the determinants of these students' disadvantages in terms of track attendance, the section presents findings on potential causes of achievement differences between students from immigrant families and students from native families at different levels of analysis.

PISA'S POTENTIAL: WHAT WE HAVE LEARNED FROM PISA ABOUT IMMIGRANT STUDENTS' EDUCATIONAL DISADVANTAGES

Immigrant students in Germany

At present, approximately one fifth of the population in Germany has an immigration background (Statistisches Bundesamt, 2010). Germany's immigrant

population is composed of four major groups and their descendants: the so-called *Guest Workers*, who were recruited by the German government from the 1950s to the 1970s in Southern and South-Eastern Europe to overcome the shortage of labour; asylum seekers and refugees, who mostly entered the country before the asylum laws became more restrictive in 1993; ethnic Germans, who came from the former Soviet Union and Eastern European countries (*Aussiedler*); and immigrants from European Union (EU) countries, who are allowed to settle in Germany due to the right of free movement of EU citizens within the EU. The *Guest Workers* as well as asylum seekers and refugees were initially expected to return to their countries after a temporary stay in Germany. Yet, many of these immigrants continued to reside in Germany and brought their family members to join them. For the *Aussiedler*, in contrast, permanent residence was intended from the start. Unlike members of other immigrant groups, they were automatically granted German citizenship and various measures were taken to support their integration.

The fact that Germany is an immigration country was denied for a long time. To some extent, this was possible because the actual number of immigrants living in the country was unknown. As pointed out above, the *Microcensus* used to define immigrants strictly on grounds of their citizenship rather than on the basis of their migration history (Statistisches Bundesamt, 2006). As a result, a large proportion of immigrants remained invisible. *Aussiedler* and their descendants, who are automatically granted German citizenship, as well as other naturalized immigrants and their children could not be identified in the data. Similarly, the school statistics distinguished between students with German citizenship and students holding the citizenship of another country (referred to as “foreign students” in the remainder of this article). In 2000, the *Microcensus* reported a proportion of 8 per cent foreign students in German schools (Statistisches Bundesamt, 2001).

The PISA study established a more comprehensive indicator of immigration background by recording students' and their parents' countries of birth. Based on these data, the study revealed a much higher proportion of immigrant students attending German schools than the official statistics. In the first PISA cycle, Baumert and Schümer (2001) found that 22% of 15-year-old students in Germany had at least one foreign-born parent. This number was largely confirmed in the more recent PISA assessments. For instance, PISA 2006 reported a proportion of 19% immigrant students (Walter & Taskinen, 2007).

The *Microcensus* survey subsequently adapted the questions it uses to capture respondents' immigration background. Since 2005, it identifies immigrants on the basis of their own, their parents', and their grandparents' country of birth, independent of citizenship (Statistisches Bundesamt, 2009). The survey showed that, in 2009, 32 per cent of the population in Germany under 15 years of age had an immigration background (Statistisches Bundesamt, 2010).

Partly due to the relatively high numbers of immigrants revealed by PISA and subsequently the *Microcensus*, it became more difficult to uphold the claim that Germany is not an immigration country. As a result, public debate has shifted to the question of integration, asking how well integrated immigrants are in Germany and what should be done to improve their integration. Despite this shift in focus, the discussions continue to be controversial and heated.

Determinants of disparities in track attendance

It was evident long before PISA that foreign students were considerably less successful in terms of track attendance and school-leaving certificates than students with German citizenship. In the school year 2000/2001, for instance, only 14 per cent of the foreign students living in Germany attended the highest track (*Gymnasium*) of the tripartite secondary school system in comparison to 32 per cent of the German students. Approximately 44 per cent of the foreign students, in contrast, visited the lowest secondary school track (*Hauptschule*) while only 19 per cent of the Germans attended this track. Every fifth foreign student but only every tenth German student left the school system without a school leaving certificate in the year 2000 (Statistisches Bundesamt, 2001).

Analyses of the Socio-Economic Panel (SOEP) data revealed some factors that partly explain these disparities between students from immigrant families and students from native families. In addition to the socio-economic status of the families, the educational level of the parents, and their duration of stay in Germany, the intention to return to the country of origin as well as cultural orientations emerged as important factors predicting immigrant students' track attendance (Diefenbach, 2002). However, because the SOEP does not include achievement data, the claim that tracks attendance differences may also be due to discrimination, with immigrant students having a lower chance of attending the highest tracks even when their achievement levels are comparable to those of German students (Gomolla & Radtke, 2002), could not be explored.

Based on data of the first PISA cycle, Baumert and Schümer (2001) showed that 15-year-old students whose parents were both born in Germany had a 4.4 times higher chance of attending the academic track (*Gymnasium*) than did immigrant students whose parents were both born abroad. In line with analyses of the SOEP data, these disparities could partly be explained by the lower socio-economic status of the immigrant families. Yet, when adolescents with the same socio-economic background were compared, students from native families were still 2.7 times more likely to attend the *Gymnasium* than their peers from immigrant families. Comparing students with the same level of reading achievement in German, however, did not reveal any significant differences between immigrant students and adolescents from native families in their chances of attending a *Gymnasium* (see also Walter & Taskinen, 2007). This pattern of findings suggests that immigrant students' disadvantages in track attendance are primarily due to their disadvantages in achievement, and it thus contradicts the idea that secondary school track decisions are systematically biased against immigrant students (e.g., Gomolla & Radtke, 2002).

Subsequent longitudinal studies clearly confirm this conclusion. A study conducted by the *Max Planck Institute for Human Development* examined a representative sample of elementary school students at the transition from elementary school to secondary school in Germany (Maaz, Baumert, Gresch & McElvany, 2010). Using data from this study, Gresch and Becker (2010) compared the chances of making the transition to the highest secondary school track (*Gymnasium*) for students from the two largest immigrant groups in Germany (children of Turkish descent and *Aussiedler*) and students from native families. As expected, they found that

children from both immigrant groups had lower chances of entering the *Gymnasium* than their peers from native families. This disparity could be accounted for by the lower socio-economic status of the immigrant families. Given the same socio-economic status, immigrant children's transitions did not significantly differ from their non-immigrant peers. When the students' achievement levels were additionally taken into account, the effect was even reversed. Given the same level of achievement, children from families of Turkish descent had a higher chance of attending the *Gymnasium* than students from native families ($OR^2 = 3.35, p < .05$). For students with an *Aussiedler* background, the effect tended into the same direction but failed to reach significance ($OR = 1.32, n.s.$). When socio-economic status, achievement level, and the school's recommendation for future track attendance were simultaneously taken into account, both immigrant groups had significantly higher chances of attending a *Gymnasium* ($OR = 4.83, p < .05$ for students from Turkish families; $OR = 2.38, p < .05$ for students from *Aussiedler* families).

This pattern of results thus corroborates that immigrant students' underrepresentation in higher school tracks can be explained by their disadvantages in socio-economic status and achievement rather than biased track decisions (see also Kristen & Dollmann, 2009; Müller & Stanat, 2006). However, it certainly is possible that immigrant students are systematically disadvantaged at the level of instruction. Teachers tend to adapt their instruction to the average achievement level of the class (Hattie, 2001). In addition to students' actual achievement level, stereotypes may influence their instruction (Walter & Stanat, 2008). There is some indication that negative stereotypes about Turkish immigrants exist in Germany (Schofield, 2006). As social psychological research has shown, such schematic views of groups can influence behaviour even in persons who do not actually believe that the stereotypes are true (e.g., Nelson, 2002). Teachers may, for example, have lower expectations for students of Turkish descent, which, in turn, might affect what they demand from these students in daily instruction. This could result in reduced learning outcomes and, consequently, lower these students' chances of making the transition to the higher tracks. The extent to which this is in fact the case, however, has yet to be explored.

Determinants of immigrant students' disadvantages in achievement

In order to examine the conditions of immigrant students' disadvantages in achievement systematically, multiple levels of analysis need to be taken into account. To organize the analyses, it is helpful to draw on a conceptual framework that differentiates between distal and proximal factors determining school-related competence development (Stanat, 2006a). The framework conceptualizes the developing individual as embedded in several interacting layers of his or her environment (see figure 1). Ultimately based on Bronfenbrenner's (1979) ecological model of human development, it was developed by an OECD task force on teaching and learning (cf. Baumert, Blum & Neubrand, 2004). In this article, we will focus on four of the layers included in the model: the national/societal level, the school level, the individual student level, and the level of teaching and learning. For the first three layers, we will present analyses of the PISA data on determinants

of immigrant student success. For the teaching and learning layer, we will demonstrate that other study designs than international student achievement studies such as PISA are more informative.

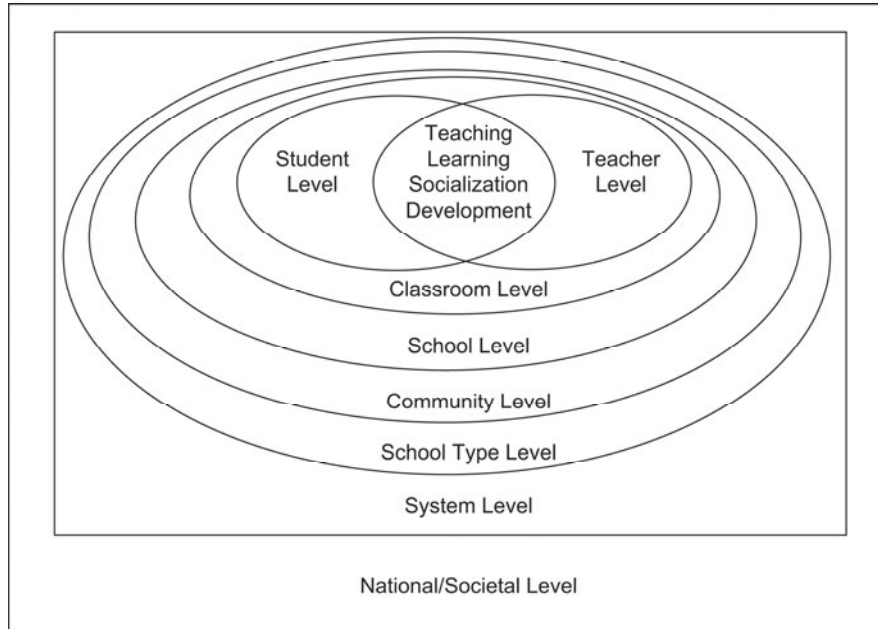


Figure 1. Levels of analyses exploring disparities between students from immigrant families and students from native families.³

National/Societal level. Past PISA studies demonstrated that, in most countries, immigrant students reached lower levels of reading literacy than students from native families. At the same time, the disparities varied considerably between countries, with the gap being particularly large in Germany (Baumert & Schümer, 2001; OECD, 2007; Stanat & Christensen, 2006). The PISA metric is calibrated such that the mean across participating OECD countries is 500 points with a standard deviation of 100 points. In PISA 2006, first generation immigrant students (both student and parents born in another country) in Germany lagged 70 points behind their native peers in reading, and second generation immigrant students (student born in Germany, parents born in another country) reached 83 points less than students from native families⁴ (OECD, 2007). The performance gap across all OECD countries amounted to 52 points for the first generation and 42 points for the second generation of immigrants. In some countries, especially in traditional immigration countries, the disadvantages were considerably smaller or non-existing. In Australia, for instance, first generation immigrant students reached exactly the same average scores as native students and second generation immigrant students scored even 7 points higher than students from native families. First generation

immigrant students in Canada reached only 19 points less than their native peers, while the second generation scored exactly the same as students without an immigrant background.

It is sometimes assumed that the international variations in performance gaps between immigrant students and students from native families are due to differences in integration policies and practices. These arguments frequently fail to take into account that countries also differ in terms of their immigration policies and practices, which are likely to affect the integration process considerably. Immigration policies determine the composition of immigrant populations by defining who is allowed to move to a given country for what reasons and for how long. Bourhis and colleagues stress the interdependence of immigration and integration policies (Bourhis, Moise, Perreault & Sénécal, 1997). According to these authors, immigration policies and practices shape the context in which the integration and acculturation of immigrant populations take place.

Immigration policies differ considerably between countries. Traditional immigration countries, such as Canada or Australia, base immigration decisions on indicators of qualifications immigrants will bring to the labour market. Canada, for instance, has a highly restrictive immigration policy and admits immigrants based on a system assigning points to such characteristics as level of education and English language proficiency. Immigrants are only admitted if they reach a specified number of points. In Germany and other central European countries, by contrast, no comparable systems are in place. In fact, many immigrants in Germany are unskilled workers who were recruited as *Guest Workers* to carry out physically demanding work in industry, such as assembly line production, requiring low qualification levels. As a consequence, the immigrant populations in different countries vary considerably in terms of resources crucial for successful participation in the labour market and in the educational system.

Although it is impossible to separate clearly the effects of immigration policies from the effects of integration policies within a country, PISA can give some indication of their relative importance. The extent to which immigrant populations differ in terms of relevant resources can be estimated based on their socio-economic background and level of education. Especially for the first generation of immigrants, these indicators should to a large degree reflect the position of the families at the time they entered the receiving country. The reduction in disparities between immigrant students and students from native families that results after statistically controlling for these factors can be attributed to the composition of immigrant populations which, in turn, is at least partly determined by immigration policies and practices.⁵

Using data from PISA 2006, [Figure 2](#) shows the results of regression analyses with the language spoken at home as the predictor and reading literacy as the outcome. The dark bars depict the size of the achievement gap between students speaking another language at home and students speaking the test language at home. The differences vary considerably between countries, with nearly 110 points on the PISA scale in Belgium and only 16 points in Australia. After controlling for the socio-economic status and educational background of the families, the disparities are reduced. As the shaded bars show, however, the gap between the

groups continues to be substantial in some countries. This suggests that the international variations in disparities between immigrant students and students without an immigrant background cannot be fully attributed to countries' immigration policies and, hence, the composition of their immigrant populations. Integration policies and practices are likely to play a role as well. For example, countries differ in the approaches they employ to support second language acquisition for immigrant students (Stanat & Christensen, 2006). However - because school systems vary with regard to a large number of confounded factors - to what extent these and other practices related to integration explain the relative achievement levels of immigrant students is difficult to determine.

School level. In addition to features of the national/societal and system levels, school level factors influence student achievement as well. It can be expected that some schools are more successful than others in facilitating immigrant students' educational success. One aspect of the school context that research has explored systematically is the composition of the student body.

Theoretical frameworks of integration and acculturation describe these processes along two dimensions: integration into the majority community and integration into the ethnic community within the receiving country (Berry, 1980; Esser, 2006). A situation in which immigrants are not integrated into the majority community but well integrated into their ethnic community is referred to as segregation. At the residential and school levels, a high concentration of immigrants of the same origin and a low proportion of persons without an immigrant background is often used as an indicator. Esser (2001, 2006) proposes that such a pattern impedes social integration and school success of immigrants, as segregated environments are associated with fewer opportunities and reduced motivation for the acquisition of the majority language. This, in turn, is expected to hinder school-related competence development.

Based on German data from the PISA 2000 study, Stanat (2006b) explored the extent to which the proportion of immigrant students not speaking German at home influences student achievement over and above the effects associated with family background characteristics (factors controlled for at the individual level: socio-economic status, parental level of education, material possessions, cultural resources, cultural activities, communicative practices, home language, and age at migration). In addition, students' basic cognitive abilities were controlled as a proxy for their prior knowledge at the time they entered the respective school⁶ (see also Baumert, Stanat & Watermann, 2006). Results of multi-level analyses initially seemed to support the hypothesis that a higher proportion of immigrant students speaking another language than the language of instruction at home is negatively related to competence development: reading achievement tended to be lower in schools with higher proportions of students whose home language is not German. The size of this effect was small, however. A one per cent increase in the proportion of students not speaking German at home was associated with a decrease in achievement of only half a point on the PISA scale. Even more importantly, further analyses indicated that the effect was not specifically tied to

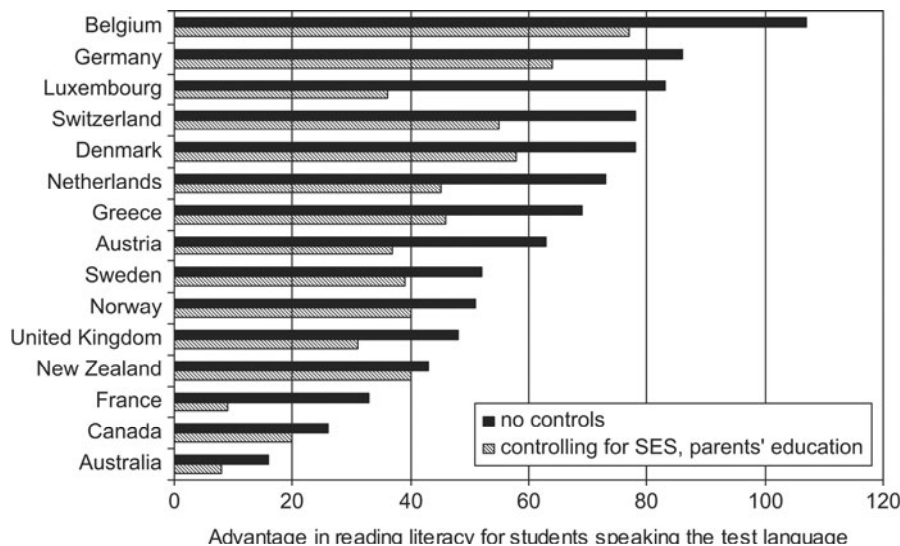


Figure 2. Regression coefficients from analyses with language spoken at home as the predictor (0=another language spoken at home, 1=test language spoken at home) and reading literacy as the criterion (PISA 2006).

the linguistic composition of the student body. In Germany, immigrant status and social status are typically confounded, such that the student body in schools with high proportions of immigrant students also tends to be socially disadvantaged. Controlling for the mean socio-economic status of students' families as an indicator for the social composition of the school reduced the effect of linguistic composition considerably. The social composition effect was significant indicating that, all other things equal, students tend to reach lower levels of achievement in schools with higher proportions of students from families with low socio-economic status (see also Bryk, Lee & Holland, 1993; Coleman, 1966; Sammons, Thomas & Mortimore, 1997; Zimmer & Toma, 2000). Further controlling for mean cognitive ability at the school level reduced both the linguistic and the social composition effects, such that they were no longer significant. The effect of the cognitive composition of the class, in contrast, was large. This suggests that the average level of prior knowledge within a school mediates the effects associated with the proportion of students not speaking German at home and the proportion of students from socially disadvantaged families.

According to Esser's (2006) argument, it should be especially detrimental for competence development of immigrants if students with the same linguistic background attend the same school. To explore this assumption, Walter and Stanat (2008) examined whether the proportion of students from particular immigrant groups in schools exerts an independent effect on students' reading achievement over and above family background factors controlled for at the individual level and the attended school type controlled for at the school level. No independent effect of

the proportion of students whose parents were born in the former Soviet Union was found. However, a negative effect of the proportion of students of Turkish descent emerged. Further controlling for the proportion of students speaking Turkish at home did not reduce this effect. This finding contradicts Esser's assumption that composition effects occur when immigrant students within a school or class share the same language background. After the mean socio-economic status and cognitive abilities within schools were also taken into account, the proportion of students of Turkish descent still had a significant effect on achievement, but only if it reached or exceeded 40 per cent.

Apart from the composition effect associated with high proportions of students from Turkish families, Stanat, Schwippert and Gröhlich (2010) largely confirmed these findings based on data from a large-scale longitudinal study which was carried out in Hamburg (Bos, et al., 2007). The unit of analysis in this investigation was the classroom. In the analysis, the authors also examined the impact of the proportion of immigrant students in classrooms on the development of reading competence. Reading competence was assessed at the end of grade four and again at the beginning of grade seven. In Hamburg, students change school after grade four as they transit from the comprehensive elementary school to the tracked system of secondary school. Thus, the study included an assessment of reading competence shortly before students entered the school in which they were tested again in grade seven. Results of multi-level analyses, controlling for prior achievement and family background factors at the student level (migration status, family language, socio-economic background, parents' level of education, material possessions, cultural resources, and cultural activities), showed that the proportion of students in a classroom speaking another language than German at home as well as the proportion of students having at least one foreign-born parent exerted negative effects on reading achievement. When these two composition effects were included in the analysis simultaneously, only the effect of the linguistic composition was significant. This finding seemed to confirm the language-related mechanisms proposed by Esser in predicting composition effects. However, the impact of the proportion of students not speaking German at home was considerably reduced when the families' socio-economic status at the class level was taken into account as well. After the average level of students' prior reading achievement in classrooms was also controlled, the linguistic composition effect was no longer significant. The size of the socio-economic composition effect decreased considerably as well, yet it continued to be significant.

To examine the assumption that the effect of linguistic composition should be most pronounced when immigrant students in a school share the same first language, the analyses were repeated focusing on the proportion of Turkish-speaking students. The same pattern of results emerged: no linguistic composition effects occurred after the socio-economic composition of the class and students' mean prior level of reading achievement were controlled.

Taken together, there is little evidence for the assumption that high proportions of immigrant students, students not speaking the language of instruction at home, or immigrant students speaking a particular language at home affect student achievement above and beyond the effects of social composition and average prior

achievement of the student body. Analyses of the PISA data and subsequent longitudinal analyses confirming the findings thus helped to qualify an assumption that is frequently made by educational research, educational policy and practice as well as by the general public.

Individual level: Family background and language spoken at home. The PISA data also allow for analyses exploring different individual and family-level explanations for immigrant students' disadvantages in competence development. PISA identified pronounced achievement disparities between immigrant students and students from native families (Baumert & Schümer, 2001). In the first PISA cycle, students whose parents were both born abroad scored more than two thirds of a standard deviation (73 points) below students whose parents were both born in Germany on the reading literacy scale. This disparity is equivalent to the learning development of approximately almost two school years. Controlling for the socio-economic status of their families and the students' duration of stay in Germany reduced the difference to 33 points, which still amounts to about one year of competence development. Only after the language spoken at home was also taken into account did the disparities between immigrant students and their native-language peers vanish. Thus, the family language seems to be an important determinant of immigrant students' disadvantage in achievement.

The results from more recent PISA cycles confirmed and further qualified these findings. Due to a national extension involving an oversampling of immigrant students, the German PISA 2003 data allow for analyses distinguishing between different immigrant groups. Walter, Stanat and Segeritz (in press) differentiated immigrants according to their country of origin (former Soviet Union, Turkey, Poland, former Yugoslavia, and Italy) and generation status (first generation immigrant students and second generation immigrant students). Regression analyses of the data revealed significant disadvantages in reading achievement⁷ for almost all immigrant students, the only exception being second generation immigrant students whose families came from the former Soviet Union or Poland (see [table 1](#), model I). The disadvantages were especially pronounced for immigrant students whose families had emigrated from Turkey, the former Yugoslavia, or Italy.

Comparisons of first generation and second generation immigrant students revealed very different patterns for the five immigrant groups. For students whose families emigrated from the former Soviet Union and the former Yugoslavia, the achievement gaps were considerably smaller in the second generation of immigrants than in the first generation. This indicates that structural assimilation processes in educational outcomes occurred (Alba & Nee, 2003; Esser, 2001 and 2006). For students whose families came from Turkey or Italy, in contrast, the achievement gap continues to be large in the second immigrant generation. These findings could potentially reflect segmented assimilation processes (Segeritz, Walter & Stanat, 2010). Unlike classical assimilation theory (e.g. Gordon, 1964) and new assimilation theory (e.g. Alba & Nee, 2003), which assume that immigrants continually adapt to the mainstream society of the receiving society over the course of generations, the theory of segmented assimilation (Portes & Rumbaut, 2001; Zhou, 1997) suggests that assimilation can occur into different segments of receiving societies. One path –

the so-called downward assimilation – leads to socially disadvantaged segments. According to this view, then, ethnic disparities between immigrants and members of the mainstream receiving society will not necessarily vanish over time but may also consolidate. The stagnation of some immigrant groups in Germany might reflect such a downward assimilation process. To determine whether this is in fact the case, however, would require data on third immigrant generation students which are not yet available in sufficient amounts.

In the attempt to identify determinants of disparities between immigrant students and students from native families, Walter, Stanat and Segeritz (in press) included various individual background variables into a series of regression models (see table 1).

Table 1. Results of regression analyses predicting students' reading achievement (PISA 2003)

	Model I		Model II		Model III		Model IV	
	B	(SE)	B	(SE)	B	(SE)	B	(SE)
Constant	519	(1)	519	(1)	515	(1)	528	(2)
Second generation								
Former Soviet Union	14	(15)	14	(15)	15	(16)	14	(17)
Turkey	-109	(7)	-90	(7)	-52	(7)	-50	(8)
Poland	-16	(10)	-9	(11)	0	(11)	-1	(10)
Former Yugoslavia	-70	(18)	-48	(17)	-19	(15)	-18	(16)
Italy	-110	(14)	-91	(13)	-49	(13)	-49	(14)
First generation								
Former Soviet Union	-57^a	(4)	-42^a	(4)	-9	(5)	-10	(5)
Turkey	-124	(11)	-98	(11)	-49	(12)	-47	(12)
Poland	-21	(8)	-7	(8)	14	(7)	14	(7)
Former Yugoslavia	-112	(13)	-85	(13)	-59	(14)	-54	(14)
Italy	-107	(25)	-81	(23)	-36	(24)	-43	(23)
Family language not German			-35	(5)	-34	(5)	-34	(5)
Socio-economic status ¹					26	(1)	22	(1)
Parents' level of education ²					10	(1)	8	(1)
Cultural possessions							13	(1)
Material possessions							3	(1)
Number of children in family							-5	(1)
R ²	0.075		0.079		0.179		0.196	

Note. Adapted from Walter, Stanat & Segeritz (in press).

¹: HISEI, z-standardized. ²: in years of education. ^a: significantly different from adolescents of the second immigrant generation ($p < .05$).

Coefficients significantly different from zero are in bold type ($p < .05$).

For the whole chapter: house style on T for table? Confirming results from the first PISA cycle, the language spoken at home emerged as a powerful predictor of reading achievement (model II). After the language spoken at home was controlled, the achievement disadvantages associated with an immigration background were reduced for 14 to 26 points on the PISA reading scale for all immigrant groups except for second generation students from the former Soviet Union and Poland. The introduction of families' socio-economic status and parents' level of education decreased the effects of students' immigrant background substantially (model III). Further taking cultural possessions, material possessions, and the number of children in the family into account did not add much to the explanation of the disparities observed between immigrant students and students from native families (model IV). When all background variables were considered simultaneously, the language spoken at home was the strongest single predictor of immigrant students' reading achievement. Again, this finding indicates that the learning opportunities for acquiring the language of instruction are crucial for the emergence of disadvantages associated with an immigration background (Stanat, 2008), suggesting that schools and other educational institutions need to strengthen their efforts aimed at developing students' proficiency in German as a second language (see below).

Individual level: Motivation and aspirations. Additional factors that have been shown to influence educational success are school-related motivational orientations and aspirations. The *immigrant optimism* hypothesis suggests that immigrants – especially when they were socio-economically disadvantaged in their country of origin – perceive migration as an opportunity for social advancement (Kao & Tienda, 1995). To the extent that they consider education important for reaching this goal, individuals who recently immigrated should display particularly high levels of school-related motivation and aspiration. In analyses of the immigrant students' school experiences and outcomes, these factors have rarely been explored. As the PISA data set contains comprehensive information on motivational orientations and aspirations, it is possible to take them into account.

Using data from the national extension of PISA 2003, Stanat, Segeritz and Christensen (2010) examined students' instrumental motivation for learning mathematics, their aspiration to complete tertiary education, and their occupational aspirations. The findings revealed general evidence for immigrant optimism. Given the same achievement level and socio-economic background, students of Turkish descent showed higher levels of instrumental motivation, were more likely to aim for tertiary education, and were reported to strive for an occupation with higher social prestige than students from native families. While students of Polish origin were not more motivated in mathematics, their educational and occupational aspirations were also higher than those of students from native families. For students whose families had emigrated from the former Soviet Union, finally, higher levels of instrumental motivation in mathematics as well higher aspirations were found to be most consistent in the first immigrant generation.

A similar pattern emerged in most other countries participating in the PISA study with substantial proportions of immigrant students. Given the same level of

achievement, immigrant students were often more motivated for learning mathematics as well as more likely to aspire for tertiary education and higher-status occupations than their peers from native families (Stanat et al., 2010). Thus, the achievement disadvantages of immigrant students do not seem to be due to a lack of motivation or aspirations. In fact, the positive motivational orientation of adolescents from immigrant families represents a resource on which schools can build in the attempt to reduce these students' disadvantages in achievement and educational success in general.

PISA'S LIMITATIONS: WHAT THE STUDY CANNOT TELL US

Methodological constraints of cross-sectional designs

Cross-sectional studies on student achievement like PISA can provide valuable information on the situation of immigrant students and on potential determinants of their educational success. However, cross-sectional designs are limited in several respects. Data from cross-sectional studies offer a snap-shot of individuals' learning outcomes, but they do not capture the processes that led up to these outcomes (Blossfeld, Schneider & Dollmann, 2009). Panel studies focusing on educational processes, in contrast, allow for analyses of developments over time. With such longitudinal data it is possible to explore how characteristics of learners and learning environments shape learning developments as well as how the effects of life-events and interventions unfold. Moreover, transition processes, which are crucial for educational success over the lifespan (Ditton & Krüsken, 2006; Maaz, Baumert, Gresch & McElvany, 2010), can best be studied with longitudinal research designs examining the situation before and after the transition.

A related shortcoming of cross-sectional designs consists in its limited potential for drawing causal inferences. Based on cross-sectional data, causality can only be established in terms of *Causation as Robust Dependence*, i.e., by way of controlling potentially confounded variables (Blalock, 1970). According to this approach, a causal relationship is assumed if a relationship between two variables persists even after additional variables considered relevant in this context were introduced in the statistical model. Yet, it is impossible to ensure that all relevant third variables have been taken into account. Therefore, inferences about causality can only be drawn provisionally from such findings (Shadish, Cook & Campbell, 2002; Goldthorpe, 2001). Experimental designs with random assignment to treatment and control groups are generally regarded as the best approach to establishing causal relationships (e.g., Holland, 1986; Rubin, 1974). However, for practical or ethical reasons, randomization is often impossible in educational research. Therefore, several statistical techniques have been developed that allow for causal inferences based on cross-sectional data. For instance, *propensity score matching* (Rosenbaum & Rubin, 1983; Pearl, 2009) or *regression discontinuity analyses* (Cook & Campbell, 1979; Thistlethwaite & Campbell, 1960) aim at providing unbiased estimates of treatment effects even when the group exposed to the treatment and the control group differ in terms of relevant covariates. Thus, cross-sectional designs may produce valuable indications for potential causal

relationships. To the extent possible, however, the findings from these analyses should be confirmed with longitudinal and experimental studies (Blossfeld, 2009; Blossfeld, et al., 2009; Boruch, de Moya & Snyder, 2002). As pointed out above, some of the most important findings on the educational situation of immigrant students in Germany identified with the PISA data could be replicated in longitudinal analyses.

Improving immigrant students' achievement: Measures at the teaching and learning level

Student achievement studies such as PISA aim primarily at identifying relative strengths and weaknesses of school systems. They are, however, limited in helping to determine what can and should be done to improve teaching and learning processes to overcome observed problems. Accordingly, PISA does not tell us how effective efforts aimed at improving the achievement of immigrant students should be designed. Answering this question requires empirical evidence from different types of studies. Approaches to supporting immigrant students need to be developed on the basis of theories and evidence on teaching and learning processes. Subsequently, these approaches need to be operationalized and tested in the field. To gain reliable information on their effectiveness, finally, randomized field trials are the preferred method as they combine high levels of experimental control with high levels of external validity (Shadish, Cook & Campell, 2002).

As pointed out above, an important target point for interventions aimed at improving the educational situation of immigrant students seems to be the development of students' proficiency in German as a second language. Reading literacy in the language of instruction proved to be an important predictor of attending a higher track in secondary schooling. Furthermore, the learning opportunities for German language acquisition in students' families seem to affect their achievement in reading as well as in other domains. Therefore, educational institutions need to provide high-quality support for immigrant students with limited proficiency in the language of instruction. Which approach to second language teaching and learning will be most promising, however, is largely unclear.

Most international research addressing this issue in the past has focused on comparing bilingual and transitional programs on the one hand with submersion and immersion approaches on the other hand. While bilingual and transitional programs provide – either permanently or temporarily – instruction to immigrant students in their native language (L1) as well as in the language of instruction (L2), submersion and immersion approaches use only the L2 in instruction (with or without additional support in L2). The studies that have been carried out to explore the relative effectiveness of these approaches typically have serious methodological limitations that impair their interpretability (Limbird & Stanat, 2006; Rossell & Kuder, 2005; Slavin & Cheung, 2005; Söhn, 2005). The evidence is mixed and allows only the tentative conclusion that bilingual instruction has neither negative effects nor particularly positive effects on L2 development (Limbird & Stanat, 2006; Rossell & Baker, 1996; Söhn, 2005).

In Germany, as well as in most other OECD countries, bilingual support programs for immigrant students are rare (Stanat & Christensen, 2006). By far the most widely applied approach is immersion/submersion with systematic support in L2. Currently, a multitude of monolingual support programs aimed at promoting immigrant students' second language competence are implemented in German schools. Yet, little is known about the effectiveness of the different approaches and their underlying mechanisms (Limbird & Stanat, 2006; Redder et al., 2010). Although the situation of immigrant students in Germany seems to be improving somewhat over time, they continue to be highly disadvantaged in terms of educational achievement and attainment. Therefore, systematic intervention studies examining the processes and effects of different monolingual support programs are urgently needed.

A first intervention study exploring the effects of second language support was the *Jacobs Summer Camp Project* which compared the outcomes of two monolingual approaches to L2 teaching and learning with a randomized field trial. Immigrant students who had recently finished third grade attended the programs for three weeks during the summer break. The literature on foreign language teaching and learning distinguishes between implicit and explicit approaches to instruction (DeKeyser, 2003; Hulstijn, 2005; Nunan, 1999). While implicit approaches (or "Focus on Meaning") emanate from the assumption that learners acquire language competences without a deliberate focus on rules, explicit approaches (or "Focus on Form(s)") draw the learners' attention to language structures and rules. In the summer camp, one of the two implemented programs was designed to promote language learning implicitly, by engaging students in active and meaningful communication. This approach was operationalized with improvisational theatre activities in which students participated in the mornings and in the afternoons. The second approach involved explicit language learning. In this program, the children also participated in the theatre activities in the afternoons, yet they received systematic grammar instruction in the afternoons.

At the simplest level, it can be corrected by saying "... In the afternoons, but they also received...." However I simply do not know if this is true, in actuality: in the experiments and research practice itself. The baseline group included in the study, finally, did not attend the summer camp at all.

The results of the Jacobs Summer Camp Project indicate that the combination of implicit and explicit language support was more effective than the purely implicit approach. Students who attended the combined program outperformed students of the untreated baseline group in grammar and reading at the beginning of fourth grade. Students who received the implicit program only, in contrast, did not significantly differ in their L2 performance from the baseline group. In the direct comparison, students with explicit support outperformed students with implicit support in grammar but not in reading and vocabulary. After three months, the advantage of the explicit approach was still visible in terms of effect sizes, yet the differences between the treatment group with explicit support and the baseline group reached significance only in reading, and no significant differences in the outcomes of the implicit and explicit support emerged (Stanat, Baumert & Müller,

2005; Stanat, Baumert & Müller, 2008; Stanat, Becker, Baumert, Lüdtke & Eckhardt, submitted).

Building on the Jacobs Summer Camp Project, another project is currently carried out. It aims at further developing the two approaches to language support in L2 and to test its effectiveness as it is implemented in school settings over a longer period of time (Rösch & Stanat, in press).

CONCLUSION

As illustrated in this chapter, international large-scale assessment studies on student achievement, such as PISA, offer valuable information on the status of immigrant students' educational integration. The data allow for comparisons of this situation across countries participating in the studies. Moreover, the conditions of immigrant students' educational success can be explored at various levels of analyses, helping to identify possible target points for interventions. However, at the level of teaching and learning, the analytical potential of large-scale assessments is limited. As proficiency in the language of instruction seems crucial for immigrant students' educational success, further high-quality studies are needed that examine the effectiveness of different approaches to second language support.

Another important feature of PISA is that assessments are carried out every three years. This allows for estimations of the extent to which outcomes of school systems have changed over time. Each cycle of the project focuses on one of the three assessments domains (reading, mathematics, and science) which is measured more comprehensively than the other two. The major domains were reading in 2000, mathematics in 2003, and science in 2006. In 2009, the project came full circle and the data can be expected to yield a reliable estimate of the extent to which school systems were able to improve achievement levels and to reduce achievement gaps over the last nine years. The publication of these results is scheduled for December 2010.

NOTES

¹ In this article, the term "immigrant" refers to persons who themselves immigrated to the receiving country (first generation) as well as to persons whose parents immigrated but who themselves were born in the receiving country (second generation). The new *Microcensus* (see below) also identifies the third generation as immigrants (persons whose grandparents immigrated but who themselves and whose parents were born in the receiving country). In the existing PISA data sets, however, this group is too small to be analysed separately.

² OR = odds ratio.

³ Adapted from Stanat, 2006a, p. 101

⁴ The finding that second generation immigrant students reached lower test scores than first generation immigrant students does not imply that the situation of immigrants in Germany is getting worse across generations. Instead, the pattern is largely due to the composition of the two immigrant groups. The second generation includes many children of *Guest Workers* who are predominantly of Turkish descent and tend to be particularly disadvantaged in terms of socio-economic background,

A. EDELE AND P. STANAT

- education, and achievement. The first generation, in contrast, largely consists of *Aussiedler* whose situation tends to be more favourable overall (e.g., Segeritz, Walter, & Stanat, 2010).
- ⁵ The longer a family has lived in a country, the more its socio-economic and educational situation should also be affected by the country's integration policies and practices. Thus, depending on how long ago the immigration took place, controlling for socio-economic status and educational background will to some extent also capture these effects. Such an analysis is therefore prone to overestimate the impact of immigration policies and to underestimate the impact of integration policies.
 - ⁶ As a cross-sectional study, the PISA data set does not contain information on students' prior achievement at the time they entered the school. This information, however, is necessary to estimate the effects of school-level factors on students' learning outcomes over time. Using basic cognitive ability as a proxy for prior achievement is based on the assumption that it is a good predictor of learning development and highly stable over time. Yet, basic cognitive ability is certainly also affected by schooling (Ceci, 1991) so that controlling for this factor should result in an underestimation of composition effects. Results from longitudinal analyses reported by Baumert, Stanat and Watermann (2006), however, indicate that the degree of underestimation is negligible.
 - ⁷ The results for mathematics achievement are very similar to those for reading achievement.

REFERENCES

- Alba, R., & Nee, V. (2003). *Remaking the American Mainstream*. Cambridge, Mass./London: Harvard University Press.
- Baumert, J., Blum, W., & Neubrand, M. (2004). Drawing the lessons from PISA 2000 - Long term research implications: Gaining a better understanding of the relationship between system inputs and learning outcomes by assessing instructional and learning processes as mediating factors. *Zeitschrift für Erziehungswissenschaft*, 7, 143–157.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb [Family background, educational participation, and competence acquisition]. In Deutsches PISA-Konsortium (Ed.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323–410). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus [School structure and the formation of differential learning and developmental environments]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden.
- Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. Padilla (Ed.), *Acculturation: Theory, models, and some new findings* (pp. 9–25). Boulder, CO: Westview.
- Blalock, H. M. (1970). *Causal models in the social sciences*. Chicago: Aldine.
- Blossfeld, H.-P. (2009). Causation as a Generative Process. The Elaboration of an Idea for the Social Sciences and an Application to an Analysis of an Interdependent Dynamic Social System. In H. Engelhardt-Wölfler, H.-P. Kohler, & A. Fürnkranz-Prskawetz (Eds.), *Causal analysis in population studies: Concepts, methods, applications* (pp. 83–109). Berlin: Springer.
- Blossfeld, H.-P., Schneider, T., & Dollmann, J. (2009). Methodological advantages of panel studies. *Journal for Educational Research Online*, 1(1), 20–30.
- Boruch, R., de Moya, D. & Snyder, B. (2002). The importance of randomized field trials in education and related areas. In R. Boruch and F. Mosteller (Eds.), *Evidence Matters. Randomized field trials in education research*. Washington, DC: Brookings Institution Press. No page numbers.
- Bos, W., Bensen, M., Gröhlich, C., Guill, K., May, P. & Rau, A. (2007). *KESS 7 - Kompetenzen und Einstellungen von Schülerinnen und Schülern, Jahrgangsstufe 7 [KESS 7 – Competences and attitudes of seventh grade students]*. Hamburg. Technically, no Publisher is listed.

PISA'S POTENTIAL FOR ANALYSES OF IMMIGRANT STUDENTS' EDUCATIONAL SUCCESS

- Bourhis, R. Y., Moise, L. C., Perreault, S., & Senécal, S. (1997). Towards an interactive acculturation model: A social psychological approach. *International Journal of Psychology*, 32(6), 369–386.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge: Harvard University Press.
- Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.
- Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27(5), 703–722.
- Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis for field settings*. Chicago, IL: Rand McNally.
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.
- Diefenbach, H. (2002). Bildungsbeteiligung und Berufseinmündung von Kindern und Jugendlichen aus Migrantenfamilien: Eine Fortschreibung der Daten des Sozio-ökonomischen Panels [Educational participation and transition to work life of children and adolescents from immigrant families: A follow-up of the Socio-Economic Panel data]. In Sechster Kinder- und Jugendbericht (Ed.), *Migration und die europäische Integration: Herausforderungen für Kinder- und Jugendhilfe* (pp. 9–70). München: Verlag Deutsches Jugendinstitut.
- Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I [The transition from primary to secondary schools]. *Zeitschrift für Erziehungswissenschaft*, 9, 348–372.
- Esser, H. (2001). *Integration und ethnische Schichtung. MZES Arbeitspapiere 40 [Integration and ethnic stratification. MZES working paper 40]*. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten [Language and integration: The social conditions and consequences of immigrants' language acquisition]*. Frankfurt am Main: Campus Verlag.
- Goldthorpe, J. H. (2001). Causation, statistics, and sociology. *European Sociological Review*, 17(1), 1–20.
- Gomolla, M., & Radtke, F.-O. (2002). *Institutionelle Diskriminierung: Die Herstellung ethnischer Differenz in der Schule [Institutional discrimination: The generation of ethnic differences in school]*. Opladen: Leske + Budrich.
- Gordon, M. (1964). *Assimilation in American life: The role of race, religion, and national origins*. New York: Oxford University Press.
- Gresch, C., & Becker, M. (2010). Sozial- und leistungsbedingte Disparitäten im Übergangverhalten bei türkischstämmigen Kindern und Kindern aus (Spät-)Aussiedlerfamilien [Social and achievement related disparities in school transition of immigrant students from Turkish and *Aussiedler* families]. Hans-Goerg to check this translation please! In K. Maaz, J. Baumert, C. Gresch, & N. McElvany (Eds.), *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (pp. 181–200). Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Hattie, J. A. C. (2001). Classroom composition and peer effects. *International Journal of Educational Research*, 27, 449–481.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
- Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition*, 27(2), 129–140.
- Kristen, C., & Dollmann, J. (2009). Sekundäre Effekte der ethnischen Herkunft: Kinder aus türkischen Familien am ersten Bildungsübergang [Secondary effects of ethnic origin: Children from Turkish families at the first school transition]. In J. Baumert, K. Maaz & U. Trautwein (Eds.), *Bildungsentscheidungen* (pp. 205–229). Wiesbaden: VS Verlag.

- Limbird, C., & Stanat, P. (2006). Sprachförderung bei Schülerinnen und Schülern mit Migrationshintergrund: Ansätze und ihre Wirksamkeit [Language support for immigrant students: Approaches and their effectiveness]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 257–307). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lohmann, H., Spieß, C.K., Groh-Samberg, O., & Schupp, J. (2008). Analysepotenziale des Sozio-ökonomischen Panels (SOEP) für die empirische Bildungsforschung [Analytical potentials of the German Socio-Economic Panel (GSOEP) for empirical educational research]. *Zeitschrift für Erziehungswissenschaft*, 12(2), 252–283.
- Maaz, K., Baumert, J., Gresch, C., & McElvany, N. (Eds.). (2010). *Der Übergang von der Grundschule in die weiterführende Schule: Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten [The transition from primary to secondary school: Achievement-based equity and regional, social, and ethnic disparities]*. Bonn: Bundesministerium für Bildung und Forschung, Referat Bildungsforschung.
- Müller, A. G., & Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Jugendlichen aus der ehemaligen Sowjetunion und der Türkei [Academic success of immigrant students: Analyzing the situation of adolescents from the former Soviet Union and Turkey]. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 221–256). Wiesbaden: VS Verlag.
- Nelson, T. D. (Ed.). (2002). *The psychology of prejudice*. Needham Heights, MA: Allyn & Bacon. Italics needed for the book title.
- Nunan, D. (1999). *Second language teaching and learning*. Boston: Heinle, & Heinle.
- OECD (Ed.). (2007). *PISA 2006 Volume 2: Data*: OECD. No place of publication. Probably insert: Paris.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press
- Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. University of California Press, Berkeley.
- Redder, A., Schwippert, K., Hasselhorn, M., Forschner, S., Fickermann, D. & Ehlich, K. (2010). Grundzüge eines nationalen Forschungsprogramms zu Sprachdiagnostik und Sprachförderung [Main features of a national research program on language diagnostics and language support]. *ZUSE-Diskussionspapier Nr. 1*. Hamburg: Hamburger Zentrum zur Unterstützung der wissenschaftlichen Belreitung und Erforschung schulischer Entwicklungsprozesse (ZUSE).
- Rösch, H. & Stanat, P. (submitted). Bedeutung und Form (BeFo): Fachbezogene und sprachsystematische Förderung in Deutsch als Zweitsprache [Meaning and form (BeFo): Subject-integrated and language-systematic support in German as a second language].
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for the causal effects. *Biometrika*, 70(1), 41–55.
- Rossell, C. H. & Baker, K. (1996). The effectiveness of bilingual education. *Research in the teaching of English*, 30, 7–74.
- Rossell, C. H., & Kuder, J. (2005). *Meta-Murky: A rebuttal to recent meta-analysis of bilingual education*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688–701.
- Sammons, P., Thomas, S. & Mortimore, P. (1997). *Forging links: Effective schools and effective departments*. London: Paul Chapman.
- Schofield, J. W. (2006). *Migrationshintergrund, Minderheitenzugehörigkeit und Bildungserfolg. Forschungsergebnisse der pädagogischen, Entwicklungs- und Sozialpsychologie [Migration background, minority group membership and academic success. Research results from educational, developmental and social psychology]*. Berlin: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI), Wissenschaftszentrum Berlin für Sozialforschung (WZB).

- Segeritz, M., Walter, O. & Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? [Patterns of educational success of immigrant students in Germany: Evidence of segmented assimilation?]. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62, 113–138.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Slavin, R. E. & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, 75(2), 247–284.
- Söhn, J. (2005). Zweisprachiger Schulunterricht für Migrantenkinder. Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg [Bilingual instruction for immigrant children: Results of evaluations of its effects on second language acquisition and school success]. Berlin: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI), Wissenschaftszentrum Berlin für Sozialforschung (WZB).
- Stanat, P. (2006a). Disparitäten im schulischen Erfolg: Analysen zur Rolle des Migrationshintergrunds [Disparities in school success: Research on the role of immigration background]. *Unterrichtswissenschaft*, 34(2), 98–124.
- Stanat, P. (2006b). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft [School achievement of adolescents with immigrant background: the role of the student body composition]. In J. Baumert, P. Stanat, & R. Watermann (Hrsg.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*. (pp. 189–219). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stanat, P. (2008). Heranwachsende mit Migrationshintergrund im deutschen Bildungswesen [Adolescents with immigration background in the German education system]. In Cortina, K. S.; Baumert, J.; Leschinsky, A.; Mayer, K. U., & Trommer, L. (Hg.) *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick* (pp. 691–705). Hamburg: Rowohlt.
- Stanat, P., Baumert, J. & Müller, A. G. (2005). Förderung von deutscher Sprachkompetenz bei Kindern aus zugewanderten und sozial benachteiligten Familien. Evaluationskonzept für das Jacobs-Sommercamp Projekt [Supporting German competences in students from immigrant and socially disadvantaged families. The evaluation concept for the Jacobs Summer Camp Project]. *Zeitschrift für Pädagogik*, 51(6), 856–875.
- Stanat, P., Baumert, J. & Müller, A. G. (2008). Das Jacobs Sommercamp Projekt in Bremen [The Jacobs Summer Camp in Bremen]. In A. Ballis, & K. H. Spinner (Eds.), *Sommerschule - Sommerkurse - Summer Learning. Deutsch lernen in außerschulischem Kontext* (pp. 14–32). Hohengehren: Schneider Verlag.
- Stanat, P., Becker, M., Baumert, J., Lüdtke, O. & Eckhardt, A. G. (submitted). *Effects of implicit and explicit language support for second-language learners: Results of a randomized trial*.
- Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performances and engagement in PISA 2003*. Paris: OECD.
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb. Längsschnittliche Überprüfung eines umstrittenen Effekts [The effect of the immigrant proportion in class on competence development. A longitudinal examination of a disputed effect]. *Migration, Identität, Sprache und Bildungserfolg. Zeitschrift für Pädagogik*. 55. Beiheft, pp. 147–165.
- Stanat, P., Segeritz, M. & Christensen, G. (2010). Schulbezogene Motivation und Aspiration von Schülerinnen und Schülern mit Migrationshintergrund [School-related motivation and aspiration of students with immigration background]. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (pp. 27–54). Münster: Waxmann.
- Statistisches Bundesamt (Ed.). (2001). *Bildung und Kultur [Education and Culture]. Schuljahr 2000/2001*. Stuttgart: Statistisches Bundesamt.

A. EDELE AND P. STANAT

- Statistisches Bundesamt (Ed.) (2006). *Strukturdaten zur Migration in Deutschland [Structural data on migration in Germany]*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (Ed.). (2009). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2007 [Population and employment: The population with immigration background - Results of the Microcensus 2007]*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (Ed.). (2010). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2009 [Population and employment: The population with immigrant background - Results of the Microcensus 2009]*. Wiesbaden: Statistisches Bundesamt.
- Thistlethwaite, D. L. & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309–317.
- United Nations Department of Economic and Social Affairs (UN DESA) (2009). *International Migration 2009*. Retrieved 19.10.2010. from http://www.un.org/esa/population/publications/2009Migration_Chart/ittmig_wallchart09.pdf.
- Walter, O. & Stanat, P. (2008). Der Zusammenhang des Migrantenanteils in Schulen mit der Lesekompetenz: Differenziertere Analysen der erweiterten Migrantenstichprobe von PISA 2003 [The relationship between the proportion of immigrant students and reading competence: In-depth analyses of the extended immigrant sample in PISA 2003]. *Zeitschrift für Erziehungswissenschaft*, 11(1), 84–105.
- Walter, O., Stanat, P. & Segeritz, M. (in press). Schulbezogene Kompetenzen, motivationale Orientierungen und Bildungsaspirationen von Jugendlichen mit Migrationshintergrund in Deutschland: Differenzierte Analysen zu PISA 2003 [School-related competences, motivational orientation and educational aspirations of adolescents with immigration background in Germany: In-depth analyses on PISA 2003]. In W. Lauterbach, & R. Becker (Eds.), *Integration durch Bildung? Bildungserwerb von jungen Migranten in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Walter, O. & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland: Ein Vergleich mit ausgewählten OECD-Staaten [Achievement and attitudes towards education in adolescents with immigration background in Germany: A comparison with selected OECD-countries]. In PISA-Konsortium Deutschland (Ed.), *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 337–366). Münster: Waxmann.
- Zhou, M. (1997). Segmented assimilation: Issues, controversies, and recent research on the new second generation. *International Migration Review*, 31, 975–1008.
- Zimmer, R. W. & Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, 19(1), 75–92.

Aileen Edele
Department of Education and Psychology
Freie Universität Berlin

Petra Stanat
Institute for Educational Quality Improvement
Humboldt-Universität zu Berlin

JULIO CARABAÑA

**WHY DO THE RESULTS OF IMMIGRANT STUDENTS
DEPEND SO MUCH ON THEIR COUNTRY OF ORIGIN
AND SO LITTLE ON THEIR COUNTRY OF
DESTINATION?**

INTRODUCTION

In spite of the many criticisms to which PISA can and should be legitimately subjected, it is difficult to deny that this large and highly successful OECD project is a blessing for research in Comparative Education, as it provides strictly comparable data on inputs, processes and results for most educational systems in the world. The outcomes of analyzing these enormous data sets are rather disappointing for those who had expected from it quick and solid adjudications among rival theories. First the PISA reports, and then the many reanalyses carried out on the PISA data, have found only very weak relations between students' outcomes and characteristics of educational systems which are usually thought of paramount importance (Carabaña, 2008). It is from these tenuous links that multilevel regressions extract cross-country coefficients that, when statistically significant, are quite hard to interpret even by the best willed officials and scholars.

I will here address the question of the effects that schools in the destination countries have on the academic results of immigrant's children. Immigration countries have generally better schools and better results than countries of emigration. It is therefore easy to imagine how the desire to improve the schooling conditions of their children might be one important pulling factor for emigrants. However, against all expectations of subjects and observers, immigrant students score in the PISA tests rather like students in their countries of origin than like native students in the countries of destination. In spite of the allegedly better schools of the host countries, the learning of the newcomers remains at the level of their origin countries not only in the first generation, but also in the second and even later on. The question I will discuss here is: *why do the results of immigrant students depend so much on their country of origin and so little on their country of destination?*

Since the best theorists of scientific research insist upon the importance of firmly establishing the facts, this introduction will be followed by three parts. The first one consists of some considerations about research designs, the second deals with the fact that the learning outcomes of immigrant students are very similar to the learning outcomes of their con-nationals in their countries of origin, and the third part will try to explain this fact, or rather, will be an attempt to look for hypotheses not obviously at odds with the fact in question.

A PLEA FOR DETAILED, COMPARATIVE RESEARCH DESIGNS

By producing and giving easy access to an unprecedented amount of strictly comparable data on the most important world educational systems, PISA has opened the (almost entirely new) possibility of carrying out research about immigrant students using a design of the type ‘one origin-various destinations’.

It must be said that this type of research depends on the singular coding of the countries immigrant students and their parents come from, and that this information seems to have been of rather marginal interest for PISA managers. Participant countries were only obliged to collect the information needed to construct the category ‘immigrant’, and many of them (like Spain, France, Canada and the USA) disregarded country names, just noting ‘foreign country’. It is, thus, not thanks to the central PISA project, but to PISA country teams, that information about countries of origin has been collected and preserved.

Differences in academic achievement and attainment between groups defined by their national origins (sometimes called ‘ethnic groups’) have been a subject of scientific inquiry for years, first in the United States and later in Europe. To mention just a few, there are European studies on the difference between South Asians and Mahgrebis in France (Costa-Lascoux, 1996), on the difference between Asians and Latinos in Sweden (Jonsson, 2002), on the difference between South Europeans and Mahgrebis in The Netherlands (Lindo, 2000), on the difference between Jews and other groups in Hungary (Karady, 1987), on the difference between Askhenazis, Arabs and Sefardis in Israel (Shavit, 1990), etc. Even in Spain, where immigration is rather recent, albeit intense, such studies have been produced (Siguán, 1998; Siguán, 2003). Most of these studies assume that natives and immigrants should show equal academic achievement; upon finding that immigrants learn less than natives, they try to explain this difference first through socioeconomic factors and then relating the residual, if any, to attitudes, mores, folkways and other cultural characteristics. For instance, Lindo relates the low achievement of Moroccan children in the Netherlands to the fact that ethnic traditions press their mothers to stay at home, and so they develop weak social relations (Lindo, 2000:16).

All these studies share *per force* one and the same comparative design of the type ‘various origins-one destination’, for which Table 1 could be an example. PISA opens unique opportunities for extending this type of design to several countries, taking advantage of common methods and measures, instead of, as before, having to pick up one study after another. This way of extending a research design is in the spirit of comparative research that takes units of analysis case by case, looks for similarities and differences among them and finally groups them into types.

There is, however, another way of making use of large data sets, that is to aggregate them and to estimate synthetic parameters (means, regression coefficients) by way of statistical methods (of data synthesis, rather than of data analysis), thus considering countries as a sample. This has been the way PISA official reports (of the highest quality in this genre, it must be said) have been composed, in line with hegemonic trends in educational research that also prevail in most academic reanalyses of the PISA data.

WHY DO RESULTS OF IMMIGRANT STUDENTS

This dominance of the statistical methodology helps to explain the disdain for registering the names of singular countries in favour of directly coding them into more general and abstract categories. It also helps to explain why part of the potential of PISA data for comparative research still remains unexplored.

Table 1. A 'various origins-one destination design'. Regression of Pisa math scores on zones of origin and sociodemographic variables. Belgium, Pisa 2003

	MODEL 1		MODEL 2	
	B	SE	B	SE
FRENCH SB	-42,94	2,21	-42,95	2,05
GERMAN SB	-31,59	13,97	-31,15	12,96
POBELGME	-24,27	4,69	-23,07	4,36
POHOLAND	-27,22	17,85 ^{ns}	-5,86	16,93 ^{ns}
POFRANCI	-40,20	12,06	-18,36	11,56 ^{ns}
POEUROOR	7,82	21,06 ^{ns}	30,72	19,81 ^{ns}
POTURQUI	-111,05	9,49	-60,60	9,36
POMAGR	-85,00	9,00	-36,09	8,68
POAFRI	-68,95	12,62	-53,61	11,76
POOTR1	-39,75	8,38	-10,55	7,99 ^{ns}
POOTR2	-64,19	9,74	-40,29	9,26
GEN1	-21,16	8,28	29,66	12,33
FISCED			3,73	0,73
HISEI			1,84	0,07
LEX			-26,23	6,18
EDADINM			-5,89	1,17
(Constant)	574,13	1,38	449,77	4,43
Rsquared	0,11		0,24	

SOURCE: PISA 2003 DATA SET. ns=<.05

Variable labels: French SB: 'French -speaking Belgium; German SB: German speaking Belgium; Pobelgme: Country of Origin Belgium, foreign mother; Poholand: Country of Origin Netherlands; Pofrance: idem France; Poeuroor: idem East Europe; Poturkey: idem Turkey; Pomagr: idem Maghreb; Poafri: idem Africa; Pootr1: idem other Europe; Pootr2: idem other; Gen1: first generation; Fisced: Father's Educational Level; Hisei: highest socioeconomic index (father or mother); Lex: foreign language spoken at home; Edadinm: age at immigration.

Certainly, PISA reports the differences between natives and immigrants in *several countries*, making clear how much they vary from one country to another, from null in Australia to more than one standard deviation in Central Europe. Moreover, PISA also reports that, within each country, the results for different immigrant groups vary considerably. "For example, in New Zealand, immigrant students from Samoa demonstrate significantly lower scores than their native peers (by 81 score points), while there are no significant performance disadvantages for immigrant students from the United Kingdom or China" (OECD, 2006, p. 53).

But instead of further investigating the causes of these differences between countries, PISA leaves aside this diversity of particulars in favour of a global statistical approach where students from all countries are ruthlessly amassed into only two categories of ‘immigrants’: first and second generation. So, in the Report just mentioned above (a monograph called *Where immigrant students succeed*), the central question is the difference in the mean score of immigrant and native students, ignoring the difference in differences finding, i.e., the fact of inter-immigrant differences according to their origins. Its explanation is offered through a set of variables (first and second generation, parental education in years of schooling, father’s occupational status, foreign language spoken at home and age at immigration) which completely ignore the fact of inter-immigrant differences, thus reducing singularities to abstract variables for the sake of statistical generalization. The results (OECD 2006, Table 3.5) are that, indeed, such variables do reduce the gap between natives and immigrants of first and second generation, but in proportions depending on the country of destination. (Table 2 offers an illustration, to be compared with Table 1).

Table 2. A design conflating countries of origin into two immigrant categories. Belgium, Pisa math scores, 2003

	MODEL 1		MODEL 2	
	B	SE	B	SE
GEN1	-82,36	6,04	-6,92	12,9 ^{ns}
GEN2	-80,45	4,90	-45,03	5,03
FISCED			3,18	0,75
HISEI			1,84	0,07
LEX			-28,90	6,09
EDADINM			-5,7	1,18
(Constant)	557,00	1,12	434,42	4,5
R squared	0,06		0,18	

Source: PISA 2003 DATA SET.

Variable labels: Gen1: first generation; Gen2: second generation; Fisced: Father’s Educational Level; Hisei: highest socioeconomic index (father or mother).Lex: Foreign language spoken at home. Edadinm: age at immigration.

Academic papers analyzing PISA data share this tendency towards what Wright Mills called ‘abstract empiricism’, driven by econometric modelling and statistical packages. So, one that comes under the title ‘Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries’ (Marks, 2005) does in fact calculate the mean coefficients of socioeconomic factors, sociocultural factors and school variables across all the twenty countries. But perhaps the best available example of this ‘path-dependent blindness’ is a series of papers by Dronkers and others. All of them address the question from the point of view of the *countries of origin* (Levels & Dronkers (2008); Dronkers & Levels

WHY DO RESULTS OF IMMIGRANT STUDENTS

(2007); Levels, Dronkers & Kraaykamp (2008); Heus, Dronkers & Levels (2009); Heus & Dronkers (2008), Heus & Dronkers (2009), but only the very last one (Dronkers & de Heus, 2008) looks at it from the point of view of each *singular country* of origin.

Table 3. Showing how Close PISA Has Come to an ‘One Origin-Several Destinations’ Design Comparison of Performance Levels for Immigrant Students whose Families Came from Turkey and the Former Yugoslavia

	Turkey		Former Yugoslavia		Native Students
	N	Mean Score	N	Mean Score	Mean Score
<i>Austria</i>	141	423	276	456	515
<i>Belgium</i>	140	421			546
<i>Denmark</i>	53	424			520
<i>Germany</i>	197	405	45	448	525
<i>Luxembourg</i>			92	421	507
<i>Switzerland</i>	142	436	408	460	543

Note: All differences between native and immigrant students are statistically significant. Source: OECD 2006, [Table 2.9](#).

Table 2.9 (OECD, 2006). Comparison of Performance Levels for Immigrants Students Families Came from Turkey and the former Yugoslavia

	Immigrants Students								Native Students	Difference in Mathematics Performance Between native Student and Turkish Immigrant students	Difference in Mathematics Performance between native Student and Immigrant Students from the Former Yugoslavia			
	Turkey				Former Yugoslavia									
	Participating Students	Performance on the Mathematics Scale	Participating Students	Performance on the Mathematics Scale	Participating Students	Performance on the Mathematics Scale	Participating Students	Performance on the Mathematics Scale						
	N	Weighted %	Mean Score	SE.	N	Weighted %	Mean Score	S.E	Mean Score	S.E	Diff. er.	SE	Diff. er.	SE
Austria	141	25,9	423	(8.9)	276	47,2	456	(6.7)	515	(3.3)	-92	(9.1)	-59	(7.6)
Belgium	140	14,8	421	(13.1)	C	C	c	c	546	(2.5)	-125	(12.8)	c	c
Denmark	53	32,1	424	(12.4)	c	C	c	c	520	(2.5)	-95	(12.3)	c	c
Germany	197	32,1	405	(10.8)	45	7.0	448	(17.0)	525	(3.5)	-120	(11.6)	-78	(17.0)
Luxembourg		c	c	c	92	7.3	421	(10.2)	507	(1.3)	c	c	-86	(10.2)
Switzerland	142	8,5	436	(10.4)	408	23.0	460	(7.3)	543	(3.3)	-106	(10.3)	-82	(8.0)

As a result of this ‘econometric’ bias, PISA reports leave aside designs of the type ‘one origin, various destinations’ even if they have come very close to them. That is what happened in *Where immigrant students succeed* (OECD, 2006), which presents evidence on the similarity of results of Turkish and Yugoslavia emigrants to different countries of Central Europe (Table 3), but stops short of comparing their scores with those of their non-emigrant conationals.

Perhaps symptomatic of some confusion is that the report on immigrants reads these data in quite a different way than the report on the 2006 science study

According to the first, both groups, Turks and Yugoslavians, have significantly lower scores than natives in the host countries and their mean scores are fairly similar across countries (OECD, 2006, p. 53). Whereas according to the second report:

“Nor can they [the relative performance levels of students with an immigrant background, J. C.] *be attributed solely to the country of origin* [My italics, J. C.]: for example, a more detailed analysis of the PISA survey shows that immigrant students from Turkey performed 31 points better in mathematics in Switzerland than they did in the neighbouring country Germany “. (OECD, 2008, p. 179).

Let us, therefore, follow this path of inquiry leading to origin countries that has been abandoned, in my view rather precipitously, in PISA reports.

THE FACTS, OR *EXPLANANDA*

When the country of emigration has participated in PISA, a comparison of emigrants with non-emigrants becomes feasible for several countries. With some exceptions, emigrants reproduce the PISA scores of their aboriginal counterparts wherever they go. Table 4 presents scores extracted from PISA 2003, and Table 5 from PISA 2006. Considered are only countries of origin participating in PISA whose names have been coded in the countries of destination.

Similarities in scores by country of origin are striking by simple observation. No less stunning is to observe the distances between natives and immigrants in destination countries. Let us briefly examine how statistical analyses confirm the similarities by country of origin. Assuming simple random sampling (rule of thumb: 95% confidence interval is at 40 points for n=25, 30 points n=45, 20 points for n=100, 14 points for n=200, 10 points for n=400, etc) just 20 of the 44 emigrant students’ means are statistically different from those of their original populations (in bold in Table 5).

To evaluate these dissimilarities, it should be taken into account that immigrants are not a direct target in PISA sampling, but a collateral product of general sampling, what widens sampling errors. To remember how casual immigrants’ sampling might be, Table 5 maintains void cells for Greek and Dutch emigrants to Australia. A look at both Tables 4 and 5 shows that several groups appear in or disappear from the samples between 2003 and 2006.

Moreover, PISA samples of students are drawn from clusters (schools), a fact that enlarges the sampling errors. Differences in PISA 2003 (Table 4) can be taken to confirm that differences observed in 2006 (Table 5) are not statistical artefacts. At least 8 of the emigrant students' mean scores – among the 22 marked in bold – do not pass the test.

Suppose, now, that we still have fourteen differences not produced by sampling designs, but by real differences in really different populations. These fourteen differences might only mean that immigrants differ from their *countries*, not from their particular *regions* of origin. For instance, Serbians in Austria and Switzerland mirror better the scores of Serbia than those of the former Yugoslavia (note that Croatia's scores are far above Serbia's). French emigrants in Belgium are strikingly below the general French mean, but the reason could be that most of them come from a rural region close to the border (Hirtt, 2006). Something like this could also happen to Russians in Finland and Greece, to Germans in Switzerland or to Swedes in Norway. But this is just a conjecture that would need to be documented.

It is well known that emigrants never constitute a random sample from their country populations. Certain regions, social classes, ethnic groups and even personality traits often are over- or under-represented among them. Part of the deviations from the country of origin means observed in Tables 4 and 5 might simply derive from homologous deviations in immigrants' social background. Economic emigrants from developed countries tend to be of higher socioeconomic background than their counterparts from developing countries. The children of those workers and peasants that emigrated from Italy, Portugal, Serbia and Spain into Central-Europe score below their origin countries' means, but perhaps at one level with the children of the same social classes who have not moved abroad. We could then expect that controlling for socioeconomic and cultural characteristics of the parents would reduce still further the differences observed in Tables 4 and 5. Dronkers & Heus (2008) have actually calculated that controlling for cultural possessions at home, home educational resources, parental education and parental occupation, somewhat reduces the gap in question. On the other side, selective immigration policies could explain that emigrants to Australia outscore sometimes their native co-nationals, even if they remained in China (Table 6).

To resume, if similarities between youngsters' scores from the same country are at the first glance strong, they become still stronger as comparisons gain in accuracy. For our purposes, it is enough to make clear that scores of pupils from the same country living at home and abroad in different countries are very similar, when not fully equal. In other words, *emigration hardly affects students' PISA scores, which remain at the level of the country of origin and do not come closer to the level of the destination country.*

Table 6. Average scientific performance and n of Chinese children per zone of destination and zone of origin. Pisa 2006

		COUNTRY OF DESTINATION AND ITS MEAN PISA SCORE							
ZONE OF ORIGIN		AU		NZ		MACAO		HONGKONG	
NAME	PISA SCORE								
		527		530					
CHINA MAINLAND	530	562	252	547	117				
CHINA	530					513	3218	542	1854
HONG-KONG	542	572	77			530	31		
MACAO	511							537	30
TAIPEI	532								

Source: PISA 2006 DATA SET. Keys: AU: Australia; NZ: New Zealand

THE HYPOTHESES, OR *EXPLANANTES*

1. Countries of destination can be excluded as an *explanans*. The fact that immigrants score more like the children in their countries of origin than like the children in their countries of destiny may be surprising for conventional wisdom, that expects schools to make a difference, but it is in line with most of the literature, including the PISA reports (Carabaña, 2008). It is, however, possible to consider the question in other way. Most countries of immigration in Tables 4 and 5 clearly belong to one of two groups, Australia and New Zealand on the one side and Central-Europe on the other side, all with high PISA scores. There is no surprise in the fact that migrating among these countries leaves PISA scores unaffected. What asks for explanation is the unchanged performance of immigrants from countries with lower scores. No such countries send emigrants to Australia and New Zealand, except Greece in Table 4, whose emigrants' score rises in Australia. Therefore, in Tables 4 and 5 the case of immigrants in Central-Europe coming from South Europe seems the only one needing explanation.

It would be unwarranted, however, to reduce the question to Europe, focusing the research on traits common to Central European countries (like differentiated educational systems, Germanic languages or a penchant for discrimination), on traits common to the guest-worker countries (being Mediterranean, perhaps?) or on the interaction between both. Against this approach speak, first, the cases of pupils from Poland and Russia abroad in Europe as well as the isolated but significant cases of pupils from Jordan in Qatar and from the several Chinas everywhere (Table 6); and, second, the fact that South European emigrants to Central Europe do not homogenize their scores, but maintain the differences between their origin countries.

Although the evidence may not be fully compelling, the most straightforward and likely hypothesis backed up by figures in [Tables 4 and 5](#) (see the end of this chapter) is the absence of any systematic effect of host countries on the academic results of immigrant students.

2. We must, therefore, turn our attention to the origin countries in search for something that makes immigrant children from low-score countries resilient to the benefits of the better schools, the wealthier economies and the richer social environments in the countries of destination, at least in Central Europe. It has to be something that immigrants carry from their origin to their destination countries, and that they preserve in such a way that it continues to determine the scholastic achievement of their children, even of those born, grown up and educated in the host countries. What can it be?
 - a. Nothing material is so lasting and influential. Wealth, income or living conditions are exactly what emigrants leave behind when they arrive to new countries. In any event, it would be strange if these factors were as efficient, or even more, in the destination as in the origin country.
 - b. The same holds for social, political or economic institutions of the country of origin. The corresponding variables in the host countries, to which students are daily exposed, have no effect on scholastic achievement. It would be really strange if they had any effect acting from the past and without direct contact to the students. And, in fact, statistical analyses have confirmed this obvious conjecture. “At the origin level, political and economic features were shown not to influence the educational performance of immigrant children originating from these countries. That is to say, no support was found for the idea that immigrant children from politically unstable countries perform less in science than their counterparts from politically stable countries” (Heus & Dronkers, 2009, p. 15).
 - c. It would sound rather contradictory to argue otherwise for schools, claiming that the influences of schools that students never or hardly attended fully eclipse the effects of the schools which students have actually attended.
 - d. At first glance, personal characteristics are better *explanantes* than macro level traits. Again, wealth and income are known to count little, and in any event, material possessions are usually higher in the host country. But parental education, occupation and the so-called ‘cultural capital’ are known to be of paramount importance for academic achievement and are for the most part carried from the aboriginal country. However, although in the ‘various origins- one destination’ design, and in the simpler ‘immigrant and non-immigrant’ approach, differences in the social, economic and cultural status of the families account for differences in PISA scores between natives and immigrants, when we shift to ‘one origin, various destinations’ designs, controlling for these variables either brings mean scores of emigrants closer to the mean of the origin country (Dronkers & de Heus, 2008) or has no systematic effect on them.
 - e. Cultural factors are by far the preferred explanations for anthropologists, sociologists and even economists lacking better ideas. Moral traditions, like Confucian ethics, fatalism or protestant ethos, are collegially favoured by

sociologists, be it in Durkheim's tradition of 'explaining the social by the social' or in the track of Max Weber's famous explanation of the rise of capitalism.

Table 7 Illustrates this point for Turks in Central Europe

	MODEL 1		MODEL 2	
	B	SE	B	SE
AUSTRIA	-41,54	6,50	-40,05	6,10
BELGIUM	-8,54	6,82	-12,87	6,39
SWITZERLAND	1,07	5,15 ^{ns}	0,20	4,81
GERMANY	-1,93	6,14 ^{ns}	-6,25	5,74
DENMARK	-48,99	9,84	-46,51	9,19
LIECHTENSTEIN	-31,56	28,50 ^{ns}	-20,90	26,62
HISCED			11,34	0,66
HISEI			0,82	0,08
(Constant)	429,34	1,20	365,35	2,93
R square		0,01		0,14

Source: PISA data, 2006. Hisced, Hisei: see [Table 1](#).

Many theorists believe that values and attitudes, deeply rooted in traditions, folkways and mores, cultivated in newly formed communities, explain the differential success of ethnic groups (Terrén, 2004). Prominent among them is Portes, who relates the success of immigrants in the USA to the maintenance of languages and cultural traditions inside enclave communities, like Cubans in Miami (Portes & Rumbaut, 1996; Portes & Hao, 2005).

No social scientist, however, has to my knowledge gone so far as to maintain, or yet imagine, that these values are so lasting, pervasive and influential that American Greeks, Africans, Southern-Italians, Portuguese or Hispanics at American Schools learn no more or even less than they would have learned at the schools in their home countries, in spite of the huge differences in all kind of resources between them. No social scientist has ever claimed that moral traditions are so lasting, pervasive and powerful that completely neutralize the influence of the new mores to which most immigrants are eager to adapt. (Remember that according to PISA immigrants want to succeed at school more than natives.) Not to mention the fact that national moral traditions, or cultural factors, are too diverse to all of them to act in the same perfect way as wards of aboriginal scholastic limitations.

The cognitive ability hypothesis.

One thing that fulfils the conditions of coming from the home countries, changing slowly and being a strong determinant of scholastic achievement is cognitive ability.

WHY DO RESULTS OF IMMIGRANT STUDENTS

Attempts to explain group differences in academic results by social scientists rarely question the assumption of equal cognitive abilities. One recent exception is Dronkers & de Heus (2008), although they do it rather implicitly. Their hypothesis is that guest-workers children score low in PISA tests because their fathers were negatively selected for emigration. But they do not make explicit particular traits subjected to this process of selection.

It is nevertheless a well established fact, so well known indeed that there is a tendency to forget about it, that cognitive abilities are the main factor explaining individual differences in academic results or, for that matter, in PISA tests. It should also be so for group differences, unless groups happened to contain individuals of equal cognitive ability. But, why should they? Or, more precisely, why should members of each national, ethnic or just local group have one and the same distribution of cognitive abilities?

Psychologists who have specialized in the study of group differences in mental abilities (Lynn & Vanhanen, 2002; Flynn, 1987; Colom & Andrés, 1999) have calculated mean cognitive abilities for people in most countries by laboriously collecting results from isolated studies, often quite small ones. Their findings have been harshly criticized on these grounds. However, the CI scores obtained in this way from studies carried out years ago almost perfectly match the scores of recent international systematic studies like TIMMS, PIRLS and, of course, PISA.

To such a good match contributes the fact that PISA tests measure learning ability rather than actual knowledge. Paradoxically in trying to construct 'content free tests' (about which Professor Lundgren told us in the symposium), PISA teams of experts have produced the perverse effect of measuring something close to general cognitive ability. This interpretation of PISA tests as ability tests is strongly backed up by the huge, indeed astonishing, correlation (above .80) among scores in reading, mathematics and science across individuals. And even if PISA reports never mention it, the fact has captured the attention of some authors (Weiss, 2009).

If it is admitted that PISA actually measures cognitive abilities that differ between national groups, the question still remains how resilient they are to the environmental changes brought about by emigration. Cognitive abilities depend on genes and environment in ways that are not well established. The big IQ gains in recent generations detected by Flynn and others (the so-called Flynn effect) are commonly attributed to improvements in living conditions. The fact that emigration does not improve mental abilities would mean that it does not improve living conditions to the degree, or in the way, that would produce somewhat of a Flynn effect.

SOME CONCLUDING REMARKS

Let me conclude with a couple of notes about implications of the above for research and policies.

As far as research is concerned, the fact that immigrant students maintain at least until the second generation, and even more, the PISA scores of the countries they or their parents come from, reduces the question of their distances in

achievement to native pupils to the general problem of differences in PISA scores between countries and regions. Therefore, whatever the cause of the one type of differences, it should also be taken as the cause of the other type. The reduction of both questions to only one, greatly simplifies the quest for explanations, excluding the many that could be adequate for one of the two facts, but not for the other.

The hypothesis of national differences in cognitive or learning ability (the CA hypothesis) greatly alleviates the burden that the assumption of equality in group cognitive abilities puts on the schools of both home and host countries. Schools at the home countries are no longer supposed to be so bad as their outcomes. Schools in the host countries are no longer supposed to be as good as the results of their native students. Therefore, they are free of the suspicion of depressing the results of immigrant students, or of being unable to help them to develop their full potential. According to the CA hypothesis, schools effectiveness cannot be gauged without taking into account the ability of their intakes.

One last, this time political, remark. The CA hypothesis is not a way of 'blaming the victim'. First, there are no victims. Whatever the forces, genetic or environmental, natural or social, that shape human diversity in all kinds of abilities, they are mostly unconscious and uncontrollable, and efforts to reverse them are in general strenuous and of little effect. Second, there is no blame. Humans can not be blamed for traits outside the reach of their will and effort. Immigrant students, as anybody else, should not be treated as categories whose mean achievements must be either raised or lowered, but as individuals with the duty of doing as well as they can and with the right of not doing more than they can.

REFERENCES

- Carabaña, J. (2008) *Las diferencias entre regiones y países en las pruebas PISA*. Publicación electrónica, Colegio Libre de Eméritos, Madrid, 2008. (www.colegiodeemeritos.es)
- Colom Marañón, R. & Andrés-Pueyo, A. (1999) El estudio de la inteligencia humana: recapitulación ante el cambio de milenio. *Psicothema*, 11(3), 453-476.
- Costa-Lascoux, J. (1996). Immigration: de l'exil à l'exclusion. In Paugam, S. (dir), *L'exclusion. l'état des savoirs*, Paris: La Decouverte.
- Dronkers, J. & Levels, M. (2007). Do School Segregation and School Resources Explain Region-of-Origin Differences in the Mathematics Achievement of Immigrant Students? *Educational Research and Evaluation*, 13(5), 435-462.
- Dronkers, J. & Heus, M. de (2008). Negative selectivity of Europe's guest-worker immigration? The educational achievement of children of immigrants compared with the educational achievement of native children in their origin countries. *European University Institute, Working Paper*.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101,171-191.
- Heus, Manon de & Dronkers, J. (2008). Can educational systems compensate for societal features? The effects of educational systems and societal features of both countries of origin and destination on the scientific literacy of immigrant children in western countries. *European University Institute, Working Paper*.
- Heus, M. de, Dronkers, J. & Levels, M. (2009). Immigrant pupils' scientific performance: the influence of educational system features of countries of origin and destination". *European University Institute, Working Paper*.
- Hirtt, N. (2006). Handicap culturel, mauvaise intégration ou ségrégation sociale?, 1-29. Online at: http://www.ecoledemocratique.org/IMG/pdf/pisa_immig_fr.pdf

WHY DO RESULTS OF IMMIGRANT STUDENTS

- Jonsson, J. O. (2002). The educational and labour market attainment in Sweden of Immigrants and native-born Swedes. *Paper at the ISA RC28 Meeting*, Oxford.
- Karady, V. (1987). Juifs et luthériens dans le système scolaire hongrois. *Actes de la Recherche en Sciences Sociales* 69, 67-85.
- Levels, M. & Dronkers, J. (2008). Educational performance of native and immigrant children from various countries of origin. *Ethnic and Racial Studies*, 31(8), 1404-1425.
- Levels, M., Dronkers, J. & Kraaykamp, G. (2008). Immigrant children's educational achievement in western countries: origin, destination, and community effects on mathematical performance. *American Sociological Review*, 73 (October), 835-853.
- Lindo, F. (2000). The silent success. The social advancement of Southern European labour migrants in the Netherlands. In H. Vermeulen & R. Penninx (Eds.), *Immigrant Integration. The Dutch Case* (pp. 123-152). Amsterdam: Het Spinhuis.
- Lynn, R. & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.
- Marks, Gary N., (2005). Accounting for immigrant non-immigrant differences in reading and mathematics in twenty countries. *Ethnic and Racial Studies*, 28(5), 925-946.
- OECD (2004). *Learning for Tomorrow's World – First Results from PISA 2003*. Paris: OECD.
- OECD (2006). *Where immigrant students succeed. A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- OECD (2008). *PISA 2006. Science Competencies for Tomorrow's World*. Paris: OECD.
- Portes, A. & Rumbaut, R. G. (1996). *Immigrant America: A portrait*. Berkeley: University of California Press.
- Portes, A. & Hao, L. (2005). The schooling of children of immigrants: contextual effects on the educational attainment of the second generation. *Migraciones*, 17, 7-4.
- Shavit, Y. (1990). Segregation, tracking and the educational attainment of minorities: Arabs and Oriental Jews in Israel. *American Sociological Review*, 55(1), 115-26.
- Siguán, M. (1998). *La escuela y los Immigrantes*. Barcelona: Paidós.
- Siguán, M. (2003). *Inmigración y adolescencia. Los retos de la interculturalidad*. Barcelona: Paidós.
- Terrén, E. (2004). *Incorporación o asimilación: La escuela como espacio de inclusión social*. Madrid: La Catarata.
- Weiss, V. (2009). National IQ means transformed from programme for international student assessment (PISA) scores, and their underlying gene frequencies. Online at <http://mpr.ub.uni-muenchen.de/14600/MPRA Paper No. 14600>.

Julio Carabaña
Universidad Complutense (Madrid)

J. CARABAÑA

Table 4. Average Mathematical Performance and n of Immigrant Children per Country of Destination and Country of Origin. PISA 2003

COUNTRY OF DESTINATION AND ITS MEAN PISA SCORE																
	AU	AT	BE	CH	DE	DK	EL	LV	LU	NL	NZ	IR				
COUNTRY OF ORIGIN	524	506	529	527	503	514	445	483	493	538	523	503				
NAME	PISA SCORE															
ALBANIA	381***			412	255		403	195								
CHINA	550 (HK)	570	129								541	73				
FRANCE	511			460	236	520	96									
GERMANY	508	529	45			528	94			507	65*					
GREECE	445	470	45													
ITALY	466	503	73			472	283	420	33		473	120				
NETHERLANDS	538	502	27			530	65									
POLAND	490			493	36			495	99							
PORTUGAL	466					473	206				444	603				
RUSSIA	468							466	180**		400	99	495	238		
SERBIA	437		459	272		456	403									
SPAIN	485					477	80									
TURKEY	423		433	137	429	137	437	146	413	188	424	49			484	372*
UNITED KINGDOM	532***	539	457													
YUGOSLAVIA(FORMER)															421	92*

Source: Table A1 (number) and Table 2 (scores) from Levels, Dronkers and Kraaykamp, 2008. The authors do not detail decisions made regarding missing values and priorities of possible countries of birth (841,n4). They label all European in Netherlands as Germans and all non-european as Turks(*) ** Taken from PISA 2006, table 2.8. *** Taken from PISA 2000. Country of origin scores from OECD, 2004:90.Keys: AU=Australia; AT=Austria; BE=Belgium; CH=Switzerland; DE=Germany; DK=Denmark; EL=Greece; LU=Luxembourg; LV=Latvia; NL=the Netherlands; NZ=New Zealand; IR=Ireland

WHY DO RESULTS OF IMMIGRANT STUDENTS

Table 5. Average Scientific Performance and n of Immigrant Children per Country of Destination and Country of Origin. Pisa 2006

COUNTRY OF DESTINATION AND ITS MEAN PISA SCORE																		
COUNTRY OF ORIGIN	PISA SCORE	AU	AT	BE	CH	DE	DK	EL	LV	LU	NL	NZ	FI	IS	PT	QA	NOR	
ALBANIA	376***	527	511	510	512	516	496	473	490	486	525	530	563	454	474	349	487	
BRAZIL	390							434	187						464	45		
CHINA	530****	562	252									547	117					
CROATIA	493		458	36														
FRANCE	495			448	125	507	119			505	184							
GERMANY	516		521	44	508	47	549	173		537	199	504	90*					
GREECE	473		0				419	13										
ITALY	475				443	300	415	30		430	98							
JORDAN*	422															401	133	
KOREA	522	514	69									528	76					
NETHERLANDS	525		0		522	95												
POLAND	498		523	25	439	94		497	77									
PORTUGAL	474				454	241				420	799							
RUMANIA**	419		437	27														
RUSSIA	479					466	79		496	186			550	25	483	400**		
SERBIA	436		426	78		427	952											
SPAIN	488				466	119												
SWEDEN	503																465	39
TURKEY	424		380	161	414	156	425	244	411	198	374	89						
UNITED KINGDOM	515	542	490										569	200				

Source: Table A1 (number) and table 2 (scores) from Heus and Dronkers, 2009. The authors do not detail decisions made regarding missing values and priorities of possible countries of birth. They label all European in Netherlands as Germans and all non-european as Turks(*). ** Added from PISA 2006 set of data. *** Taken from PISA 2000. **** Imputed from Hong Kong, Taipei and Macao. Country of origin scores from OECD, 2007. Keys: AU=Australia; AT=Austria; BE=Belgium; CH=Switzerland; DE=Germany; DK=Denmark; EL=Greece; FI=Finland; U=Luxembourg; LV=Latvia; NL=the Netherlands; NZ=New Zealand; IS=Israel; PT=Portugal; QA:Qatar. NOR=Norway

SECTION V

EXTREME VISIONS OF PISA: GERMANY AND FINLAND

HANNU SIMOLA AND RISTO RINNE

EDUCATION POLITICS AND CONTINGENCY

Belief, Status and Trust Behind the Finnish PISA Miracle

Media visibility and the political use of global rankings have highlighted the topicality and relevance of comparative studies in education. This popularity has not entailed the development of theoretical instruments in the field, however. Conversely, non-historical and decontextualised concepts such as *efficiency*, *accountability* and *quality* are colonising the educational world undisputed and uncontested, largely due to the fact that they have been internationally advocated. Comparative education is still suffering from certain methodological deficits and serious under-theorisation. (See, e.g., Marginson & Mollis, 2001; Schriewer, 2006; Dale, 2009; Cowen, 2009; Simola, 2009)

The theoretical tradition in comparative education research is not too strong, which may be one reason for the success of the ahistorical and decontextualised conceptualisations in the field. Likewise, functionalistic comparisons based on different system models have become the mainstream among transnational organisations such as the World Bank, the OECD and the EU. This rather mechanistic kind of paradigm has been the bane of comparative research in the past.

There has also been heavy criticism of the solely quantitative comparative type of research, and case-study methodology has found its place. One of the pioneers in this context was Charles Ragin (1987; 1989; 1992), who tried to put right the antinomies of the quantitative and qualitative approaches through so-called “analytic induction”, taking into account the diversity of the causes and the reasons for social change in different nations. One of the most interesting approaches in comparative research is the so-called “patterned mess” – suggested by Michael Mann (1986; 1993), among others, in his comparative analysis of sources of social power. These approaches are very well suited to comparisons of Higher Education politics in different countries, for instance, because HE institutions have usually operated in a state of “organized anarchy” (Clark, 1993; see Kivinen & Rinne 1995, pp. 231, 241).

António Nóvoa and Tali Yariv-Mashal’s observation of a few years ago seems still to be valid:

The problem is that the term comparison is being mainly used as a flag of convenience, intended to attract international interest and money and to entail the need to assess national policies with reference to world scales and hierarchies. The result is a ‘soft comparison’ lacking any solid theoretical or methodological grounds. (Nóvoa & Yariv-Mashal, 2003, p. 425)

The problem is not restricted to the field of comparative education, of course. Susan Strange (1997), a prominent representative of the approach known as international political economy, sharply criticised ‘neo-institutionalists’ and ‘comparativists’ for reiterating policy agendas aiming at national success in the global struggle for competitiveness. This ‘unbearable narrowness of the national view’ (Kettunen, 2008) could be seen as a professional illness emanating from the comparative policy studies of our times.

Roger Dale (2009, p. 123; cf. Beck, 2006) refers to three fundamental problems in comparative studies in education: methodological nationalism, methodological statism and methodological educationalism. The nation and the nation-state are still seen as the only real and final policy unit, and the very concept of education is taken for granted. Instead of ‘models’ and the convergence or divergence among them, we should be more interested “in the webs of structural power operating throughout the world system than in comparative analysis of discrete parts of it, bounded by territorial frontiers dividing states” (Strange, 1997, p. 182). Education is still most often seen only as a question of increasing competences and qualifications among nation-state citizens in the face of global competition among knowledge-based economies. Decades ago John W. Meyer (1986, pp. 345-346) warned about ‘functional blinders’ that permit us to take schooling as a self-evident rational system and create a moralist discourse - among not only educationalists but also sociologists of education.

This narrowness of the national view easily creates a blind spot in terms of how interactions and comparisons reconstruct the national or the local: how transnational interactions and crossings constitute the national parties of these relationships, and here we come to the crucial role of comparative practices as a mode of reflexivity that (re)shapes individual and collective agency (Strange, 1997). In pursuance of an understanding of such a complex phenomenon as the relationship between the global, the regional, the national and the local in education policy formation it is vital to consider the theoretical conceptualisations from a both/and rather than an either/or point of view. A good and illuminating example here is the controversy among researchers of nationalism and the frequently observed confrontation between understanding nationalism as ‘the invention of traditions’ by the elite (e.g., Hobsbawm, 1990) or as creating prerequisites and limits for ethnic identities (e.g., Smith, 1995). From the perspective of comparative research, nationalism as an elite strategy and nationalism as a socio-cultural frame are both valid approaches. Comparative actions (such as the PISA studies) should be analysed *both* as economic, political and cultural practices (see, e.g., Nóvoa & Yariv-Mashal, 2003; Simola et al., forthcoming) *and* as international exhibitions of national competitiveness in the global educational market place.

A Finnish researcher of modern history, Pauli Kettunen (forthcoming), emphasises that criticism of the nation-state-centred view on globalisation should not just declare it outdated, but should rather take it seriously as an influential mode of thought and action, and recognise how it is embedded in the structures of globalised economic competition. Such a critical ambition means getting beyond the train of thought that contrasts the profound internal permanence of national agency with the drastic change in the external environment. Historicity means the temporal

multi-layeredness of institutions and discourses that constrain and enable agency. It also means the contingency of each action situation, in which the actors must handle the tension between experiences and expectations. Making comparisons and making histories are crucial modes of reflexivity in social action, and this also applies to constructions of collective agency, not least those evolving in the framework of the nation-state society and influencing the making of the welfare state.

PROMISING THEORETICAL CONCEPTS FOR COMPARISON

One of the most promising remedies for theoretical deficit in research appears to lie in emphasising *historicity* (cf. Kettunen, forthcoming) or the *socio-historical analysis of complexity* (Charle et al., 2004; Schriewer, 2006; 2009). This rich domain of knowledge could also be characterised as a history of the transnational (*transnationale Geschichte*, Conrad, 2006) or entangled history (*histoire croisée*, Werner & Zimmermann, 2006). To put it simply, the point is that it is not enough to study dependencies and interactions among national states, or the border-crossing transfer of ideas and concepts.

What, then, are the most promising theoretical concepts from the perspective outlined above? In what follows, three promising analytical dimensions are very briefly outlined: first, bringing the theoretical concepts of path dependency, convergence and contingency together; second, tracing the history of *problématiques* or asking what is the problem the new education policy is meant to solve; and finally, analysing national and local interpretations and translations as hybrids.

First, path dependency and convergence have often been seen as simplistic dualism in comparative studies: the former covers major national specificities and the latter refers to international tendencies. The approach essentially underestimates both the insecurity and openness of the horizon of expectations and the relative freedom of more or less conscious and informed actors. This deficit is even more assuredly fatal in these global and late-modern times characterised as the ‘Era of Contingency’ (Joas, 2008; Joas & Knöbl, 2009), in which the difference between the already-done and the yet-to-be-done is vital and things are increasingly not necessary or impossible. At the crossing of these two dimensions – path dependency and contingency on the one hand, and path dependency and convergence on the other – we might find histories and comparisons as forms of reflexivity in social practices. Contingency is one essential element in creating *Spielraum* for ‘politicking’ (Palonen, 1993 and see also 2003). Relating the past, the present and the future, or experience and expectation, and recognising and interpreting differences and similarities are inherent aspects in human agency (Kettunen, forthcoming).

Secondly, Nóvoa and Yariv-Mashal (2003, pp. 436-437) propose in their seminal paper that the very focus of comparative research in education should be on *problématiques* rather than on ‘facts’ and ‘realities’ that, by definition, are incomparable in a strong sense. They can be contrasted in order to highlight differences and similarities, but it is hard to go further. Therefore, they claim, only problematisations can constitute the basis for complex comparison. Problems are anchored in the present but possess a history and anticipate different possible

futures. They are also located and relocated in places and times, through processes of transfer, circulation and appropriation. Furthermore, they can only be elucidated through the adoption of new zones of observation that are inscribed in a space delimited by frontiers of meaning, and not only by physical boundaries.

Finally, the concept of hybridisation could be suggested to cover different, more and less conscious interpretations and translations of travelling, borrowed and learned policies in education. The concept is used here as introduced by two eminent U.S. historians of educational reform, David Tyack and Larry Cuban (1995). They emphasised the underrated influence of teachers, or as they put it of 'street-level bureaucrats', in educational reforms. In this sense they concluded that there should be much more research on how schools change reforms rather than vice versa. Another conclusion was that school reforms in the U.S, have always brought about change, but rarely the change that was intended.

This fits well with Stephen J. Ball's eminent semi-classic characterisation of the distance and controversies between policy writing and its implementation:

National policy-making is inevitably a process of *bricolage*; a matter of borrowing and copying bits and pieces of ideas from elsewhere, drawing upon and amending locally tried-and-tested approaches, cannibalising theories, research, trends and fashions, and not infrequently a flailing around for anything at all that looks as though it might work. Most policies are ramshackle, compromise, hit and miss affairs that are reworked, tinkered with, nuanced and inflected through complex processes of influence, text production, dissemination and ultimately recreation in contexts of practice. [...] In short, national policies need to be understood as the product of a nexus of influences and interdependencies, resulting in 'interconnectedness, multiplexity and hybridisation' [...] that is, 'the intermingling of global, distant and local logics'. (Ball, 2001a, b)

In the very same spirit, Norman Fairclough and Ruth Wodak (2008, p. 112) felicitously characterise the processes to be studied as:

[...] a complex and interrelated series of relationships between strategies and their contingent implementation in structures, imaginaries and their contingent operationalisation in practices and institutions, and implemented/ operationalised strategies/imaginaries and ideologies and legitimations.

By way of simplification one could state that path dependency, convergence and contingency belong to the structural dimension of action whereas problematisation and hybridisation are connected to its agency/strategic and actor/tactical dimensions. Along this trident distinction one could argue for connecting theory-rich discoveries such as travelling and embedded policies, vernacular or indigenous globalisation (Ozga & Jones, 2006; Ozga & Lingard, 2007) and commonality within difference and exogenous trends (Marques Cardoso, 1998; Sweeting & Morris, 1993) as structural, whereas level-specific policy technologies, techniques and mechanisms (Simola, 2009) could be seen from the agency/strategic perspective, and externalisations (Steiner-Khamsi, 2004; Schriewer & Martinez, 2004) and indigenous foreigners (Popkewitz, 2009) as part of the actor/tactical dimension.

Here we concentrate on contingency as one promising conceptualisation of the historical approach to comparative studies in education.

CONTINGENCY AS UNCERTAINTY AND FREEDOM, COINCIDENCE AND
SPIELRAUM

The German sociologist Hans Joas (2008) has characterised our time as an “Age of Contingency”. It seems plausible that the concept of contingency is able to capture something essential in our society in that it carries attributes such as post-traditional (Giddens), postmodern (Bauman) and the risk society (Beck). Following Niklas Luhmann’s definition, Joas defines contingency as follows:

A fact is contingent if it is neither necessary nor impossible – something that is but does not have to be. I think this definition is useful because it makes clear at the outset that the best way to understand the meaning of contingency is to see it as a counter-notion to another idea, namely ‘necessity’. Thus the precise meaning of the term ‘contingency’ depends on the precise meaning of the term ‘necessity’ that it presupposes. If ‘necessity’ referred, as in pre-modern philosophy, to the idea of a ‘well-ordered cosmos’, ‘contingency’ referred to the incompleteness and imperfection of the merely sensual and material world on the one hand, and to the liberty and creativity of God’s unrestrained will on the other. (Joas 2004, p. 394)

The concept of contingency carries a double meaning: on the one hand it signifies *coincidence* or *conjunction*, and on the other it is *free will* or *volition* (Joas 2008, p. 209). In the former sense it refers to the uncertainty and ambivalence and in the latter sense to possibilities and the *Spielraum* of the actor.

The former dimension, the uncertainty side of contingency, so to say, emphasises the fact that our history and living are essentially haphazard and random: things often happen by accident. Nevertheless, as the US sociologist Howard S. Becker states:

[...] social science theory looks for determinate causal relationships, which do not give an adequate account of this thing that ‘everyone knows’. If we take the idea of ‘it happened by chance’ seriously, we need a quite different kind of research and theory than we are accustomed to. (Becker 1994, p. 183)

In this latter sense, the freedom side of contingency could be seen as an ability to handle and face the contingent characteristics of reality; as “the art of playing with the contingency” as the eminent Finnish political scientist Kari Palonen puts it:

Polity and policy refer to attempts to regiment (polity) or to regulate (policy) the contingency characteristic of politics as action. As opposed to them, politicization refers to opening new aspects of contingency in the situation and thus expanding the presence of the political in it. Politicking may be interpreted as the art of playing with the contingency, using it both as an inescapable moment of the situation to be considered in any case and as an instrument against opponents less ready to tolerate or make use of the presence of the contingency. (Palonen 1993, p. 13)

The concept of contingency is put forward in at least a few Finnish texts as an explanation for the Finnish PISA success. For example, Simola (2005, pp. 465-466) concluded his historical and sociological remarks on this success as follows: “The Finnish ‘secret’ of top-ranking may [...] be seen as the curious contingency of traditional and post-traditional tendencies in the context of the modern welfare state and its comprehensive schooling.”

Similarly, Jarkko Hautamäki, the head of the Finnish PISA 2006 Team, recently stated: “The major point to know [for understanding the Finnish comprehensive school] is that the new system was indeed comprehensive. This was both a necessity [...] and a chance encounter, a lucky constellation of political, economical and social conditions.” (Hautamäki et al., 2008, p. 197; cf. Kupiainen et al., 2009)

We elaborate the analysis a little further in this chapter: what might be that “curious contingency of traditional and post-traditional tendencies”, or the “chance encounter, a lucky constellation” that would make the Finnish PISA success understandable, at least to some extent?

Our aim here is to develop and even to test contingency as a theoretical instrument for comparative understanding. Only theoretical conceptualisations can constitute a basis for complex comparisons. Therefore, our main objective is to see whether the concept of contingency could shed new light on and promote a deeper as well as a broader understanding of the national phenomenon known as the ‘Finnish PISA success’, rather than to try to explain it in comparative terms.

In fact, what we are trying to illustrate are three rather common beliefs emanating from the recent national discussion in Finland. All of these beliefs seem to be rather distinctive compared to other nations’ beliefs and discussions, and they certainly have at least some generative roots in Finnish national history. The first is that the Finns share a strong belief in schooling, the second that teaching is rather highly appreciated as a profession in Finland, and the third that the Finnish comprehensive school enjoys rather high trust on the part of parents, authorities and politicians. All three are national ‘truths’ in a way (e.g., Simola, 1998; Simola et al., 1998; Heikkinen et al., 1999), widely accepted even though there is not too much empirical research evidence behind them. They are definitely constituent parts of the national self-understanding in terms of education. In this chapter we address the question of why these beliefs exist rather than whether they are true or not. Finally, we also impugn traditional functionalist and rationalist explanations of comparative research in education.

A HIGH BELIEF IN SCHOOLING

There are astonishingly few comparative studies on Finnish education, even related to the other Nordic countries. Nevertheless there is a strong national consensus that, on the international level, Finns appreciate education, or schooling to be more precise, very much. Therefore, the belief in schooling as an agent for social equality and as a cornerstone of continuity and consensus in Finnish education policy has remained stronger than in many other Western countries.

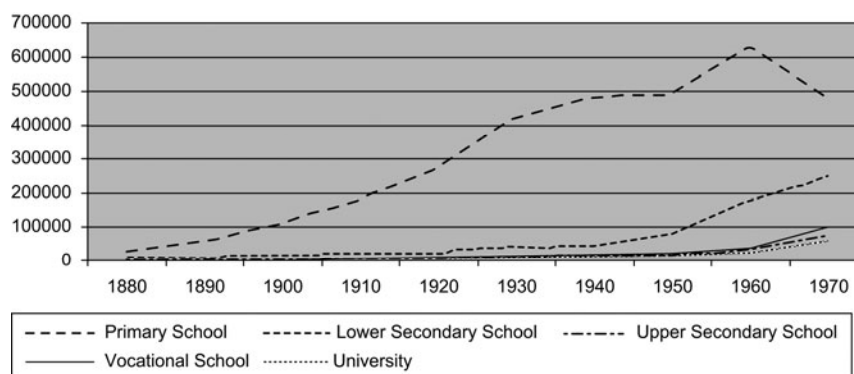
Our hypothesis is that the high belief in schooling is an outgrowth from the contingent conjunction of three social changes that all came exceptionally late in

Finland: the expansion of schooling, the modernisation of the occupational structure and the construction of the welfare state.

Finland was among the last countries in Europe to establish compulsory education. Six-year elementary education was made compulsory by law only in 1921, in the same year as in Thailand, whereas the relevant legislation was in force in Denmark in 1814, in Sweden in 1842 and in Norway in 1848. Moreover, primary-school expansion was slow even after the law, and compulsory education was not fully functional and did not cover all children across the whole country and among all social groups until just before WW2. (Rinne & Salmi, 1998, p. 27; Ramirez & Boli-Bennett, 1982; Rinne, 1984)

All this is indicative of the fact that the Finnish success story in education is historically very recent: whereas almost 70 per cent of the younger generation nowadays aspire to a higher-education degree, among their grandparents about the same proportion obtained the full elementary-school certificate. [Table 1](#) clearly illustrates the late blooming of Finnish education.

Table 1. The expansion of schooling in Finland 1880-1970

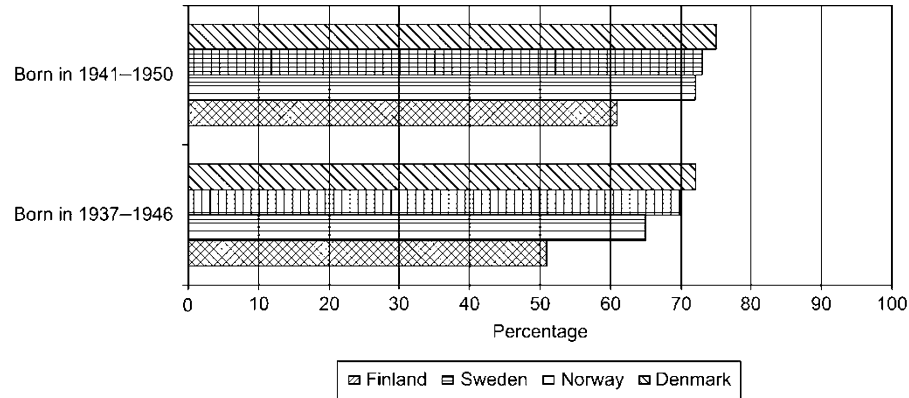


Sources: Kivinen 1988, p. 48; Kivinen, Rinne & Ahola 1989, p. 62; Kerr, forthcoming
 Nb. The figures in other forms of education than that of the Primary and Lower Secondary Schools are very low and near to each other.

The late development of the educational system at the secondary level in Finland and the low percentage of participation in secondary education compared to the other Nordic countries are clearly visible in the following [Table 2](#). In 2001 only about half of 55-65-year-olds had a certificate of secondary education (51%) compared to 65-72% in the other Nordic countries. The differences were still remarkable - well over ten percents in 2005 - compared to the other Nordic countries

Because of the late historical formation and widening of the educational system, the educational gaps between older and younger generations in their educational levels are among the widest in Europe.

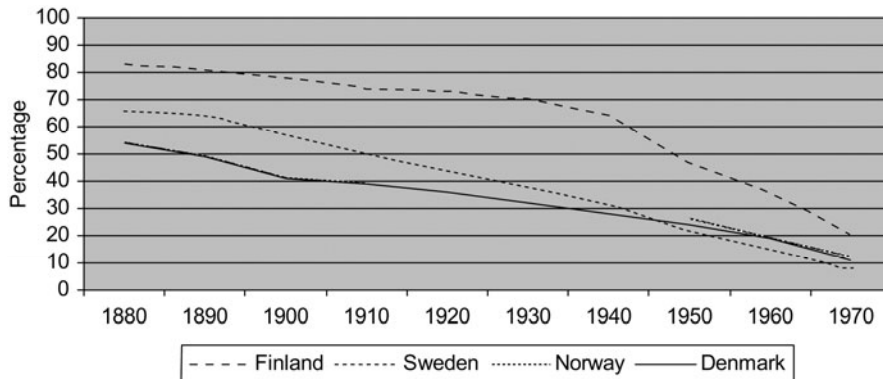
Table 2. Two Nordic population cohorts aged 55-64 years with at least an upper-secondary education



Source: *Education at a Glance* 2002, p. 37; 2007, p. 37

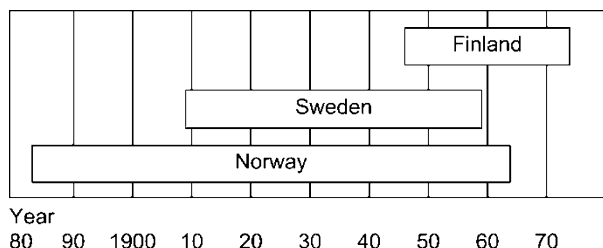
The same is also true of the modernisation of the occupational structure. Finland belongs to the group of European nations that has most recently left behind their agrarian society and life style. The process of industrialisation and urbanisation was sluggish until the Second World War, compared with Central Europe and the other Nordic countries. In 1945, 70 per cent of the Finnish population lived in rural areas, and almost 60 per cent were employed in agriculture and forestry. Following the great migration in the 1960s, by 1970 half were living in cities and 32 per cent were employed in industry and construction (cf. e.g., Alapuro et al., 1987). [Tables 3 and 4](#) contrast the late but rapid change in the Finnish occupational structure with the changes in other Nordic countries.

Table 3. Change of working population in agricultural work in Nordic countries 1880-1970



Source: Pöntinen 1983, p. 46. Nb. The change in Norway and Denmark is during the table very similar (from 1880 to 1910 exactly the same) and therefore the table in their cases is nearly unreadable.

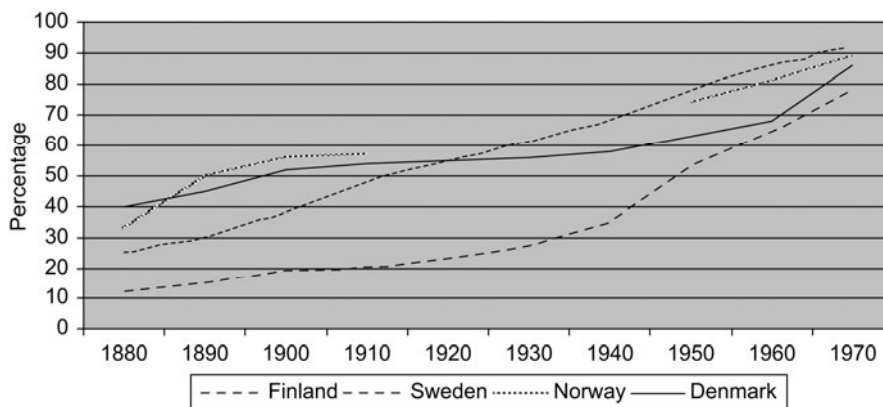
Table 4. Change of working population in industrial and service work in Nordic countries 1880-1970



Source: Pöntinen 1983, p. 46

Whereas the demise of agrarian labour took place over 80 years in Norway, and over 50 years in Sweden, it happened in Finland within 20 years. No wonder, then, that the construction of the welfare state began a decade later than in the other Nordic countries.

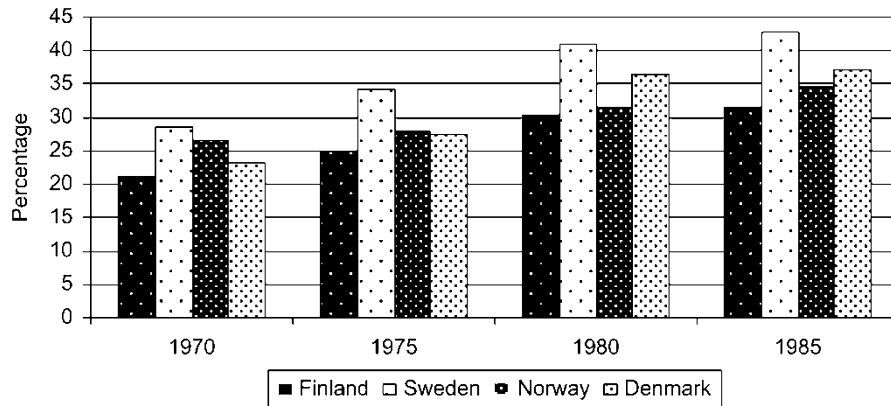
Table 5. The timing and rapidity of the changes in occupational structure in three Nordic countries: the period during which the agrarian labour force decreased in proportion from 50 to 15 per cent



Source: Karisto et al., 1998, p. 64

Expansion of the welfare state after WW2 meant an upheaval in the labour markets of the Nordic countries. Public-sector employment in Finland grew from 20 to over 30 per cent between 1970 and 1985. Typical of the Finnish model was that the growth began later but also continued longer than in the other Nordic countries (Tables 6 and 7).

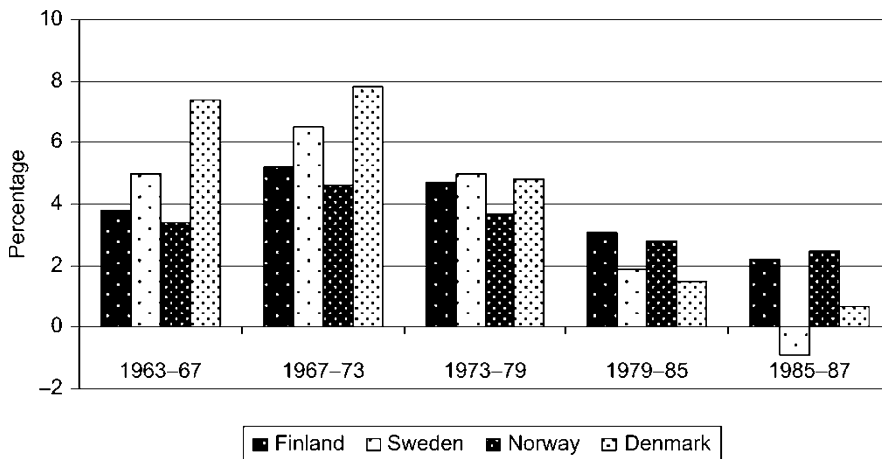
Table 6. Public employment in Nordic countries 1970-1985



Source: Kosonen 1998, p. 152

We could therefore conclude that the high belief in schooling resulted from the contingent conjunction of its late expansion, the late modernisation of the occupational structure and the late construction of the welfare state. These social changes happened in most countries successively rather than one at a time. It may be that this rare conjunction created a strong collective experience of causality between progress in formal education and simultaneous social advancement.

Table 7. Growth of the work force of the public sector in the Nordic countries 1963-1987



Source: Alestalo 1991, p. 8

In fact, the eminent Finnish sociologist of education Ari Antikainen (1990) referred to the very same phenomena when he wrote that the overall rise in student enrolment brought increasing numbers of students from the lower classes, even though their proportion of the total number remained low. This might be “a shared experience among the common people”, who also have their own experience of

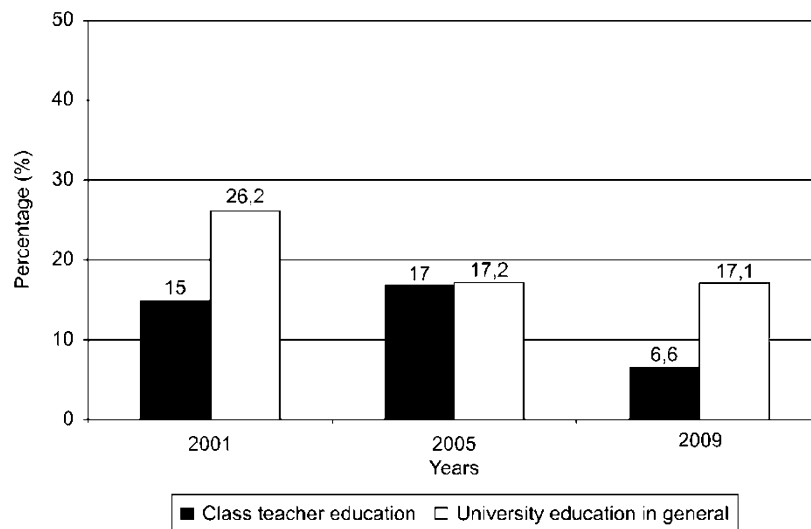
education as a real resource in the rapid transformation of Finnish society, not least as a channel of migration from rural areas and agriculture to the cities in the period of the ‘Great Migration’, 1960-1975.

THE HIGH STATUS OF TEACHERS IN COMPREHENSIVE SCHOOLS

Although the belief in schooling is not well documented, the popularity of the teaching profession among Finnish students is a fact. Year after year teaching has retained its position as one of the most popular careers in terms of university entrance examinations (see e.g., Jussila & Saari, 2000; Kansanen, 2003). According to a survey among candidates for the matriculation examination (i.e. final-year students in upper-secondary school), teaching is clearly the number-one career choice and overtakes traditionally favourite professions such as medicine, law, psychology, engineering and journalism (*Helsingin Sanomat*, 11 February, 2004). It is also clear that teaching in primary schools is the most popular choice: the rate of acceptance for training is around 10 per cent.

Table 8 below shows clearly and more concretely that Finland has no shortage of students wanting to become teachers, and there are plenty of applicants for university places. During the 2000s only about every seventh, or at least less than every tenth aspiring teacher was admitted to the university teacher-training programme. In fact it has become clearly more popular than studies at university in general, and consequently it has become more difficult to get into university for teacher training than to take almost any other academic course of study. This situation is quite unique in the world.

Table 8. The percentages of accepted students versus applicants for teacher training and for university education in general in the 2000s



Source: Kumpulainen & Saari 2005, pp. 10 and 12; Kumpulainen 2009, pp. 20 and 22; OPH 2009; OPM 2010

One major reason for the popularity of primary-school teaching as a profession is the Master's-level qualification required of all teachers,¹ which is still unique in international terms. The teacher's career in Finland, even at primary-school level, is no *cul-de-sac* or second-class honour, but is on a par with all other professions requiring higher university degrees, which on an international level corresponds to the M.A. This qualifies them academically for doctoral studies, for example. It has also made teaching in primary school an accepted profession and a standard career choice among the offspring of those in the upper social strata.

Although the high level of teacher education is lauded as one of the essential factors behind the Finnish PISA success, no one has thus far and in this connection referred to the contingent coincidence and conjunction of the establishment of Master's-level teacher education.

Our second hypothesis is that Master's-level teacher education was realised due to the coincidence of teacher-education reform and the General Degree Reform of Higher Education.

The reform in teacher education was planned in close connection with the Comprehensive School Reform of the 1970s. The 1971 Act on Teacher Education transferred primary-school teaching to the universities. However, the degree programme was still on a three-year basis and on the Bachelor's level, i.e. at the level of a lower university degree. It is remarkable that there were no state-committee or other authoritative texts proposing the elevation of primary-school-teacher training to the Master's level before 1978. On the contrary, a late-stage teacher-education committee (KM 1974, KM 1975) headed by an influential professor of education suggested in 1975 that even the four-year degree model should *not* incorporate Master's-level studies in education.

It was only the coincidence with the General Degree Reform of Higher Education launched in 1977-80 that brought teacher education up to the Master's level, and this was also a coincidental consequence of the abolishment of the Bachelor's level,² with a few exceptions, from Finnish universities. Since 1979 primary-school teachers have complete their MA studies in four to five years, with educational science as their major. What is remarkable here is that the final decision was made as part of the General Degree Reform of Higher Education, and the thousands of pages of committee reports and memoranda written since the late 1960s by specialists in teacher education were ignored. (Simola 1995, pp. 184-185; 1993)

This gives us reason to conclude that this clearly essential decision behind the popularity and status of the Finnish teaching profession was apparently realised due to the contingent coincidence of the General Degree Reform of Higher Education with the implementation of the reforms in teacher education (Simola 1993; Simola et al., 1997).

HIGH-TRUST CULTURE

The former social-democratic head of the National Board of Education, Erkki Aho, the main driver of the comprehensive-school reform between 1973 and 1991, made a statement that it was in his period, in the 1980s that the belief and high trust in schooling became a consensual belief in Finland:

The gradual shift toward trusting schools and teachers began in the 1980s, when the major phases of the initial [comprehensive school] reform agenda were completely implemented and consolidated in the education system. In the early 1990s, the era of a trust-based culture formally began in Finland. (Aho et al., 2006, pp. 12 and 132)

To anyone familiar with Finnish schooling this definitely sounds too lofty and too smug to be true. There is clear counter-evidence, too. Perusal of the thousands of pages of state committee and memoranda material between the 1860s and 1990s, and since the implementation of the Comprehensive School Reform in the 1970s, revealed only one exception in which classroom teachers were *not* seen as the very obstacles to developing education and thus as the objects *par excellence* of the reform (Simola, 1995).

It is true, however, that the supervision of work done in Finnish schools is minimal by international standards. All traditional forms of control over the teacher's work had disappeared by the beginning of the 1990s. The school inspectorate, a detailed national curriculum, officially approved teaching materials, weekly timetables based on the subjects taught and a class diary in which the teacher had to record what was taught each hour—all these traditional mechanisms were abandoned. Furthermore, Finland has never had a tradition of nationwide standardised testing at the comprehensive-school level. Indeed, according to a Eurydice report (2004), Finnish teachers at comprehensive schools seem to have the greatest freedom from evaluative control among their European colleagues. All this can be interpreted as very high trust in the work of teachers and the culture of schools, which may legitimate the rare rather autonomous position of the semi-profession of teachers and school welfare institutions. (Rinne, Simola, Mäkinen-Streng, Silmäri-Salo & Varjo, in press)

One should nevertheless keep in mind that the aim of the reform in the 1990s was not to free teachers but rather to restructure the steering of education. Traditional means of normative control were to be replaced by evaluation, realised by the municipal and national authorities. This was clearly expressed by the then Secretary General of the NBE: evaluation is a pivotal element in the new steering system since it “replaces the tasks of the old normative steering, control and inspection system” (Hirvi, 1996, p. 93).

Then something unexpected and stunning happened again. The recession of 1991–93 heralded the deepest peacetime crisis in Finland's economy. It is widely accepted that without shifting decision-making to the local level the municipalities could not have been required to cut spending as much as they did during the recession. Thus the new decentralised and deregulated mode of governance was moulded around the economic principles of savings and cutbacks. The process of decentralisation and deregulation started in the late 1980s, but in the thick of the recession the new legislation radically increased local autonomy and strengthened the judicial position of the municipalities.³ The decentralisation level of educational administration in Finland became one of the highest in Europe (Temmes et al., 2002, pp. 129 and 92).

The radical decentralisation and deregulation spawned two competing coalitions in the national Quality Assurance and Evaluation (QAE) field of compulsory schooling, neither of which has real normative power over the municipalities and schools. On the one hand the Ministry of Education (ME) and the National Board of Education (NBE) consider QAE from the perspective of the education system and the associated legislation, and on the other the Association of Finnish Local and Regional Authorities (AFLRA) and the Ministry of the Interior – often accompanied by the Ministry of Finance – see it in terms of municipal service production and legislation. Both of these coalitions have attempted to assume the leading role in determining the discourse of evaluation in the context of education.

The frustration seemed to be most evident among our interviewees from the NBE, whereas in AFLRA there appeared to be a kind of complacent acceptance of the predominant situation. One high-ranking NBE official explains his/her feelings:

“[...] we have no jurisdiction to touch anything, we have no legislation about it, we have no mechanisms, we have nothing. This, in a nutshell, is our biggest weakness.” (Simola et al., 2009, p. 171)

Given that all these proposals were directive rather than obligatory, it is no wonder that their implementation on the municipal level varies widely. The Finnish Parliamentary Committee for Education and Culture concluded in 2002:

The evaluation work done has had very small effects at the level of municipalities and schools. Nation-level evaluations have been implemented to a creditable extent, but there is no follow-up on how these evaluations affect the actions of the evaluated and the development of the schools. [...] Many municipalities are at the very beginning in the evaluation of education. (CEC. 2002)

Therefore we venture to suggest a dimension of contingency here, too, although different than before. In this case an intervening conjunction – the deep economic recession and the radical municipal autonomy linked to it – circumvented and extinguished the reform intentions. Ironically enough, it seemed to create unintended side effects: trust and freedom.

CONTINGENCY AS FREEDOM AND SPIELRAUM

We have outlined the uncertainty side of contingency: contingency as coincidence and conjunction. As mentioned above in section 1, there is also a freedom dimension: contingency as free will and *Spielraum*. The first version of this chapter was presented in Finnish in Spring 2009 at a seminar involving the most eminent policymakers in Finnish education from the 1970s to the 1990s. It was a surprise that, with only one exception⁴, they sharply rejected all references to coincidence and conjunction: in their eyes it was all part of the purposeful, hard-headed, rational and also successful struggle to implement a consciously and carefully planned educational policy. There was no room for accident or coincidence.

It seemed to me that the policymakers found it hard to understand that accepting a certain randomness in life did *not* necessarily lead to the abandonment of a

certain amount of freedom for the actors, rather the contrary⁵. By way of illustration we might ask, for example, what the freedom or *Spielraum* of the policymakers was and how they capitalised on it in these three historical cases. Let us just sketch out two notions.

First, there were different *levels of conjunction*. In the first case three historical processes (change in the occupational structure, the expansion of mass education and the construction of the welfare state), which ‘normally’ happen with certain time lags, were crammed into a very short period of time. In the second case two relatively separate reforms in the different educational sectors (Teacher Education Reform and the General Degree Reform in Higher Education) coincided. In the third case three reforms in different policy sectors (comprehensive-school governance, municipal autonomy and tightness of economy) were concurrent. What is common to them all is the fact that these conjunctions were not planned or foreseen by the contemporary actors, and the consequences were unexpected.

Secondly, the policymakers *reacted differently* in all these cases. The ‘great conjunction’ in post-war Finland was considered by many contemporary politicians to be a state of emergency: in a poor, semi-agrarian society it was to find place for the ‘Big Age Groups’, born after WW2. The scale of the problem was huge. It was seen as an even bigger task than the recent settling of 450,000 refugees, more than 15 percent of the population, from the part of Karelia that was lost to the Soviet Union in WW2. The storage function of schooling, its capacity to keep part of the age group away from the labour market, was actualised in Finland in a drastic way. The policymakers reacted decisively and the late but quick expansion of the education system began. In the second case of Master’s-level teachers, the policymakers were rather passive and finally agreed to the decision made by the higher decision makers as a part of a bigger higher-education reform. The implementation of evaluation-based governance in comprehensive schooling was quite similar. The policymakers concerned did not see any alternatives as long as the decision on municipal autonomy was beyond their jurisdiction. What is similar in all these cases, however, is that there was certain freedom or *Spielraum* for the policy actors.

CONCLUDING REMARKS

Our aim in this chapter was twofold. On the one hand we emphasise the importance of ‘truths’ - i.e., consensual taking for granted and self evidence - in understanding the political history of education, and on the other we highlight the concept of contingency as fruitful theorisation for analysing the emergence of these truths.

We have argued, first, for an interpretation according to which no understanding of Finnish comprehensive schooling is possible without taking into account at least three deep-rooted national beliefs: the belief in schooling as an essential source of welfare, the belief in teachers as rather solid and stable suppliers of this common good, and finally the belief in schools as institutions that deserve a certain autonomy, trust and industrial peace free from trendy quality-assurance and evaluation systems. Secondly, we have tried to show that these beliefs have been constructed through historical processes in which both rational actors and coincidental factors have always met, converged and intertwined.

Hans Joas' construction of the ambiguous concept of contingency⁶ as dialectics between uncertainty and freedom, coincidence and *Spielraum* is useful. From the point of view of the actor, awareness of contingency means, on the one hand, that things are increasingly not necessary or impossible, and on the other that it is precisely this that not only makes the change possible but also acts for it. Thus and paradoxically, the concept opens up the field of meaningful action in today's seemingly chaotic and intricate world. One could even say that it may 'save' the agency in these complex and late-modern times. In this sense, it could be seen as essential within the neostructuralist project, a channel through which to bring subjectivity, history and meaning back to the discussion in the wake of postmodernist and poststructuralist nihilism (Frank, 1989; Heiskala, 2003).

What can we say, then, about the concept of contingency in relation to its Siamese twin comparative education research – under-theorisation and the unbearable narrowness of the national view – outlined at the beginning of this chapter? We could conclude that linear and causal explanations are not enough to enhance understanding of deviant trajectories in policy and politics, for example, not even in the case of one single nation state. Operating through functionalist and system models only, whether emphasising transnational or national trends and efforts or focusing solely on rational decisions and choices, does not give theoretically adequate instruments for comparative research.

Are the contingent events and conjunctions not exactly the facts and realities that, by definition, are incomparable in the strong sense implied at the beginning of this chapter? Yes, indeed. Therefore, contingency shall not be seen as an explanation but rather as a concept that should always be taken into account: there may always be an element of coincidence and freedom in human action. Elsewhere (Simola et al., forthcoming) we have tried to vitalise standard conceptions of comparative study such as path dependency and convergence by bringing in contingency. It seems that contingency may really have a certain analytical power in linking and bonding the dimensions of structure and agency in particular. This might not be as radical as it appears. It may merely be reminiscent of Bismarck's words, "Politics is the art of the possible", or the slogan of an ice-hockey coach, "Only the good teams are lucky"?

NOTES

- ¹ The only exception relates to kindergarten teachers, whose training was moved to the Bachelor's level in early 1990s.
- ² The Bachelor's Degree was re-introduced in Finnish universities in the late 1990s because of the Bologna process but, understandably, there was no real discussion on lowering the qualification demands of teachers at that time, although the whole idea behind the Bologna process was in that direction.
- ³ Act on Central Government Transfers to Local Government (Law 707/1992) and the Local Government Act (Law 365/1995)
- ⁴ In the case of Master's-level teacher education only Jaakko Numminen, Chief Secretary at the Ministry of Education in 1973–1995, conceded that, "Simola was perhaps right".
- ⁵ ... or these senior policymakers simply may represent the generation that never became familiar with the 'organised uncertainties' of risk management in the present-day world; see e.g., Power 2007
- ⁶ Franz Josef Wetz (1998) has presented seven different readings of contingency.

REFERENCES

- Aho, E., Pitkänen, K. & Sahlberg, P. (2006). *Policy development and reform principles of basic and secondary education in Finland since 1968*. Washington, D. C.: The World Bank, 12; 132
- Alapuro, R., Liikanen, I., Smeds, K. & Stenius, H. (Eds.) (1987) *Kansa liikkeessä* [The Nation in movement]. Vaasa: Kirjayhtymä.
- Alestalo, M. (1991). Julkisen sektorin laajentuminen [Expansion of the Public Sector]. In M. A. Salo (Ed.) *Sosiaalipolitiikka 1991* [Social Policy in 1991]. Helsinki: Sosiaalipoliittinen yhdistys. 7–36.
- Antikainen, A. (1990). The rise and change of comprehensive planning: The Finnish experience. *European Journal of Education*, 25(1), 75–82.
- Ball, S. J. (2001a). Global policies and vernacular politics in european education opening address to the National Conference for Finnish Educational Research in Turku, Finland, November 23–25th 2000. In A. Jauhiainen, R. Rinne & J. Tähtinen (Eds.), *Koulutuspolitiikka Suomessa ja ylikansalliset mallit*. Published in Finnish “Gloaalit toimintaperiaatteet ja kansalliset politiikat eurooppalaisessa koulutuksessa”. Suomen kasvatustieteellinen seura: Kasvatusalan tutkimuksia 1. Turku: Painosalama Oy, pp. 21–44.
- Ball, S., J. (2001b). Global policies and vernacular politics in education. *Curriculo Sem Fronteiras*, 1(2), xxvii–xliii.
- Beck, U. (2006.). *The cosmopolitan vision*. Cambridge: Polity Press.
- Becker, H. S. (1994). “FOI POR ACASO”: Conceptualizing Coincidence. *Sociological Quarterly*, 35(2), 183–194.
- CEC. 2002. *Committee for Education and Culture*, SiVM 11/2002.
- Charle, C., Schriewer, J. & Wagner, P. (2004.). *Transnational intellectual networks: forms of academic knowledge and the search for cultural identities*. Frankfurt: Campus
- Clark, B. R. (1993). *Diversification of higher education: viability and change*. A paper presented at conference on Diversification in higher education. *The Second Conference on Policy Change in Higher Education*. 13–15 August 1993, Research unit for the sociology of education, University of Turku, Finland.
- Conrad, C. (2006). Vorbemerkung. Thema: Sozialpolitik transnational. *Geschichte und Gesellschaft*, 32(4), 437–444.
- Cowen, R. (2009). The transfer, translation and transformation of educational processes and their shape-shifting? *Comparative Education*, 45(3), 315–327.
- Dale, R. (2009). Studying globalisation and Europeanisation in education: Lisbon, the Open Method Coordination and beyond. In R. Dale, & S. Robertson (Eds.), *Globalisation and Europeanisation in education* (pp. 121–140). Oxford : Symposium Books.
- EURYDICE. 2004. *Evaluation of schools providing compulsory education in Europe*. European Commission. Directorate-General for Education and Culture. Brussels: Eurydice, European Unit.
- Fairclough, N. & Wodak, Ruth (2008) The Bologna Process and the knowledge-based economy: a critical discourse analysis approach. In Jessop, B., Fairclough, N., & Wodak, R. *Education and the knowledge-based economy in Europe*. Rotterdam: Sense.
- Frank, M. (1989). *What is neostructuralism?* Minneapolis: University of Minnesota Press.
- Hautamäki, J., Harjunen, E., Hautamäki, A., Karjalainen, T., Kupiainen, S. & Laaksonen, S. (Eds.). (2008). *PISA06 Finland: analyses, reflections and explanations*. Helsinki: Ministry of Education.
- Heikkinen, S., Silvonon, J. & Simola, H. (1999). Technologies of truth: Peeling Foucault’s triangular onion. *Discourse*, 20(1), 141–157
- Heiskala, R. (2003). *Society as semiosis: neostructuralist theory of culture and society*. Frankfurt am Main: Verlag Peter Lang.
- Helsingin Sanomat*, 11 February, 2004
- Hirvi, V. (1996). *Koulutuksen rytmivaihdos. 1990-luvun koulutuspolitiikka Suomessa* [Changing Pace of Education. Education Policy of the 1990s in Finland]. Helsinki: Otava.

- Hobsbawm, E. J. (1990.). *Nations and nationalism since 1780: programme, myth, reality*. Cambridge: Cambridge University Press.
- Joas, H. (2004). Morality in an Age of Contingency. *Acta Sociologica*, 47, 392–399.
- Joas, H. (2008). Kontingenssin aikakausi [The era of contingency]. *Sosiologia*, 45(3), 203–212.
- Joas, H., & Knöbl, W. (2009.). *Social theory: twenty introductory lectures*. New York: Cambridge University Press
- Jussila, J. & Saari, S. (Eds.) (2000). *Teacher education as a future-moulding factor. International evaluation of teacher education in Finnish universities*. Publications of Finnish Higher Education Evaluation Council 9: 2000. <http://www.minedu.fi/minedu/education/finheec/finheec.html>, read by 12.5.2004
- Kansanen, P. (2003). Teacher education in Finland: Current models and new developments. In B. Moon, L. Vlăsceanu, & C. Barrows (Eds.), *Institutional approaches to teacher education within higher education in Europe: Current models and new developments*. Bucharest: Unesco – Cepes, 85–108.
- Karisto, A., Takala, P. & Haapola, I. (1998). Matkalla nykyaikaan: elintason, elämäntavan ja sosiaalipolitiikan muutos Suomessa [On the way to the modern: changing standard of living, way of life and social policy in Finland]. Porvoo: WSOY.
- Kerr, S. (forthcoming) Suomen koulutusmenojen ja väestön koulutustason kehitys 1900-luvulla [Developments of Finnish Schooling Cost and Educational Level of the Population in 20th Century]. In P. Kettunen & H. Simola (Eds.) *Suomen kasvatuksen ja koulutuksen historia. Osa III: Tiedon ja osaamisen Suomi*. [History of Education and Schooling in Finland. Part III: Finland of Knowledge of Know-How] Helsinki: SKS.
- Kettunen, P. (2008.). *Globalisaatio ja kansallinen me: kansallisen katseen historiallinen kritiikki*. [Globalisation and us: a historical critique of the national gaze] Tampere: Vastapaino.
- Kettunen, P. (forthcoming). The National Welfare State as Transnational Historical Construction. In K. Petersen, & P. Kettunen (Eds.), *Beyond the Welfare Models*. Cheltenham: Edward Elgar.
- Kivinen, O. (1988). Koulutuksen järjestelmäkehitys [The System Developments of Education]. Turku: Turun yliopiston julkaisuja C:67.
- Kivinen, O. & Rinne, R. (1995). Vertailevan korkeakoulututkimuksen metodologisia haasteita [Challenges of comparative studies higher education]. In J. Nieminen (Ed.) *Menetelmävalintojen viidakossa. Pohdintoja kasvatuksen tutkimuksen lähtökohdista* [In the jungle of methods]. Tampere: Tampereen yliopiston kasvatustieteiden laitoksen julkaisuja B:13, 227–244.
- Kivinen, O., Rinne, R. & Ahola, S. (1989). Koulutuksen rajat ja rakenteet [The limits and structures of education]. Helsinki: Hanki ja jää.
- KM 1974:101. *Vuoden 1973 opettajankoulutuskomitean välimietintö* [The preliminary report of the Teacher Education Committee of 1973].
- KM 1975:75. *Vuoden 1973 opettajankoulutustoimikunnan mietintö* [The final report of the Teacher Education Committee of 1973].
- Kosonen, P. (1998). Pohjoismaiset mallit murroksessa [Changing models of Nordic countries]. Tampere: Vastapaino.
- Kumpulainen, T. (2009). *Opettajat Suomessa 2008*. [The teachers in Finland 2008]. Helsinki: Opetushallitus.
- Kumpulainen, T. & Saari, S. (2005). *Opettajat Suomessa 2005*. Helsinki: Opetushallitus.
- Kupiainen, S., Hautamäki, J. & Karjalainen, T. (2009). *The Finnish education system and PISA*. Helsinki: Ministry of Education Publications 46. <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2009/liitteet/opm46.pdf?lang=en>, read by 25.6.2010
- Mann, M. (1986). *The sources of social power. A history of power from the beginning to A.D. 1760*. New York: Cambridge University Press.
- Mann, M. (1993). *The sources of social power. The rise of classes and nation-states, 1760–1914*. New York: Cambridge University Press.
- Marques Cardoso, C. (1998). *Reforms of School Management and Governance: Portugal and England 1986–1996*. London: King’s College.

- Meyer, J. W. (1986). Types of explanation in the sociology of education. In J. Richardson G. (Ed.), *Handbook of theory and research for the sociology of education* (pp. 341–359). Westport, Conn: Greenwood Press.
- Marginson, S. & Mollis, M. (2001). The door opens and the tiger leaps: Theories and reflexivities of Comparative Education for a Global Millennium. *Comparative Education Review*, 45(4), 581–615.
- Nóvoa, A. & Yariv-Mashal, T. (2003). Comparative research in education: a mode of governance or a historical journey? *Comparative Education*, 39(4), 423–439.
- OECD (2002). *Education at a Glance*. Paris: OECD.
- OECD (2006). *Education at a Glance*. Paris: OECD.
- OPH (2009). *Yliopistojen hakija- ja opinto-oikeusrekisteri HAREK* (Excel-taulukko 29.9.2009). Helsinki: Opetushallitus.
- OPM (2010). *Kota-tietokanta* [The KOTA Database]. Helsinki: Opetusministeriö. <https://kotaplus.csc.fi/online/Etusivu.do> (read by 8.4.2010).
- Ozga, J. & Jones, R. (2006). Travelling and embedded policy: the case of knowledge transfer. *Journal of Education Policy*, 21(1), 1–17.
- Ozga, J. & Lingard, B. (2007). Globalisation, education policy and politics. In B. Lingard, & J. Ozga (Eds.). *The RoutledgeFalmer Reader in Education Policy and Politics*. Abdingdon: Routledge.
- Palonen, K. (1993) Introduction: from policy and polity to politicking and politicization. In K. Palonen & T. Parvikko (Eds.) *Reading the Political: Exploring the Margins of Politics* (pp. 6–16). Helsinki: The Finnish Political Science Association.
- Palonen, K. (2003). Four times of politics: Policy, polity, politicking, and politicization. *Alternatives*, 28(2), 171–186.
- Popkewitz, T. S. (2009). National imagineries, the indigenous foreigner, and power: Comparative educational research. In J. Schriewer (Ed.), *Discourse formation in comparative education* (3rd rev. ed.) (pp. 261–294). Frankfurt am Main: Peter Lang.
- Power, M. (2007). *Organized Uncertainty: Designing a World of Risk Management* Oxford University Press.
- Pöntinen, S. (1983). *Social mobility and social structure: a comparison of Scandinavian countries*. Helsinki: Societas scientiarum Fennica.
- Ragin, C. (1987). *The Comparative Method. Moving behind Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, C. (1989). New Directions of Comparative Research. In M.L. Kohn (Ed.) *Cross-National Research in Sociology* (pp. 57–76). Mewbury Park, CA: Sage.
- Ragin, C. (1992). Casing and the process of social inquiry. In C. Ragin & H. Becker (Eds.) *What is a Case? Exploring the Foundations of Social Inquiry*. New York: Cambridge University Press.
- Ramirez, F. & Boli-Bennett (1982). Global pattern of educational institutionalization. In P. G. Altbach, R. F. Arnove & G. P. Kelly (Eds.) *Comparative Education*. New York: Macmillan.
- Rinne, R. (1984). *Suomen oppivelvollisuuskoulun opetussuunnitelman muutokset vuosina 1916–1970* [The Changes on Curricula of the Compulsory School between 1916 and 1970]. Turku: Turun yliopiston julkaisu C:44.
- Rinne, R. & Salmi, E. (1998). *Oppimisen uusi järjestys* [The New Order of Schooling]. Tampere: Vastapaino.
- Rinne, R., Simola, H., Mäkinen-Streng, M., Silmäri-Salo, S. & Varjo, J. (forthcoming) *Arvioinnin arvo. Suomalaisen perusopetuksen laadunarviointi rehtoreiden ja opettajien silmin* [The value of Evaluation. The principals' and teachers' view on QAE in basic schooling in Finland]. Research in Educational Sciences. Finnish Educational Research Association.
- Schriewer, J. (2006). Comparative social science: characteristic problems and changing problem solutions. *Comparative Education*, 42(3), 299–336.
- Schriewer, J. (2009). Preface to the third edition. In J. Schriewer (Ed.), *Discourse formation in comparative education* (3rd rev. ed., vii–xii). Frankfurt am Main: Peter Lang.

H. SIMOLA AND R. RINNE

- Schriewer, J. & Martinez, C. (2004). Constructions of Internationality in Education. In G. Steiner-Khamsi (Ed.), *The global politics of educational borrowing and lending* (pp. 29–53). Teachers College Press.
- Simola, H. (1993). Educational Science, the State and Teachers. Forming the corporate regulation of teacher education in Finland. In T. S. Popkewitz (Ed.), *Changing patterns of power: Social regulation and teacher education reform in eight countries* (pp. 161–210). Albany, NY: State University of New York Press.
- Simola, H. (1995). *Paljon vartijat. Suomalainen kansanopettaja valtiollisessa kouludiskurssissa 1860-luvulta 1990-luvulle.* (The Guards of Plenty. The Finnish school teacher in state educational discourse from the 1860s to the 1990s). Helsinki: Helsingin yliopisto. Opettajankoulutuslaitos. Tutkimuksia 137.
- Simola, H. (1998). Firmly Bolted into the Air: Wishful rationalism as a Discursive Basis for Educational Reforms? *Teachers College Record*, 99(4), 731–757. (Republished in Ball, S. J. (Ed.) (2000). *Sociology of Education. Major Themes, vol. IV* London: Routledge / Falmer. 2112–2138.)
- Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41(4), pp. 455–470.
- Simola H. (2009) Trans-national technologies, national techniques and local mechanisms in Finnish university governance: a journey through the layers. *Nordic Educational Research*, 29(1), 6–17.
- Simola, H., Kivinen, O. & Rinne, R. (1997). Didactic closure: Professionalization and pedagogic knowledge in Finnish teacher education. *Teaching and Teacher Education*, 13(8), 877–891.
- Simola, H., Heikkinen, S. & Silvonon, J. (1998). Catalog of possibilities: Foucaultian history of truth and education research. In T. S. Popkewitz & M. Brennan (Eds.), *Foucault's challenge: Discourse, knowledge, and power in education* (pp. 64–90). New York: Teachers College Press.
- Simola, H., Rinne, R., Varjo, J., Kauko, J. & Pitkänen, H. (2009). Quality Assurance and Evaluation (QAE) in Finnish Comprehensive Schooling – a national model or just unintended effects of radical decentralisation? *Journal of Education Policy* 24(2), 163–178.
- Simola, H., Varjo, J. & Rinne, R. (forthcoming) Against the Flow – Path dependence, convergence and contingency in understanding the Finnish QAE model. *Education et sociétés. Revue internationale de sociologie de l'éducation*.
- Smith, A. D. (1995). *Nations and nationalism in a global era*. Cambridge: Polity Press.
- Steiner-Khamsi, G. (Ed.). (2004). *The global politics of educational borrowing and lending*. Teachers College Press.
- Strange, S. (1997). The future of global capitalism: or, will divergence persist forever? In C. Crouch, & W. Streeck (Eds.), *Political economy of modern capitalism mapping convergence and diversity*. London: Sage.
- Sweeting, A. E. & Morris, P. (1993). Educational reform in post-war Hong Kong: Planning and crisis intervention. *International Journal of Educational Development*, 13(3), 201–216.
- Temmes, M., Ahonen, P. & Ojala, T. (2002). *Suomen koulutusjärjestelmän hallinnon arviointi* [Evaluation of the Finnish education administration] Helsinki: Opetusministeriö.
- Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia a century of public school reform*. London: Harvard University Press.
- Werner, M. & Zimmermann, B. (2006). Beyond comparison: 'Histoire Croisée' and the challenge of reflexivity. *History and Theory*, 45(1), 30–50.
- Wetz, F. J. (1998) Kontingenzt der Welt – Ein Anachronismus, In G. v. Graevenitz & O. Marquard (Eds.) *Kontingenzt. Poetik und Hermeneutik XVII*. München: Fink., 81–106.

Hannu Simola,
University of Helsinki

Risto Rinne,
University of Turku

DANIEL TRÖHLER

CONCEPTS, CULTURES AND COMPARISONS

PISA and the Double German Discontentment¹

In the October/November issue of 2002 the University of Heidelberg newsletter *Unispiegel* announced a series of public lectures dedicated to the question: “Are we still a people of poets and thinkers?”, with the subtitle providing the information that the university’s *Studium Generale* lecture series in winter semester 2002/2003 would focus on educational questions (*Bildungsfragen*) (Unispiegel, 2002). Ten different scholars were invited, including even one scholar from abroad, as the announcement proudly emphasized. The speakers were philosophers, historians, politicians and writers – and none of them were from the educational sciences.

The *Unispiegel* announcement of this series of lectures gives the poor German PISA results presented to the public a year previously in 2001 (Deutsches PISA-Konsortium, 2001) as the reason for this initiative. Also, the announcement states that these results have alarmed Germany and that all over the country causes and culprits were being sought after and identified and that a lot of reform ideas were being formulated: “Although the bad ranking of the PISA survey concerns the realm of education and schooling, the self-doubts go far beyond. A whole nation wonders: Are we still the people of poets and thinkers?” (Unispiegel, 2002, para. 1). PISA, in other words, had shattered the national prospect of idiosyncratic singularity, the residual identity of a country with a troubled history.

The estimation that PISA signified a cultural crisis was no illusion. No other country has reacted to PISA as fiercely as Germany. It is illuminating to go to the different country websites of Amazon.com – www.amazon.de, www.amazon.fr, www.amazon.co.uk. If you enter “PISA” at www.amazon.co.uk, you find games, sandals, novels and guidebooks but nothing on the OECD survey. If you search in www.amazon.fr, the same holds for French publications; however, there are at least two English and one German publication on the survey, and in Spain, where there is no Amazon.com but www.fnac.es, there is one single book to buy, the official OECD report of the 2006 survey in Spanish. In Portugal, there is no publication about the survey at all. The picture changes dramatically when you search for PISA in Germany (www.amazon.de): You find over a dozen publications connected to the survey, including the official publications, in-depth analyses, PISA for small children, PISA for adults and PISA training programs to enhance knowledge; even a crash course on PISA has been produced. In other words: In Germany, PISA has obviously created a market, for it has created customers with specific demands and specific supply.

D. TRÖHLER

In this paper I want to analyze the emergence of these broad public and academic discussions and a market by treating them as a specific cultural phenomenon. My general thesis is that this manifold phenomenon arises out of a situation that was caused by the clash of two very different cultural self-understandings – which are of a religious nature, in the end, and the result is a cultural disorientation that triggers off the emergence of this PISA market. I am going to demonstrate my thesis in five steps. First, I introduce the German debate by looking at the concepts (1.); then I identify the clash of cultures as the background of the debate (2.). In the third step I do some archaeology on the PISA ideology (3.) in order to excavate a fundamental problem of PISA (4.). Finally, I interpret the German PISA dispute as a double discontent, resulting from different religious aspirations (5.). Before I start, though, let me say in advance that in Germany as a rule you are either pro PISA or against it, and if you oppose it, you usually do so by defending an ideal called *Bildung*. I will take a more differentiated standpoint – but that is the risk of asking a Swiss and not a German to discuss the phenomenon of PISA in Germany.

CONCEPTS IN A DOMESTIC DEBATE: COMPETENCE – *BILDUNG* – KNOWLEDGE

At the centre of the PISA survey there is a distinction that caused some confusion in Germany. This distinction is the one between different characteristics of knowledge, namely, the distinction between useful and useless knowledge. There is knowledge that is “merely learned,” and there is learned knowledge than can be used in the (future) lives of students (OECD, 2001, p. 14). Because PISA wants to look at “young people’s ability to use their knowledge and skills in order to meet real-life challenges,” the focus is not on what students learn at school on the basis of their curriculum and textbooks (p. 16). “Assessments that test only mastery of the school curriculum can offer a measure of the *internal* efficiency of school systems. They do not reveal how effectively schools prepare students for life after they have completed their formal education” (p. 27).

The German translation, of course, follows this fundamental distinction. It says – and here I need to retranslate from the German-language version of the OECD report – that PISA does not merely focus on knowledge learned at school but on how students can apply this knowledge (OECD, 2001a, p. 14). In the German text this ability is called *Grundbildung*, and it is specified by the addition in parentheses of the English notion “literacy,” but it is also called *Kompetenz* (competence) in the German text (p. 16 f.). *Kompetenz* (competence) is the German conceptual frame within which PISA operates in Germany, and it opposes – at least to a certain degree – the concept of knowledge (*Wissen*). In this respect the German translation is quite consistent with the English version.

However, this consistency with the English version is superficial. In Germany, there is indeed a long tradition that marginalizes school knowledge as mere knowledge, but this alienation occurs for a totally converse reason than that of the OECD survey. The overall German cultural marginalization of knowledge can be easily seen in the absence of the headword *Wissen* (knowledge) in relevant German-language educational encyclopaedias and reference works. These works

do not even refer to related headwords such as *Kenntnisse* (knowledge, skills); and in those instances where *Wissen* is discussed, it only appears in compound words that can be found in numerous encyclopaedias, such as *Wissensbegierde* (desire for knowledge). This shift to the inner attitude of the learner is no coincidence and reflects the educational concept that draws the most attention in the German discussion, namely, the notion of *Bildung*. Knowledge is by nature in deficit; *Bildung*, in contrast, is the aim.

- In other words, both the OECD and the German tradition marginalize knowledge in contrast to something else, and this “something else” is called competence within the linguistic world of PISA in Germany and *Bildung* in the much older German tradition. This shared enmity against mere knowledge suggests of course the merging of competence and *Bildung*. And indeed, the assertion that competence is in fact basically *Bildung* – and that their common enemy “mere knowledge” – is made explicitly. In their outline on developing national education standards, the German PISA experts assure us that “‘Competencies’ describe nothing other than those individual skills that had been indicated by the concept of *Bildung*” (Klieme et al., 2003, p. 65, see also p. 66). Unfortunately, and in contrast with the concepts of competencies and standards, “concept of *Bildung*” is not elucidated at all, but in another publication by the same PISA experts we find the names of Wilhelm von Humboldt and Wilhelm Flitner, and with them some very general references to the German concept of *Bildung* (Deutsches PISA-Konsortium, 2001, p. 21). It is important to note, too, that the merging of competencies and *Bildung* is not solely an act by historically blind empiricists (some of the so-called empiricists were initially trained in history, too), and Heinz-Elmar Tenorth, a genuine historian of education, did the very same thing: “*Bildung* and literacy, basic skills and modes of handling higher culture do not depict disjunctive classes of knowledge and behavioural patterns but specific developments of a single and identical dimension of human practice” (Tenorth, 2008, p. 29).

So much for what may develop from of a common enmity, but whether or not this marriage of competences and *Bildung* was sustainable must be analyzed. The critics of this alliance between competencies and *Bildung* were, at any rate, not a long time coming. One journal in education (see Rekus, 2007) organized a special issue by inviting scholars to discuss the question of whether or not competence is indeed simply a new notion for the concept of *Bildung* (“Kompetenz – ein neuer Bildungsbegriff?”), and the result of this discussion is devastating for the PISA consortium – admittedly, the consortium did not participate in the discussion. The general critique discloses not only a fundamental difference between *Bildung* and competence but also a fundamental hierarchy between the two, for the latter is identified as a decline of true culture. It is interesting to note that the very same Wilhelm von Humboldt that had been used by the PISA consortium to legitimize their concept of competence is now being used for exactly the opposite purpose. Manfred Sieburg, for instance, says that one of the enormous merits of Wilhelm von Humboldt was that he had “succeeded at breaking the unfortunate chain” between education on the one hand and adjustment on the other by proclaiming

D. TRÖHLER

Bildung as immeasurable occurrence in the inner person, and that with the PISA ideology the schools would basically become training institutions again, with the aim to adjust students to the existing environment (Sieburg, 2007, p. 189).

This sharp dualism between inward *Bildung* on the one hand and simple adjustment to the existing world on the other hand was in fact the basic inspiration for the lecture series at the University of Heidelberg in 2002/2003, defending Wilhelm von Humboldt against the aspirations of the PISA ideology. The philosopher Brigitte-Sophie von Wolff-Metternich reminded the public that “*Bildung* ... is not codifiable and fixable knowledge – neither theoretically nor practically” (Wolff-Metternich, 2004, p. 68), not utilitarian nor pragmatic (p. 69), and therefore principally purposeless (p. 71). And compared to the Humboldtian theory of *Bildung*, Wolfgang Frühwald, professor of literary studies, identified the basic assumptions of PISA by even using a medical metaphor, as being the “cancer” of a “*value-for-money ideology*” – and he wrote this in English (Frühwald, 2004, p. 42). *Bildung*, as another prominent critic stated, is exactly the opposite of this “*value-for-money ideology*”, for it indicates the inward formation of a human being that in the end is called a *Persönlichkeit* (Herrmann, 2007, p. 172). The *Persönlichkeit* as result of *Bildung* is the self-sufficient mature and harmonious person, whereas PISA and its program intend to incapacitate humans in order to train them to be obedient *homo oeconomicus* (Krauz, 2007). In other words, the suggestion to mate competence and *Bildung* caused irritation and raised scepticism.

THE CLASH OF CULTURES

It would probably have been helpful to the protagonists of this domestic debate to recognize that a very similar discussion occurred a century ago, when American Pragmatism, as the first non-European philosophy, began to be discussed on the old continent. Whereas American Pragmatism was received with some interest or even sympathy in different European countries such as Britain, the Netherlands or, most of all, the French-speaking part of Switzerland, it was with very few exceptions harshly rejected in Germany. One of the major reasons for this rejection was William James’ theory of truth, in which James had identified the idea of truth as a contingent function in the process of thinking – and not as an eternal idea, as German idealism did. “‘The true’ to put it very briefly, is only the expedient in the way of our thinking, just as ‘the right’ is only the expedient in the way of our behaving” (James, 1907, p. 86). Truth, in other words, is not the aim of thinking and research but its means; it is a tool of the human practice and not its transcendent aspiration. In the normative horizon of a dualistic German philosophy, this very identification of truth and utility was a slap in the face, and because James did not use his metaphors with caution he added even more fuel to the German fire, by talking about the “truth’s cash value” (p. 77). The German intellectuals pronounced sentence quickly: Pragmatism was an abject philosophy; it was labelled a “dollar philosophy” and a “despicable kitchen and handyman utilitarianism” that did not hesitate to sell truth for cash (Spranger, 1915/66, p. 37).

Within the context of the William James discussion in Germany, a German philosopher named Jacoby (1912) replied to William James' characterization of the educational goals of European and, most of all, British universities in *Talks to Teachers* (1899), which was published in German translation 1900. Jacoby took up the notion of the *Persönlichkeit* that is praised today against the PISA ideology (see above) and wrote:

The German university does not make it its task to teach a German *Herr* how to *behave* like a German *Herr*. In our tradition, that is exclusively a matter for the nursery. In contrast, the German university, to an outstanding extent, makes it its task to educate the German student to become a *Persönlichkeit* – a fact that William James, of course, does not take into account but that is nonetheless important and true. England is the land of gentlemen; Germany is the land of *Persönlichkeiten*. Gentleman and *Persönlichkeit*, however, stand essentially in hostile opposition to one another. This does not at all mean that a gentleman cannot have something *Persönliches* about him or that a *Persönlichkeit* cannot be a gentleman. But the ideal of the gentlemen clashes with the ideal of the *Persönlichkeit*, and the ideal of the *Persönlichkeit* clashes with the ideal of the gentleman. (Jacoby, 1912, p. 217)

The perception of a national orientation within these arguments – if they are arguments at all – is not misleading. Eduard Spranger, one of the mandarins of German education in the 20th century and a critic of Pragmatism, had as early as in 1902 lamented the “inner corrosion” of Germany to an industrial state or – *horribile dictu* – a social democracy or even anarchy (which in Spranger's eyes was about identical with democracy) and promoted Fichte's national ideal of a “closed national *Bildung*” (Spranger, 1902/1973; see Tröhler, 2003). And it was the same Spranger who – in his fundamental opposition to the Western world and democracy – in 1928 propagated *Bildung* as the essential German alternative to the modern world. He identified the birthplace of this alternative in German classicism, at the time around 1800; it is the time of the German poets and thinkers, as Spranger emphasized: “We call our thinkers and poets German classics. They had ‘*Bildung*’ in the full plastic sense of the word, for they were not merely literary intellectuals. That is why they were masters of life and not its wageworkers” (Spranger, 1928, p. 11). The opposition of *Bildung* and knowledge is crucial:

First of all, it is evident that the meaning of *Bildung* is not an arbitrary sum of know-how and language knowledge, of social attitudes and political dispositions ... The meaning of *Bildung* is always personality [*Personalität*], that is, *Bildung* belongs to the human being insofar he is able to represent a unitary meaningful form as opposed to the manifold intellectual contents. It is this meaning that the classics [poets and thinkers] have discovered: The human being as a meaningful form in contrast to the materials of life, the human being as a unity against the multiplicity of the manifold sensual fields of life. (Spranger, 1928, p. 12)

The role of *Bildung* creating meaning is one of the crucial arguments in Spranger's criticism of Pragmatism, and it is one of the crucial arguments of

D. TRÖHLER

today's critics of PISA. Whereas Pragmatism – and here most of all George Herbert Mead – emphasized that meaning is created by social interaction and therefore changes in different contexts, the German idealist tradition insists on an internal instance of true meaning-making, and this inward instance is what *Bildung* is all about and how it makes the person the *Persönlichkeit*. Today, the critics argue against the concept of competence by claiming that the PISA ideology refers solely to utility while neglecting the dimension of the *Sinnhaftigkeit*, something like the “reasonability” in life (Rekus, 2007, p. 156). Against this background, the merging of *Bildung* and competence by the German PISA protagonists has to be rejected. Sieburg (2007) gets to the point by starting out with a quotation from what is referred to as ‘the Klieme report’:

‘The educational standards are oriented at the general aims of education, and in principle they are convertible (operationalizable) by tasks and test charts’ [Klieme et al., 2003]. This is the inner contraction: *Bildung*, regardless of which historical or present attempt of containment of this almost impossible to comprehend concept one wants to subscribe to, resists being operationalized; *Bildung* is the character of the *Persönlichkeit*, a never ending process; *Bildung* is the meta-useful. In other words, *Bildung* is unmeasurable. (p. 186)

THE EDUCATIONAL “SYSTEM”, ITS ENGINEERS AND COGNITION PSYCHOLOGY

From the critic's linguistic and thus ideological background, the merging of *Bildung* and competences, and thus the whole setting of PISA, cannot be interpreted as anything other than upsetting, for it represents the surrender of an entire tradition, of the German *Sonderweg*, the peculiarity of German History and its ideological and cultural background. Even worse, it is the almost unconditioned capitulation of a cultural ideal to a capitalist ideology, without any ethical value save the making of money. And it is true that against the background of this normative horizon in which *Bildung* plays a crucial role, PISA's advance is being identified as an unfriendly take-over that has to be firmly rejected.

And indeed, when we look at the emergence of PISA, there are good reasons to reject this attempt to merge competency and *Bildung*. As I argued elsewhere (Tröhler, 2010), PISA roots in the late 1950s, when the launch of the Sputnik triggered off the educationalization of the Cold War, as it was expressed, for example, by former US President Hoover's reaction to Sputnik:

The trouble is that we are turning out annually from our institutions of higher education perhaps fewer than half as many scientists and engineers as we did seven years ago. The greatest enemy of all mankind, the Communists, are turning out twice or possibly three times as many as we do (...) The harsh fact is that the high schools are not preparing youngsters for the entrance requirements which must be maintained by our institutions training scientists and engineers. (“Education,” 1957)

In the frame of this ideology, little by little the Cold War had become an encompassing *educational* reform project with single facets that ironically merged into one single new agenda only after the end of the Cold War in 1989. One of the facets was the National Educational Defense Act in 1958, with its emphasis on three school subjects that today may not be unfamiliar: mathematics, sciences and foreign languages – in other words, almost the trilogy that PISA is focusing on today. A second facet was the development of the human capital theory at that very same time, and still another facet was the foundation of the OECD in 1960. The first official OECD conference, held in Washington, D.C., in 1960, was devoted to the topic “Policy Conference on Economic Growth and Investment in Education” (OECD, 1961). However, the enemy was not only the Russians but also the educational ideology that was dominant in the United States at the time and supported by the philosophers of education and the powerful teachers’ unions; it was called the Life Adjustment doctrine. Admiral Hyman G. Rickover attacked this doctrine in the name of many:

If the local school continued to teach such pleasant subjects as ‘Life Adjustment’ and ‘How to know when you are really in love,’ instead of French and physics, its diploma would be, for all the world to see, inferior. Taxpayers will begin to wonder whether they are getting their money’s worth. (“Education”, 1957)

Against this background, the educationalization of the Cold War in the United States marked a transformation of the dominant reference discipline for education, for it switched from philosophy to psychology, more precisely from a popular interpretation of Pragmatism to cognitive psychology, which was at its outset in the late 1950s – cognition theory being the most important academic reference of PISA today, as the stakeholders admit themselves (Klieme et al., 2003, pp. 23–26; Deutsches PISA-Konsortium, 2001, p. 22). The rise of cognitive psychology came along with the rise of new governance ideologies of the Cold War. These governance ideologies arose on the background of a specific historic model – the effective model of problem solving by collaborating military, scientific, and political experts during the Second World War at Massachusetts Institute of Technology (MIT) and in the context of the Manhattan Project at Los Alamos. In the superb book, *Scientists in the Classroom*, John Rudolph describes in detail how in the eyes of this ideology the idea arose that nearly any problem could be solved by cooperation of first class experts, for it had been groups of scientists, for example, who had successfully solved complex problems with radar during the war in order to detect German submarines and who developed the atomic bomb being used in Japan (see Rudolph, 2002, p. 90).

This idea of contract research became the model of efficient research for the sake of the nation defending freedom, welfare and peace, for the sake of all people, in other words – an idea that is being applied again within the OECD, PISA and other large-scale assessments. However, the shift from self-defined research to contract research implied a shift of terminology, and the main actor, the researcher, became an expert, who understood predefined problems as a complex setting of different elements constituting what was called a “system”. The scientific background

of this system perspective largely disengaged the experts from cultural constraints; they focused less on understanding in what way a system is a cultural construction or how the system works as a system and instead defined it departing from the idea of the best possible mutual arrangement of its identified elements. In other words, the systems perspective was the engineering perspective, and this perspective focused “not just on the optimum performance of a given human/technological system” but on “the entire array of possible alternatives that might be created by using existing or newly developed technologies ... from scratch” (Rudolph, 2002, p. 94).

It is exactly this idea of interpreting problems as systems and finding solutions “from scratch” – in other words, disregarding the contexts in which systems are constructed and operating – that gave cognitive psychology entry into educational governance, for cognition theory interpreted cognitive data processing in the language of mathematics, defining intellectual solution procedures using the mathematical template of algorithms. In this way, the human mind had become a computing machine that had to be maintained and supported like a complex computer. This understanding of the procedures within the human mind ideally fitted the technological systems perspective, and optimism in the feasibility of the one safe free and prosperous world grew again. The new model entered into educational governance quickly. Already at the Woods Hole Conference in 1959, where the problems of US education were discussed against the background of Sputnik, Jerome Bruner reconciled his cognitive psychology with the systems engineering perspective. Bruner noted that in order to discuss the problem of the US educational system “we introduced this subject ... by suggesting the analogy to a weapon system – proposing that the teacher, the book, the laboratory, the teaching machine, the film and the organization of the craft might serve together to form a balanced teaching system” (Rudolph, 2002, p. 99). The experts at the conference had agreed that “the goals of education ... expressed in terms of the human functions and tasks to be performed ... can be as exactly and objectively specified as can the human functions and tasks in the Atlas Weapon System” (p. 99). This ideology took 40 years to become globally dominant, ironically, however, only after the big enemy, communism, had eclipsed.

SOMETHING ABOUT THE REAL WORLD AND ITS CHALLENGES

Comparing the ideological roots of both the PISA experts and PISA critics, makes the harsh rejection in Germany of the merging of competency and *Bildung* appropriate and the series of lectures at the University of Heidelberg a matter of course. But why, we might ask, have the PISA experts been so foolish in trying to merge these two so obviously contradictory concepts? Were they attempting to use a Trojan Horse strategy, pretending that PISA was in fact what Humboldt and the German theory of *Bildung* actually wanted? Or are they just not *gebildet*, and thus confused about core concepts? There is some evidence of this, after all, for PISA uses variable and sometimes contradictory concepts: Does PISA in fact assess the performance of students (Deutsches PISA-Konstortium, 2001, p. 11), or the basic competencies of the next generation, or the student’s literacy and skills (p. 15), or the educational system (Klieme et al., 2007, p. 11)? All this is asserted – and it

makes PISA a real sitting duck for the rigorous school of both German idealism and critical theory, two ideological threads that are usually not very much in accord with each other.

Indeed, it is pretty simple to attack PISA's use of concepts as influenced by a rather meagre educational theory that was developed in the context of the human capital theory. The argument that PISA is more an empirical than a theoretical approach does not hold, first of all because I do not understand how you can count without knowing what and why you are counting. What is even worse is that PISA is in fact not as empirical as it pretends to be, which brings it closer to the non-empirical German ideology of *Bildung* and which thus supports the merging of competence and *Bildung*. When PISA looks at "young people's ability to use their knowledge and skills in order to meet real-life challenges", the focus is obviously not directed at what students learn at school on the basis of their curricula and textbooks (OECD, 2001, p. 16). In an irritating way these PISA "real-life challenges" are anything but the students' school life, and beyond that they are not only outside of school, but they are also situated in "life after" compulsory education. In other words, PISA does not ask how students master their own lives but speculates about the mastery of a future life: "Assessments that test only mastery of the school curriculum can offer a measure of the *internal* efficiency of school systems. They do not reveal how effectively schools prepare students for life after they have completed their formal education" (p. 27). The German translation says even more explicitly that PISA wants to test the ability of the different school systems to prepare students for life (PISA, 2001a, p. 30), for life as adults (Deutsches PISA-Konsortium, 2001, p. 17) – as if students were not living at all as empirical entities in the here and now. Even more irritating is the fact that PISA takes a rapidly changing world as cause of the assessment itself, at the same time pretending to know what skills will be necessary in a changed world in 10, 20 or 30 years. How can they know what kind of skills will be necessary, if they are so sure about how quickly the world changes?

I have no trouble believing that competencies that you acquire in your life can be useful later, although the success will always be at risk. But the basis of a successful later life is success in your present life, for you develop, differentiate, and adapt your skills or competencies through the learning effects of your interactions. Allow me to go back to my own youth, knowing full well that biographical introspection is certainly not the primary concern of cognitive psychology, but it gives you an idea of what real real-life challenges may be for a youngster at the age of 15. In short, the challenge is to gain esteem and recognition within the peer group and to avoid disturbing troubles at home. This latent tension may demand the highest skills. One example: At the time when I was young it was cool for young men to have long hair, and I was growing my hair. The challenge was to convince my parents, most of all my father, week after week why I did not have to go to the barber, and I had to find reasons and withstand financial baits and bear arguments, again and again. The longer my hair grew, the more respected I felt in the peer group, and the more troubles I had at home. Another thing was to get a moped and additionally to have a fast moped – bicycles were outmoded. However, the technical control mechanisms of mopeds were stipulated by the law,

D. TRÖHLER

so that the moped could not exceed 30 km/hour – an offensive speed limit in our eyes. So a real-life challenge was to develop skills to tune up the motors and to have – if ever possible – the fastest moped in town. Further skills were to convince parents, without lying too much, that spending the night with friends took place in a controlled family situation, and other skills were to get money from them to pay for music that they did not like. All these negotiations on and violations of family rules were – at least in the eyes of most parents that I knew – alright, as long as we did well at school, and doing well at school meant nothing other than succeeding at what PISA is not interested in: mastering the curriculum. In other words, the handling of private life challenges had to be paid in some sort of way by bringing home good grades, and this meant – despite what PISA says – learning the very concrete lessons in very concrete textbooks, corresponding to very concrete curricula. All these skills – the basic ones that we needed – are largely neglected by PISA, because it is obviously not interested in the present lives of students but in speculating on future lives.

THE INNER AND OUTER HARMONY AND THE DOUBLE DISCONTENT

PISA's peculiar non-empiristic empiricism is rooted in the original ideology of PISA, the Cold War ideology of the late 1950s. In this cultural milieu we find the vision to create a united harmonious world of free people. The slogan of this vision was "One World". It had been used as early as 1943 by US presidential candidate Wendell Lewis Willkie, and it conveyed the idea of a safe and united world based on the security and well-being of common people throughout the world, provided by US world leadership (Fousek, 2000, p. 79). Annoyingly, one of the former Allies, the Soviet Union, had expressed similar ambitions on its own agenda and had thus become more and more a distracting factor for the global vision of "One World" under the leadership of the United States. The Russians – by the way with a very similar technocratic and expertocratic ideology as the United States – were denounced as being ideological, whereas the United States was praised as being free of ideology, for it was seen to be in the hands of academic experts. The concept replacing the concept of ideology was development, development being the global expression of an ideology-free, expert-driven world.

The emphasis on development towards peace and freedom was by its very foundations religious, even missionary, as can be seen as early as in 1947 when former Vice-President of the United States, Henry A. Wallace, said: "By reason of history, geography and sheer economic strength America has it in her grasp to furnish that great and last peace which the prophets and sages have preached for thousands of years" (quoted in Fousek, 2000, p. 11). The religious language of salvation is not misleading but instead characteristic, as Denis Brogan, a British commentator upon the United States, noticed in 1957: "The notion of 'mission' is far wider than it was; the whole world is the parish of the United States as a government and a culture" (quoted in Gilman, 2003, p. 69). Or to quote President Harry S. Truman in 1949:

The United States is preeminent among nations in the development of industrial and scientific techniques. The material resources which we can afford to use for the assistance of other peoples are limited. But our imponderable resources in technical knowledge are constantly growing and are inexhaustible. (...) Greater production is the key to prosperity and peace. And the key to greater production is a wider and more vigorous application of modern scientific and technological knowledge (quoted in Gilman, 2003, p. 71).

The human capital theory and the OECD fit this ideology perfectly, and it is certainly no coincidence that at its first conference in Washington, D.C., in 1960, where the discussion topic was *Economic Growth and Investment in Education*, none of the keynote speakers was an educationalist (see for more details, see Tröhler, 2010).

Against this background, the exclusion of the real-life situations of students and thus of the curriculum from PISA becomes understandable. What students learn at school across the world is culturally contingent and disparate; however, the world according to PISA is the globally harmonized world of interaction: “PISA offers a new approach to considering school outcomes, using as its evidence base the experiences of students across the world rather than in the specific cultural context of a single country” (OECD, 2001, p. 27). But there exists nothing like the experiences of students across the world in contrast to experiences within the “specific cultural context of a single country”, for experiences are always situated within a specific cultural context. The neglecting of both the real-life situations of students and the culturally situated learning experiences makes it clear why PISA exponents in Germany aim at harmonizing competence and *Bildung*.

Both the PISA ideology and the German traditional *Bildung* ideology are non empirical by their foundations, and both are driven by the idea of a harmonious world as the goal of education. The German tradition aims at the harmonious inward *Persönlichkeit*, being able to give sense to the multiple outer world, and the PISA tradition aims at the harmonious “One World” of free, globally interacting and economically secure citizens. The ideological backgrounds of these two visions are not as alien as the heated debate in Germany might lead us to believe – they are both rooted in different denominations of Protestantism. The ideal of inward *Bildung* is based on Lutheranism and the ideal of the One World on Calvinism. It is no coincidence that around 1900, the president of the University of Chicago, William Raines Harper, originally a Baptist theologian, told his students that the fourth part of world history was then beginning and that it had its centre in the United States. Harper said that in this era civilization was reaching its apex: According to him, “the history of civilization has been synchronous with the development of a pure and true conception of God, and of his relation to man” – that is, the Baptist-Protestant interpretation of God and God’s relation to man. Harper saw this movement as a mandate for a mission that had been assigned to the United States by God and that had deep educational consequences, built on the “Gospel and education”. The Gospel and education would empower the United States to convert the world into one: “In this work of educating humanity to understand God and itself, America is the

D. TRÖHLER

training-school for teachers” (Harper, 1904, pp. 175 ff.). Milton’s Paradise, which had gone lost when the Puritans had to leave England in the 17th century, was finally to be regained under the leadership of purified models of freedom, democracy and prosperity based on technological and economical progress.

Against this background, the double discontent in Germany and the general upset becomes clearer. For some, PISA is the cultural catastrophe as such, for it is unmistakably situated in the outer world rather than focused on the inner world, which in turn means a betrayal of the idiosyncratic German tradition of *Bildung* and *Persönlichkeit*, a tradition of a country that through its peculiar history has a hard time defining and finding national identity. The educational idea of *Bildung* had in some way replaced what William Wallace alias Braveheart is for Scotland, William Tell for Switzerland, George Washington for the United States, Giuseppe Mazzini for Italy, Mahatma Gandhi for India, or the revolutionaries of 1789 for France – and against that background, calling *Bildung* into question means basically questioning the idea of the German nation. However, there was an analogue irritation on the side of the PISA experts, who realized how little the educational project of the “One World” has been achieved, of all things, in Germany, for in no other country were the differences in the PISA scores between the immigrants and the native students so striking as in Germany, indicating poor national unity and coherence. It is no coincidence that in contrast to the politicians, the German PISA experts placed less emphasis on the international ranking than on the catastrophic diversity of the German students based on their social and racial backgrounds. The educational system had failed in its task to integrate the immigrants and thus failed to contribute to the creation of harmony, here not of the inward person – which is an odd perception in the world of cognitive psychology, anyway – but of the outer world. It is this clash of cultures, so close and so alien at the same time, that made PISA an incomparable event in Germany. This clash is not without advantages, for at least it shows that on both sides an educational theory that might be called secular and thus academic has still a long way to go.

NOTES

- ¹ Paper presented at the international CESE conference “PISA under critical examination: Changing knowledge, changing tests, and changing schools” at the Teatro Chico in Santa Cruz de La Palma, Canary Islands, 23 to 26 November 2009

REFERENCES

- Deutsches PISA-Konsortium (Eds.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske+Budrich.
- Education: What price life adjustment? (1957, December 2). *Time Magazine*, LXX (23).
- Fousek, J. (2000). *To lead the free world. American nationalism & the cultural roots of the Cold War*. Chapel Hill, NC: The University of North Carolina Press.
- Frühwald, W. (2004). Kulturenstreit oder Von der Rolle der Sprache in der Wissenschaft. In W. Frühwald, J. Limbach, A. Schavan, D. Schipanski, M. G. Schmidt, L. Schorn-Schütte, et al., *Sind wir noch ein Volk der Dichter und Denker?* (pp. 35–45). Heidelberg: Universitätsverlag Winter.

- Gilman, N. (2003). *Mandarins of the future. Modernization theory in Cold War America*. Baltimore, MD: Johns Hopkins University Press.
- Harper, W. R. (1904). America as a missionary field. In W. R. Harper, *Religion and the higher life: Talks to students* (pp. 173–184). Chicago: Chicago University Press.
- Herrmann, U. (2007). “Bildung”, “Kompetenz” – oder was? Einige notwendige Begriffsklärungen. *engagement Zeitschrift für Erziehung und Schule*, 171–178.
- Jacoby, G. (1912). William James und das deutsche Geistesleben. *Die Grenzboten. Zeitschrift für Politik, Literatur und Kunst*, 71, 212–220.
- James, W. (1899). Talks to teachers on psychology: And to students on some of life’s ideals. New York: Henry Holt.
- James, W. (1995). *Pragmatism. A new name for some old ways of thinking*. Mineola, NY: Dover Publications. (First edition published 1907)
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., et al., (2003). *Zur Entwicklung nationaler Bildungsstandards*. Berlin: Bundesministerium für Bildung und Forschung.
- Krauz, J. (2007). Die Vereinnahmung der Person. Zu Auswirkungen und Hintergründen des Kompetenz-Konzepts. *engagement. Zeitschrift für Erziehung und Schule*, 211–225.
- OECD (1961). *Policy Conference on Economic Growth and Investment in Education. Washington, D.C., 16th - 20th October 1961*. Paris: OECD Publishing.
- OECD (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD Publishing.
- OECD (2001a). *Lernen für das Leben. Erste Ergebnisse von PISA 2000*. Paris: OECD Publishing.
- Rekus, J. (2007). Kompetenz – ein neuer Bildungsbegriff? *engagement Zeitschrift für Erziehung und Schule*, 155–160.
- Rudolph, J. (2002). *Scientists in the classroom: The Cold War reconstruction of American science education*. New York: Palgrave Macmillan.
- Sieburg, M. (2007). Kompetent oder gebildet? *engagement Zeitschrift für Erziehung und Schule*, 184–194.
- Spranger, E. (1902/1973). Gedanken zur Pädagogik. In O.-F. Bollnow & G. Bräuer (Eds.), *Eduard Spranger. Philosophische Pädagogik* (pp. 190–207). Heidelberg: Quelle & Meyer.
- Spranger, E. (1915/1966). Letter sent to Georg Kerschensteiner on 14 March 1915. In L. Englert (Ed.), *Georg Kerschensteiner [-] Eduard Spranger. Briefwechsel 1912–1931* (pp. 30–31). München/Wien.
- Spranger, E. (1928). *Der deutsche Klassizismus und das Bildungsleben der Gegenwart* (2nd ed.). Erfurt: Kurt Stenger Verlag. (First edition published 1926).
- Tenorth, H. E. (2008). Bildungstheorie angesichts von Basiskompetenzen. Über die Ignoranz gegenüber dem Selbstverständlichen im Prozess des Aufwachsens. *Zeitschrift für pädagogische Historiographie*, 14, 26–31.
- Tröhler, D. (2003). The discourse of German *Geisteswissenschaftliche Pädagogik* – A contextual reconstruction. *Paedagogica Historica. International Journal of the History of Education* XXXIX, 759–778.
- Tröhler, D. (2010). Harmonizing the education globe. World polity, cultural features, and the challenges to educational research. *Studies in Philosophy and Education* 29, 7–29.
- Unispiegel der Universität Heidelberg. (October-November 4/2002). Retrieved 21 March 2010 from <http://www.uni-heidelberg.de/presse/unispiegel/us02-4/index.html>
- Wolff-Metternich, B. S. von (2004). Was heißt heute: sich im Denken orientieren? In W. Frühwald et al., *Sind wir noch ein Volk der Dichter und Denker?* (pp. 65–76). Heidelberg: Universitätsverlag Winter.

Daniel Tröhler
 Faculty of Language and Literature, Humanities, Arts and Education
 University of Luxembourg

ROBERT COWEN

CODA

I attended, when I was a student in the Comparative Department of the Institute of Education in London, a seminar given by C. Arnold Anderson. He argued that we needed ‘international testing’ (of the IEA kind) on the grounds that until we knew the outcomes of educational systems we could not have ‘a comparative education’. I thought the argument was unusual. I was in the middle of spending several years – I had just finished an MA in comparative education – learning what comparative educationists normally thought about, and various ways of doing comparative thinking. No one had mentioned international testing and the measured outcomes of educational systems as a *sine qua non* for the epistemic existence of the field of study. Indirectly, C. Arnold Anderson had highlighted the epistemic positioning of comparative education in London. Joseph A. Lauwerys and other teachers in the department were concerned with understanding the cultural assumptions and philosophical ideas which inscribed conceptions ‘good knowledge’ in curriculum patterns, rather than, in standardised tests, measuring what children had retained of such knowledge

At the time, all of this was probably not too surprising. In England, the optimistic post-war expectations about the social benefits of the ‘objective’ testing of children had shifted into an increasingly harsh intellectual critique of the social consequences of the cultural biases in the standardised tests used for the selection of children for secondary schools – the so-called “11+” examinations. These examinations measured, among other things, IQ and decided the futures of children in state-financed schools, typically by allocating them to schools with different curricula that led into different adult occupations. In England, the contributions of mathematicians and scientists to war-time inventions (such as radar, improved submarine-tracking devices, the breaking of secret military codes, and theoretical contributions to the physics of the atomic bomb) had left the impression that the British were short of mathematicians and scientists; rather than experiencing some qualitative deficiency in terms of Soviet science. In England, it was also assumed that testing and large-scale empirical surveys were something that psychologists, sociologists and census-takers did: such activities were certainly not central to the intellectual work of comparative educationists.

Thus, the intersection of an epistemic tradition that went back to Sadler and Hans and the politics which framed the topics of English (and British) comparative education was not powerfully disturbed by the Cold War. The politics of the ‘agenda of attention’ in British comparative education were the post-1945 themes of opening up access to secondary education, educating sufficient teachers, expanding (or not

expanding - lest 'more means worse') the size of the university sector, and whether we had enough scientists and mathematicians to trigger economic growth through the application of science in the economy. Within educational studies and comparative education, the most serious fixation was with the institutional form of secondary education (separate schools, or a 'common school for all?'). Nigel Grant's work, before he found his 'Scottish voice', included an academic text on education in the USSR. This text is entirely comprehensible within the routines of narration used in the comparative education of the period: something on the history and contemporary politics, the economy, and the religious and cultural traditions of the country, along with detailed descriptions of the administration and finance of the system, school structures, curricula and models of the good pupil, teacher education, vocational education and higher education patterns.

In other words, the 'outcomes' which university comparative education in the UK was looking at were the varied patterns of educational systems themselves (including their school patterns, teacher education provision, vocational-technical education traditions, and university and higher education systems). Such configurations were – in the phrasing of Sir Michael Sadler - the result of 'intangible forces'. This theme, translated in the vocabulary of Nicholas Hans to understanding "factors" external to the schooling system - the geographic and economic circumstances, the interrelations of language and race, of religious belief and political philosophies - was the complex starting point of an emerging twentieth century tradition of comparative education. Of course those themes were also embedded in an implicit politics. 'Agendas of attention' of comparative education always contain an implicit (and sometimes an explicit) politics. For Hans these politics were the interwar struggle between fascism, communism and 'democracy' and more tacitly the characteristics of a stable and harmonious society in which contradictions between "the factors" were small (e.g. Denmark) and other societies in which contradictions in the mix of factors were potentially destabilising (e.g. Belgium; South Africa).

It was partly in counterpoint to such an epistemic frame that the argument of C. Arnold Anderson was sharp and disturbing. Of course, the idea – or the act - of counting was not in itself a shock in the English context. There had been a long tradition of 'counting' (notably of the poor in London and including a definition of the 'undeserving poor') from the time of the Tudor dynasty. The tradition, which became associated with the industrial northern cities as well as London, was sharpened by the work of persons such as Charles Booth, Henry Mayhew, Joseph Rowntree and Sidney and Beatrice Webb. Indeed the tradition – which carried until recently the term 'social arithmetic' - helped to create and shape the London School of Economics and Political Science. The political force of such social arithmetic affected social legislation, including the creation of the 'welfare state' in Britain in 1944.

However, what was sharply disturbing was the suggestion that *comparative* work depended on knowing numerical 'outcomes'. Was this true, I wondered, of comparative literature? Of comparative studies of religions? Statistics would clearly be useful and illuminating in the study of comparative history, but presumably they would not define the subject itself as a field of study (would they?).

Yet, somehow, the general epistemic position of Anderson was not only sharply disturbing; it was also familiar. It reverberated with the similar confidence which Harold Noah and Max Eckstein had, in the potential of the harder-edged social sciences (notably economics) to serve as a model for good comparative education. Anderson had merely taken the argument in a slightly different direction. Comparative education could be disaggregated into constituent harder-edged disciplines (economics, sociology and so on); comparative education (or what is now called ‘comparative and international education’) could make major and practical contributions not merely to studies of ‘development’ but to the creation of development itself; and all such forms of comparative education would progress more rapidly to the extent that the cognitive achievements of children were accurately measured cross-nationally.

Oddly enough, and paradoxically, this brings us very much up to date. PISA was - and is - technically magnificent even if notes of caution have been sounded by specialists such as Harvey Goldstein. PISA is ‘big social science’, fieldwork on a grand scale, courageously conceived, carefully constructed, and implemented and delivered with major professional skill. PISA is manifestly comparative, in the sense of offering rigorous, empirically-justifiable, statements about the ‘outcomes’ of educational systems in many nations. And of course PISA is not merely robust research but also ‘relevant’ research: it addresses crucial issues about what may (loosely) be called skill levels and permits us to reassess – against a great deal of evidence – the educational outcomes achieved by sub-groups in the population. We have a hard-etched picture of one dimension of the effectiveness and efficiency of many educational systems.

Clearly PISA is a major step forward, within a fifty-year trajectory of this kind of research. The PISA studies have contributed to the demystification of the social world and to one form of our knowledge about educational systems. Claims to the contrary need to be inspected for traces of obscurantism and irrationality – including the wilder claims about the flaws (and virtues) of PISA made by politicians and the public media in a number of countries. Let us assume therefore that it is sociologically and historically improbable that Emperors leave their palaces without putting on clothes; or at the very least being clothed by others.

This does not, however, mean that it would be impertinent for a small boy or a tall girl to ask what kinds of clothes the Empress was wearing – and this is why oddly enough, and paradoxically, we are still very much up-to-date if we rehearse what we used to know, just to be sure we know it still.

The Nominalist Trap. Comparative education as a field of study and PISA are both ‘comparative education’ because they look at education in different places – places separated by an international boundary. Perhaps not.

PISA is large scale field-work (the application of the same field-work questions in a number of places) with an emphasis on the conventional variables utilised in routine domestic tests of attainment research. PISA is a brilliant big-social-science mapping of outcomes – but in no anthropological, historical, sociological or cultural sense is it ‘comparative’ work. However, it is a contribution to comparative education because it sets new anthropological, historical, sociological and cultural puzzles about what the results mean. In itself it advances the complex forms of

university-based understanding of comparative education no more than the confidence of OECD that Lifelong Learning is a good thing. PISA is comparative education only in its nominalist traditional form (it ‘compares’ A with B with C and so on). PISA *looks like* comparative education because ‘everyone knows’ that comparative education ‘compares’ educational systems. The assumption that PISA is ‘comparative education’ because of its form is unfortunate and is based on a view of academic comparative education that is in the order of sixty years out of date, at least.

In other words, the question of the epistemic meaning of the term ‘comparative education’ requires constant rethinking from the moment it ceases to be *Auslandspädagogic* – partly so that the field may be re-theorised regularly but also so that the implicit political assumptions embedded in its knowledge are kept visible. Classifying and comparing was a political act of enlightenment, part of forms of rational emancipation in the eighteenth century; and classifying-and-comparing are epistemic and political acts in the twenty-first century also.

The Relevance Trap. Comparative education as a field of study and PISA are both ‘comparative education’ because they make a useful contribution to the ability of policy makers to form policy and they increase the likelihood of the successful transfer of good practice from one place to another (where the two places are separated by an international boundary). Perhaps not.

Part of the PISA debate – or the debate about TIMSS for example – has been simple puzzlement. Why has Finland done well? The exact answer if we ever get to know it will be of great interest; not least to the Finns. Similarly, propositions about ‘Confucian values’ and their relation to successful international test results of countries in East Asia are rapidly contradicted when some countries which do well in such tests are in Central Europe. In other words, PISA is only as relevant and as much a guide to action as politicians and governments wish to make it. PISA, like international benchmarks, and concepts of ‘world-class universities’, is a very specific form of knowledge: it is ‘ranking’ knowledge. Thus, like Olympic medals or world rankings in football, it can be used domestically as disciplinary knowledge for the governance of (educational or sporting) systems; it can be used domestically as ‘legitimation’ knowledge for justifying reform of whatever kind; and it can be used as a form of cultural triumphalism. So too can comparative education as a field of study – but then it betrays itself morally and politically.

In other words, the question of re-thinking the social use of comparative education regenerates itself in each generation. This is necessary because the international and domestic politics (within which comparative education and international testing is done) change for each generation.

The Science Trap. Comparative education as a field of study has slowly become a science and PISA is part of the new power of the empirical and comparative social sciences to predict and to shape and to improve societies. Perhaps not.

Certainly there is a long strand in the history of comparative education which has been dedicated to the creation of ‘scientific approaches’. Certainly the ‘facts’ of PISA have about as much technical objectivity as the social sciences can currently create in cross-national work. However, this still leaves us with some

rather anxious questions to discuss before we act on the basis of either ‘scientific comparative education’ or ‘the results’ of PISA.

Partly the question is in what cultural and historical context is the ‘science’ embedded and how does that affect the choice of research topics, the forms of research, and what counts as ‘a result’, whether this be in the United States of the 1920s, Nazi Germany in the 1930s, the Soviet Union in the 1950s, Britain in the post-Thatcher period, or PISA tests. Similarly, the political arithmetic and historical context and value stances involved in counting the poor in London is not the same ‘political arithmetic’ as devising measures of effective and efficient schools and publishing the results and then acting on those results. The choices of what is a ‘good school’, what counts as ‘improvement’ and which actions should be used to ‘improve’ schools are themselves politically embedded. The most extreme political vignette I have seen in education was not a lesson taught, in 1974, by a peasant in a field in China within the ‘deductive rationality’ of “learn from the workers, soldiers, and peasants” but a full hour lecture which I once heard the 1990s by an senior administrator on raising the ‘measured achievements’ in education of children in a particular State in the USA by a fraction of a decimal point – and how this was successfully done and at what cash cost.

Partly the question is who acts: the benign scientist and architect and town-planner improving the conditions of children by ‘removing’ slums or the national-level politician who asks for scientific results, declares the research to be robust and relevant and moves forward into action? There is considerable tension, in terms of the social sciences, about what kinds of research will be accepted as ‘science’ by politicians and which kind of research will be rejected. Overall, there is a considerable general tension between the application of science and theories of democratic governance. These tensions are played out strenuously and quite correctly in public debate and protest movement and governmental adjustments on the massive issues in our time (such as environmental pollution, global warming, energy and food supplies, the distribution of health care, and the structure of financial systems).

The danger, of course, is *hubris*. We, on the basis of our ‘research’, know. We will advise and politicians would do well to follow our advice. I am less and less certain of the wisdom of such a professional position (which is also a political position). I am less and less certain of the benefits – to take one example – of a Europe socially constructed and continually reformed on the basis of social-scientific research results which politicians take up a relevant and important.

In other words, academic comparative educationists - having historically aspired to be close to power and useful to policy makers and having historically aspired to construct a ‘useful science’ - should probably now pause and ask whether, contemporaneously, this position is complex enough. Perhaps we need to re-think, deliberately and explicitly, the politics and ethics of a strand of comparative education which could be called ‘applied comparative education’.

For the moment, we can certainly welcome PISA and its ilk as a fascinating form of inter-national educational research framed by the relationships of international political, economic and cultural and intellectual power and we can continue the analysis undertaken within this book: seeing PISA as a new layer within the

R. COWEN

international educational relations which have always been the concern of (and which have helped to frame) university-based academic comparative education.

Robert Cowen

Institute of Education, University of London

ANNEXES

ANNEX I

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

ANTONIO LUZÓN AND MÓNICA TORRES

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

INTRODUCTION

The present potential of data bases and search engines opens up a wide range of possibilities for research. Indeed, we are now looking not only towards the development of new fields of research, such as bibliometrics, compumetrics or webmetrics, but also the study of new research objects deriving from the use of these scientific instruments. Specifically noteworthy because of their influence in scientific literature are the so-called *Web of Science (WoS)* by Thomson-Scientific, better known by its former name of *Institute for Scientific Information (ISI)*, which is a package of several databases: *Science Citation Index (SCI)*, *Social Sciences Citation Index (SSCI)* and *Arts & Humanities Citation Index (AHCI)* by Thomson Reuters, as well as Elsevier's *Scopus*. Likewise, we must also mention fourth generation search engines such as Google's *Google Scholar* that are coming increasingly into competition with the foregoing. Together with the possibilities their application offers for searching the complete text of a paper, the bibliographical production of an author, journal or theme, or even reviewing the citations of a study, among others, these instruments have developed to validate and legitimize the academic productivity of institutions or researchers with a view to greater scientific objectivity.

This study consists in an analysis and examination of the scientific production in both scientific literature and public usage of the Programme for International Student Assessment (PISA) developed by the OECD through *WoS*, *Scopus* and *Google Scholar*, which is the subject of significant media repercussions.

Numerous documentation and bibliometrics studies have been carried out on the virtualities and limitations in use of the *WoS* and *Scopus* databases and the *Google Scholar* search engine (Noruzi, 2005; Pauly & Stergiou, 2005; Bosman, van Mourik, Rasch, Sieverts & Verhoeff, 2006; Harzing, 2007; Jacsó, 2006; or more recently Torres-Salinas, Ruiz Pérez, & Delgado, 2009). Although the two systems both search scientific publications to provide a citation index indicating the impact of diverse publications, they are also different in significant ways that are often complementary, the one making up for the failings of the other. In this sense, *Google Scholar (GS)* has the advantage of being free, thereby providing universal access to scientific information. It also provides information on all the relevant written production of researchers, such as doctoral theses and reports, among others, and not only on their publications in journals included in the *WoS* or *Scopus*. It includes publications not only in English, which predominates in the databases, but also in other languages of significant presence in the academic world. However, we must also consider that

Google Scholar (GS) does not give information on its coverage, the databases it includes, or modes of identification of publications and citations. It can also include texts not considered academic publications and is not updated with the same frequency as *WoS*.

In order to visualize more clearly the presence of the studies and documents published about PISA both in the scientific corpus and in public usage, we compare below the potential of databases regarding a topic and the degree of penetrability of open resources such as search engines regarding a question of such interest as PISA.

METHODOLOGY

As mentioned above, our study consists in the analysis and examination of the presence of a relevant subject of broad thematic spectrum in the most influential scientific databases recognised by the scientific community, such as *Web of Science* and *Scopus*, restricted to a minority sector of society. In order to compensate this restrictive factor, the presence of the PISA topic is introduced into the *Google Scholar* (GS) search engine, which has a lower scientific and visibility index, but higher metrical breadth, because its penetration index on the Web embraces documents, books, theses, specific studies, blogs and unpublished thematic elements all open to the interest of the public at large.

To this end, we began by making a descriptive analysis of the bibliographical references contained in the *Web of Science* and *Scopus* databases related to quantitative aspects (thematic areas, growth, volume), which we complemented with an evaluative analysis based on citations as differential criterion linked to the knowledge areas covered by the Social Sciences. In other words, we first delimited the preferential knowledge areas dealing with the PISA question, in order to show that this was not only a strictly educational field, but one with wider scientific consequences.

In a second phase after citation analysis, we determined what type of journals on a particular thematic field were related to PISA either through their title, descriptors or summary, as well as their impact factor as regards topicality and relevance.

Together with the theme, another important variable is productivity, defined as the number of studies published about PISA by an author or group of authors in a certain time unit. Thematic productivity can also be analyzed by taking into account the frequency of related publications in a specific journal or publication.

This quantitative analysis allows for a more far-reaching qualitative approach, consisting of the more precise description of the subject matter of journals, as well as their type, field and scope and the presence of the authors and institutions they represent. This initial analysis allows an approach to the most commonly used databases in the scientific field such as *Web of Science* and *Scopus*, as well as to the swifter, more intense data flow available in *Google Scholar* (GS), given its multidimensionality, particularly with a view to offering an overall image of the phenomenon begun by the PISA programme at all levels.

In order to visualize and represent the relations between thematic areas in *Web of Science* and *Scopus* and the scientific literature related to these fields, we made

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

use of the features of the *Pajek* network analysis programme (De Nooy, Batagelj and Mrvar, 2007; Batagelj and Mrvar, 2010) in association with bibliographical databases (Leydesdorff, De Moya and Guerrero, 2010; Dalud-Vincent and Normand, 2011).

ANALYSIS OF RESULTS

We began with an initial search for the PISA topic by limitation to the thematic field of the Social Sciences, in *Social Sciences Citation Index (SSCI)* and *Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH)*, belonging to *ISI-WoS* and *Social Sciences & Humanities*¹ of *Scopus*.

Table 1. Titles associated with thematic fields (2000-2010)

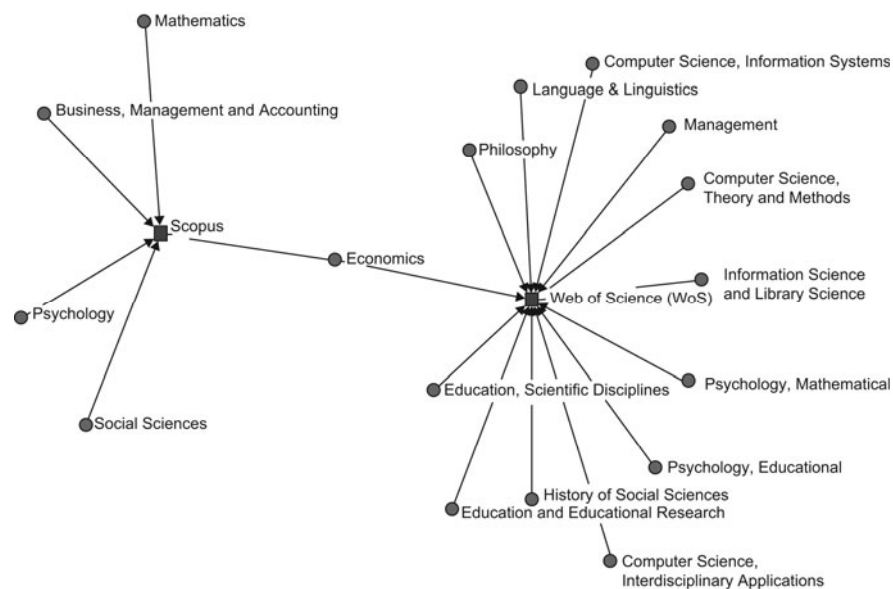
<i>Thematic area</i>	<i>Web of Science (WoS)</i>	<i>Scopus</i>
Education & Educational Research	139	
Social Sciences		153
Economics	28	28
Psychology, Educational	6	
Psychology		45
Education, Scientific Disciplines	3	
Computer Science, Information Systems	2	
Computer Science, Interdisciplinary Applications	2	
History of Social Sciences	1	
Psychology, Mathematical	2	
Mathematics		15
Computer Science, Theory & Methods	1	
Information Science & Library Science	1	
Language & Linguistics	1	
Management	1	
Business, Management and Accounting		15
Philosophy	1	
Total studies published	188	256

Source: By the authors based on data from *WoS* and *Scopus*

The first detail we noticed when searching such a broad topic as PISA is that in *Scopus* the thematic field of the *Social Sciences* forms part of the *Humanities* field, so that the citation index is larger and the search must be gradually refined until it centres on the more specific area of the Social Sciences.

The initial search found that the topic used was “PISA” (using the equation: [PISA*] and [PISA and OECD]) and the majority field in which studies on the subject have been published was Education and Educational Research in the *Web of Science*, not forgetting the broad range corresponding to the Social Sciences as represented by the *Scopus* database. This is evidence, although it is symptomatic that other fields such as Economy also have a significant presence. We must now determine which journals are those that published most studies on PISA, which are the most cited, and which authors’ studies were the most appreciated. At the same time, we use GS to visualize and compare the most cited studies and authors.

Taking the topics described above, *Google Scholar* gave a total of 14,400 search results, although only the first thousand were relevant (Torres-Salinas, Ruiz-Pérez and Delgado-López-Cózar, 2009; p. 503).²



Source: By the authors based on *Pajek*

Figure 1. Thematic fields in WoS and Scopus.

It is also significant that most of the studies located in *WoS* and *Scopus* are article in specialist journals (81%), and not others such as papers in symposia, editorials or book reviews. On the other hand, *Google Scholar* visualizes other types of studies aside from established scientific channels and irrespective of the amount of results. As already stated, the majority of relevant studies on PISA are based on the number of citations.

PISA and OECD - Google Académico

[\[PDF\] PISA 2000](#)

PK Deutsches - Basiskompetenzen von Schülerinnen und Schülern im ..., 2001 - arcor.de

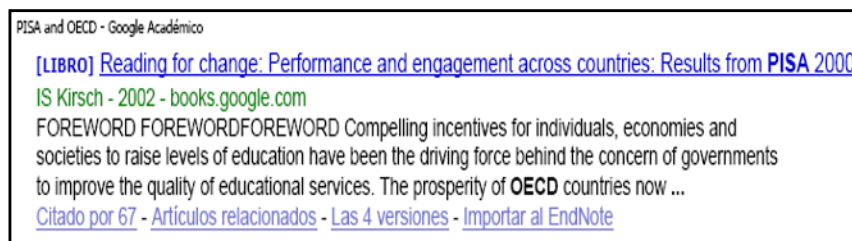
... Was ist OECD/PISA? - Die wichtigsten Merkmale im Überblick ... Abbildung 1.1: PISA-Teilnehmerstaaten An PISA 2000 teilnehmende OECD-Mitgliedsstaaten Australien Belgien Dänemark Deutschland Finnland Frankreich Griechenland Irland Island Italien ...

[Citado por 654](#) - [Artículos relacionados](#) - [Versión en HTML](#) - [Las 3 versiones](#) - [importar al CndNote](#)

Therefore, we find studies such as *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*,³ published in 2001 in Germany by the *Deutsches PISA-Konsortium*, which records the challenging figure of 654 citations

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

In addition, *Google Scholar* has the advantage for the researcher of presenting related documents or Web pages directly linked to the search options. Or documents such as *What accounts for international differences in student performance? A re-examination using PISA*, by Thomas Fuchs and Ludger Wößmann with 188 citations. Paradoxically, this text belongs to a periodical publication from the field of economy called *Empirical Economics*, 32 (2007) with a much lower citation potential in *WoS* and *Scopus*. For this reason we value the power of this search tool, which has less of an academic focus and less precision, but with a broad potential spread. Another interesting feature of GS is the possibility of consulting books and their tables of contents, which is an innovative characteristic unavailable in *WoS* or *Scopus*. It provides access to and information on a significant number of publications, even on-line, in other databases and libraries, as well as incorporating the citations received by these publications and their localisation, which is doubtless one of the most outstanding features of this search tool for the research and development of documentation and bibliometrics. (Harzing 2007 y 2007a; Mayr y Walter, 2007; Meho & Yang, 2007; Jacsó, 2006; Torres-Salinas, Ruiz-Pérez, y Delgado López-Cózar, 2009).



As a freely accessible tool, *Google Scholar* is able to familiarize the researcher with an infinite amount of documentation of different types and origins to be processed and of increasing scientific usefulness. In the specific case in hand, the resulting flow of documentation is considerable and is an indispensable tool not only for consultation of documents published in other journals, books, proceedings, symposia and doctoral theses, but also in web pages with relevant documentation.

Impact and visualization of PISA in scientific literature through WoS and Scopus

In the second phase of our analysis of studies concerning PISA published since 2000, we examine the presence of the question in citations received in journals with high impact index. We then analyze the most relevant authors according to their significance or the presence of their studies in other publications using the h-index, which is an indicator proposed by Jorge Hirsch (2005) for the number of articles of any one author with a certain number of citations⁴.

Table 2. Journals publishing scientific articles on PISA, according to WoS

Field: Source Title	Record Count	% of 165	Bar Chart
<i>Zeitschrift für Pädagogik</i>	28	16.9697 %	
<i>Zeitschrift für Erziehungswissenschaft</i>	19	11.5152 %	
<i>Oxford Review of Education</i>	6	3.6364 %	
<i>Journal of Education Policy</i>	5	3.0303 %	
<i>British Educational Research Journal</i>	4	2.4242 %	
<i>British Journal of Educational Studies</i>	4	2.4242 %	
<i>Journal of Curriculum Studies</i>	4	2.4242 %	
<i>Comparative Education</i>	3	1.8182 %	
<i>Comparative Education Review</i>	3	1.8182 %	
<i>European Journal of Education</i>	3	1.8182 %	
<i>International Journal of Science Education</i>	3	1.8182 %	
<i>Journal of Research in Science Teaching</i>	3	1.8182 %	
<i>Review of Research in Education</i>	3	1.8182 %	
<i>School Effectiveness and School Improvement</i>	3	1.8182 %	
<i>Science Education</i>	3	1.8182 %	
<i>Australian Educational Researcher</i>	2	1.2121 %	
<i>Australian Journal of Education</i>	2	1.2121 %	
<i>Educational Research and Evaluation</i>	2	1.2121 %	
<i>European Journal of Personality</i>	2	1.2121 %	
<i>International Journal of Educational Development</i>	2	1.2121 %	
<i>Zeitschrift für Psychologie-Journal of Psychology</i>	2	1.2121 %	

Source: ISI-Web of Science

This table shows the degree of production on the PISA question between the years 2000 and 2010 in journals found in the *WoS* database. They all belong to the majority thematic area *Education & Educational Research*, which recorded the highest number of studies. It is noteworthy that the two journals publishing the highest number of studies (28.4%) are German, followed by British publications (*Oxford Review of Education*, *Journal of Education Policy*, *Comparative Education* and *British Educational Research Journal*) and even North American journals (*Journal of Curriculum Studies* and *Comparative Education Review*) with higher impact index.

According to the indicator of concentration or dispersion of scientific literature, in the *WoS* database the German journals *Zeitschrift für Pädagogik* (position 54 out of 87 in the ranking of this thematic area in the *WoS*) and *Zeitschrift für Erziehungswissenschaft* (position 57) contain the highest number of studies published. This shows the social concern in German about this question, with the subsequent scientific projection concerning the results of the German education system in the OECD evaluations by PISA.

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

The *Scopus* database shows higher dispersion, because the central nucleus only contains three journals out of the 106 with studies published on PISA. There is some parallelism between the databases as regards journals with the highest number of studies published: *Oxford Review of Education*, *Educational Research and Evaluation*, *Zeitschrift für Pädagogik*.

Table 3. Journals publishing scientific articles on PISA according to Scopus

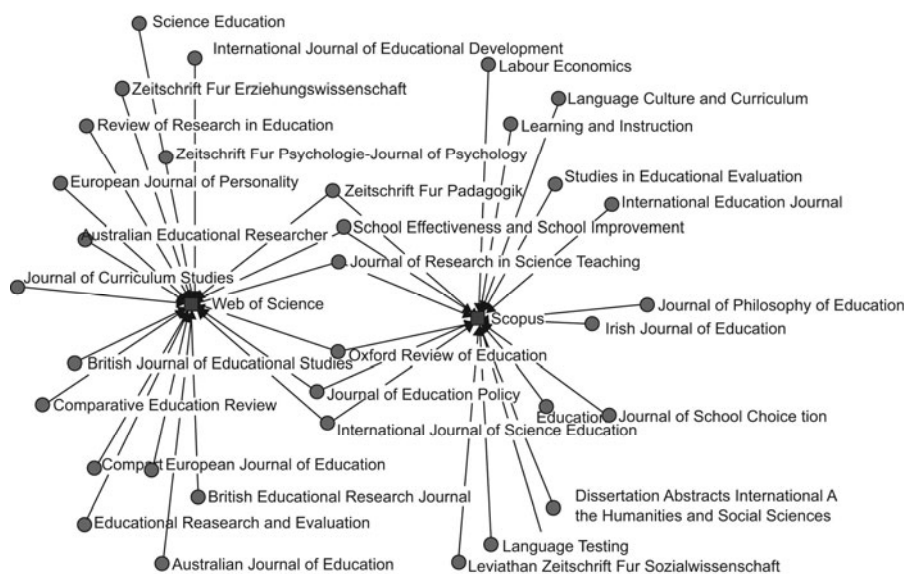
Field: Source Title	Record Count
<i>Oxford Review of Education</i>	7
<i>Educational Research and Evaluation</i>	4
<i>Zeitschrift für Pädagogik</i>	4
<i>International Journal of Science and Mathematics Education</i>	3
<i>School Effectiveness and School Improvement</i>	3
<i>Futuribles</i>	3
<i>International Education Journal</i>	2
<i>Learning and Instruction</i>	2
<i>Journal of Research in Science Teaching</i>	2
<i>Language Testing</i>	2
<i>Irish Journal of Education</i>	2
<i>Dissertation Abstracts International A the Humanities and Social Sciences</i>	2
<i>Studies in Educational Evaluation</i>	2
<i>Journal of Education Policy</i>	2
<i>Language Culture and Curriculum</i>	1
<i>Journal of School Choice</i>	1
<i>Labour Economics</i>	1
<i>Leviathan- Zeitschrift für Sozialwissenschaft</i>	1
<i>Journal of Philosophy of Education</i>	1

Source: By the authors from data in *Scopus*

Scopus also shows the frequency of publications over the ten-year period that the PISA programme has been in operation. Most of the studies are concentrated in the 2007-2009 segment, with 55% of all publications in just three years following publication of the results of the last PISA report in 2006. Two of the most cited documents are “PISA 2006: An assessment of scientific literacy,” published in the *Journal of Research in Science Teaching* (8) in 2009 and “National IQs predict educational attainment in math, reading and science across 56 nations” by Richard Lynn and Jaan Mikk in *Intelligence* (3), with a significant number of references. There are less studies critical of the PISA programme, such as “A critical examination of PISA’s assessment on scientific literacy”, published in the *International Journal of Science and Mathematics Education* 7(6); or “Governing by

numbers: The PISA 'effect' in Europe”, in *Journal of Education Policy* 24 (1), which have been well received, despite their recent publication date.

The four publications grouping the highest number of citations in *WoS* are of a different nature. An example is “Conceptual change: a powerful framework for improving science teaching and learning”, published in 2003 in the *International Journal of Science Education* 25(6), which has 52 citations, with a mean annual coefficient of 6.5. Similarly, “The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations” by the psychologist Heiner Rindermann of Magdeburg University in Germany was published in 2007 in the *European Journal of Personality* 21, and has 32 citations in all, with a mean annual coefficient of around 8. Likewise, “The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education”, by Hanu Simola, an education sociologist at the University of Helsinki, has a total of 17 citations since 2005, when it was published in the British journal *Comparative Education* 41(4).



Source: By the authors based on *Pajek*

Figure 2. Journals associated with their thematic fields in *WoS* and *Scopus*.

We might continue by including the social divisions of gender and immigration as regards factors of social inequality, which also have an important influence on the references to published studies and cognitive factors of the tests directly linked to the PISA programme, or relations with evaluation and other international tests such as TIMSS and PIRLS carried out by the IEA. We can deduce that although the thematic areas published refer to the Social Sciences, specifically the field of education and research, they can contain great diversity and thematic range, where disciplinary boundaries become unclear, but that they clearly show the evolution of

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

the mobilisation of knowledge brought about by the public debate on the PISA programme and its scientific impact on an international scale. Nonetheless, if we observe the publications throughout the ten years in which the PISA programme has been in operation, its impact is uneven over time, with periods of greater intensity of publication coinciding with the three-yearly OECD reports.

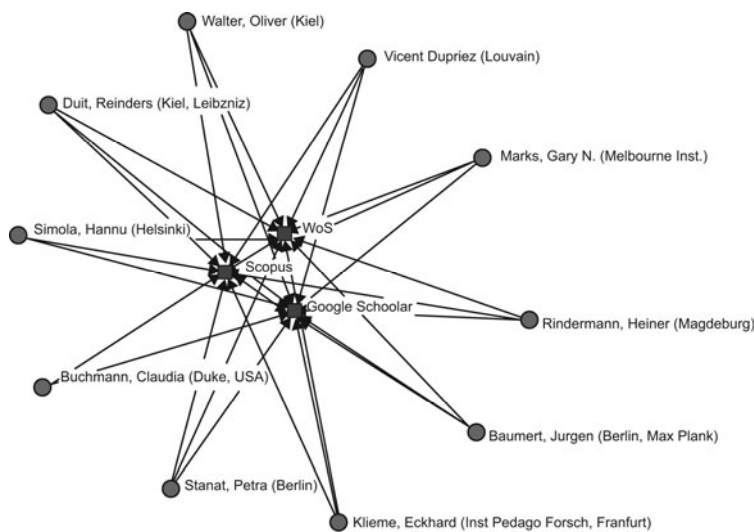
Concerning the authors, we focus on those whose range of production appears in all the databases. *WoS*, *Scopus*, and *Google Scholar*. The ten authors with scientific studies on PISA share a clear profile because of their work not only on the topic, but also on questions in the thematic area, as well as their productivity, with a considerable impact index for the decade 2000-2010. They are all related to the area of education, with a predominance of German authors. This does not mean there are no other authors with important significance indices whether because of the quality of their work or the journal where they have published.

Table 4. Profile of most cited authors of studies on PISA

Author	University	<i>WoS</i>		<i>Scopus</i>		<i>Google Scholar</i>	
		h-index	Med. cit.	h-index	Ref.	h-index	Ref.
Baumert, Jürgen	Berlin, Max Planck Inst.	13	7,72		47	>10	477
Buchmann, Claudia	Duke (USA)	8	14,07		16	>10	422
Rindermann, Heiner	Magdeburg,	8	6,23		22	>10	284
Marks, Gary N.	Melbourne Inst.	6	3,37		20	>10	1130
Duit, Reinders	Kiel, Leibniz Inst Sci Educ	5	8,17		17	>10	1492,7
Klieme, Eckhard	Inst. Int. Pädag. Forsch. (Frankfurt)	4	2,04		25	>10	387,7
Simola, Hannu	Helsinki	3	7,25		7	9	229,3
Stanat, Petra	Berlin	2	1,58		11	>10	377,8
Dupriez, Vincent	Louvain	2	1,33		9	8	193,8
Walter, Oliver	Kiel	2	1,17		16	8	193,1

Source: By the authors

In *WoS* and *Google Scholar* we have classified them according to the h index, without losing sight of the coefficient of citations of their work, while in *Scopus* they are classified according to the references appearing in their articles. Despite the differences, there is a relation between the significance and importance of the publications cited by other authors as reference.



Source: By the authors based on *Pajek*

Figure 3. Main authors associated with WoS, Scopus and Google Scholar.

We can also observe that this intensity is associated to internationally renowned research centres, such as the German Max Planck Institute or universities with an important research record. However, the possibilities offered by GS are enormous and extremely useful. It provides not only information on the best-known articles and those cited in other studies, but also valuable documentary information on the authors in relation to the thematic area of research. We can see, therefore, that the volume of documentation on the authors writing about PISA provided by *Google Scholar* is considerably higher than that of other databases, which is a factor to be taken into account because of its innumerable possibilities, despite the filtering and selection of documents it provides.

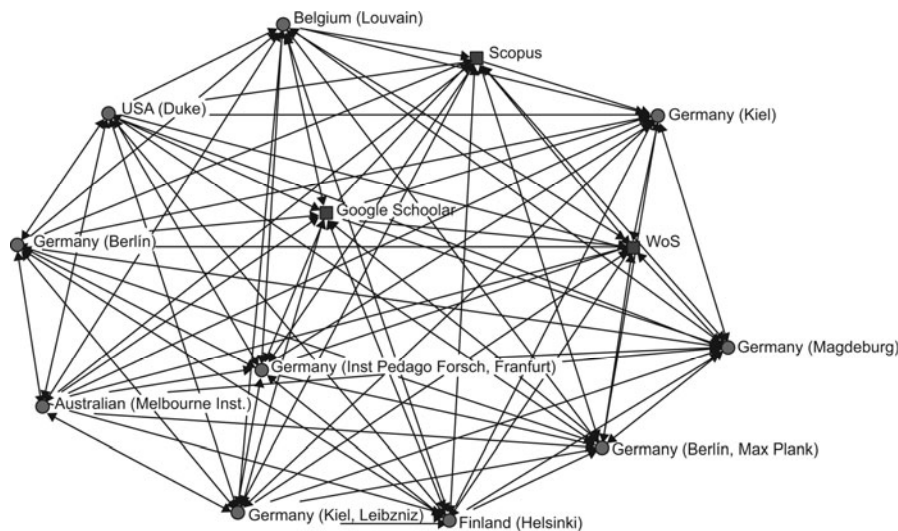
Crossing boundaries

There is no language barrier for the visualization of PISA in GS. Its preferences allow searches to be made in different languages, particularly those with more relevance for the academic world on a global basis. Although English is the predominant language, and not only in the scientific sphere, as we have seen in *Google Scholar* the PISA phenomenon has a relevant presence in other languages such as German, Italian, French or Spanish, where its visualization in other formats also takes on importance. In Italian, *Google Scholar* detected 1140 searches related to the “PISA and OECD” algorithm, of which many corresponded to books such as *Come leggono i quindicenni. Riflessioni sulla ricerca OCSE-PISA*, de Emma Nardi (2002), and the contribution of the former minister Luigi Berlinguer, now in its fourth edition with 12 citations; *Economia e società nella cultura dei giovani. Rappresentazioni e credenze degli studenti medi*, by Marcello Dei (2006) with 8

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

citations; *Pisa 2003: bravi come gli altri. Nuova luce sulle competenze dei quindicenni dal confronto fra regioni italiane ed europee*, by Lucciano Abburrà (2006), with 5 citations. As well as specific reports such as *Ocse Pisa 2003*, *La ricerca Ocse-Pisa* (2001) and articles *La comparazione internazionale OCSE Pisa* (2002), all with a medium citation index and easily accessible in *GS*.

In French, the visualization of PISA is somewhat wider. *GS* shows 1440 searches, of which the first results on PISA are noteworthy. It should also be taken into account that the French-speaking world extends well beyond France to a much wider linguistic and cultural area. Thus, the *PISA 2000 Technical Report* (2002) by Ray Adams and Margaret Wu and published by the OECD was cited 180 times, despite only being available in English. This contrasts with the OECD's publication of the first PISA report in French (2001) entitled *Connaissances et compétences: des atouts pour la vie. Premiers résultats de PISA 2000* with only three citations in *GS*. However, the recent *L'Élitisme républicain. L'école française à l'épreuve des comparaisons internationales* (2009) by the sociologists Christian Baudelot and Roger Establet, in which the mediocre results of French students are attributed to "republican élitism" or the formation of élites, has seven citations in *GS*. Other articles and reports of less significant presence on the web are "La compétence en lecture des jeunes de 15 ans: une comparaison internationale" (2002), "Notre enseignement est-il de bonne qualité? L'enquête du programme PISA" (2002), "Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français" (2006), or "Pourquoi les performances des élèves flamands et francophones sont-elles si différentes? Une analyse par la méthode des frontières stochastique" (2009), all easily available on the web and *Google Scholar*.



Source: By the authors based on Pajek

Figure 4. Countries most represented in the scientific literature on PISA in *WoS*, *Scopus* and *GS*.

The PISA reports have had significant social repercussions in Germany, with 8230 records located by GS on the web.⁵ Although some appear in English, there is also a considerable number of publications in German. An example is *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (2001) by Jürgen Baumert et al., which gives a fair idea of the impact of the PISA report in Germany, as already mentioned in the 654 references detected. The equally impressive number of 504 citations is recorded by just the single chapter entitled “Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich” by Jürgen Baumert and Gundel Shümer from the book *Pisa 2000-Die Länder der Bundesrepublik Deutschland im Vergleich*, compiled by Baumert et al., (2002). Numerous publications followed the PISA reports of 2003 and 2006, with a high percentage of citations, although rather less than after the first report in 2000. Examples are *PISA 2003: der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*, published by the German PISA-Konsortium in 2004, with 176 citations; *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*, also published by the German PISA-Konsortium with 37 citations, where the individual chapters likewise have a considerable number of citations, and *Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006* with 38 citations. The immigration question is also outstanding. The 19 citations of reports such as “Arbeitsmarkteinstieg nach dualer Berufsausbildung–Migranten und Deutsche im Vergleich” by Andreas Damelang and Anette Haas (2006) give some idea of the relevance of this question. As in other languages, GS gives access to an important documentary source of particular importance for research directly related with international evaluations and the PISA programme in particular.

In Spanish, the PISA report also has considerable presence, particularly because of the effect and influence in Latin America. Other than Spain, countries such as Argentina, Mexico and Chile concentrate considerable interest with 2830 references recorded in GS. Most of the references and citations are to be found in the monographic number on PISA of the Spanish Ministry of Education’s *Revista de Educación*, including texts by Álvaro Marchesi (2006) on the relation of PISA to Spanish educational policy, Andreas Schleicher (2006) on the principles underlying the PISA programme, and even a general overview of PISA by Turner. There are also other studies published in journals available on the web with a significant number of references, in particular “La competencia matemática en PISA”, in *PNA* (research journal of the *Didáctica de la Matemática: Pensamiento Numérico* group in the Andalusian Research Plan of the Junta de Andalucía), with 22 references; “Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003”, by Jorge Calero and Oriol Escardíbul (2007) in *Hacienda Pública Española-Revista de Economía Pública*, with 28 references; and “TIMSS y PISA. Dos proyectos internacionales de evaluación del aprendizaje escolar en ciencias” by the Schools Inspector José Antonio Azevedo (2005), published in *Eureka*, a journal on extending teaching and didactics of the sciences, with 22 references.

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

There is a lower presence of articles on PISA in Argentina, Mexico or Chile, although there are some with a moderate number of references, such as “Argentina en el Estudio PISA 2000”, “México en los resultados PISA 2003. Una interpretación no catastrofista”, “La evaluación y el diseño de políticas educativas en México”, or “Los procesos de selección en los países participantes en PISA 2003” by Brunner (2007).

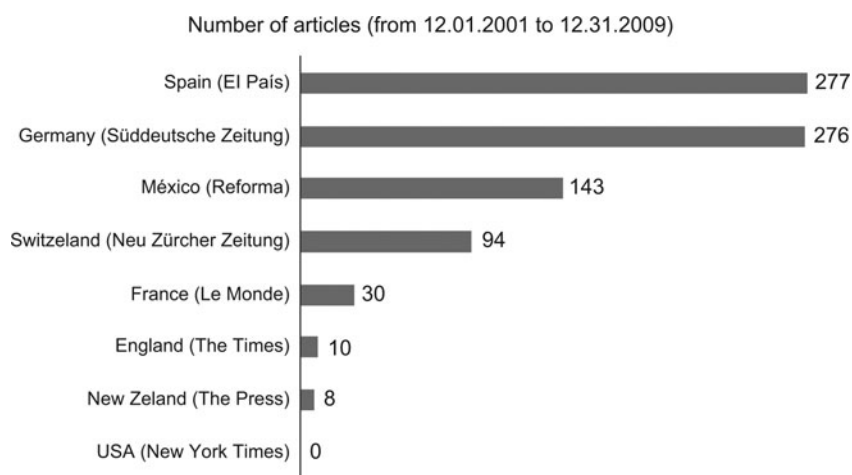
THE PISA SHOCK. THE VIEW OF THE DAILY PRESS

The daily press is not reflected in *Google Scholar* (GS), but there is no doubt that the publication of the data in the *PISA Report 2009* had worldwide media impact, as we shall show below. Indeed, after the publication of the latest report announced on the OECD’s official website for the 7th December, all the newspapers around the world carried information on the results and their repercussions in each country. The media intensity of the report and the considerable flow of news only lasted one week, with a gradual decrease of directly related news items in favour of more detailed analysis of the data by experts. This consideration of the media treatment of the *PISA Report 2009* represents a complementary tool, especially for understanding of the impact that can be caused by a broad-ranging, detailed report by the OECD on not only the reading, mathematics and scientific competence of fifteen-year-olds in the developed world, but also a whole archetype of interwoven discourses with the aim of influencing national and European educational policies (Martens, Nagel and Windzio, 2010). Moreover, it can indirectly contribute to understanding how social reality is generated and developed on the basis of journalistic treatment, as prodigiously brought out by a multitude of interrelated news items, directly linked to the educational policies put into effect over the last few years.

Field of study

The great majority of the news items consulted appeared in the week spanning 5th to 12th December 2010, when the OECD report was published. This was the interval when the highest intensity of news flow concerning PISA took place in both the Spanish national press and the international press. We consulted 1559 news items concerning PISA in the Spanish national press using the search algorithms. “PISA 2009” or “PISA and OECD” in 21 newspapers, five of which have national distribution (*El País*, *El Mundo*, *ABC*, *La Razón* and *Público*), the other sixteen being distributed in the different regions or autonomous communities of Spain (See [Table 4 Annex](#)). In the search options of the different newspapers, we consulted 1621 items during the week in question, finding 146 items directly related to PISA and its repercussions on educational policy in such a short space of time, of which we reproduce some of the most relevant headlines. In the majority of the papers these news items are open to readers’ comments, of which we found 1758 on the PISA report, with 69% of them appearing in the two days after the

report's publication. Some papers, moreover, indicate the number of visits to particular new items, with recommendations to other readers. In the case of items on PISA, some achieved up to 4962 visits in a single day and were recommended by thousands of readers in the most widely available Spanish newspapers (*El País*, *El Mundo*, *ABC*, *La Vanguardia*, *El Periódico de Catalunya*, *La Razón y Público*). The impact of PISA in the daily press in Spain has been having some relevance as we have shown and also by other researchers preoccupied about the range of different resonances of perception of the PISA experience by the public of different countries (see the database Factiva in Daniel de Olano et al., 2010, pp. 16-18 and Figure 5).



Sources: P. Knodel *et al.*, 2010 p. 17

Figure 5. Articles about PISA in the daily press in some countries.

However, in our case the field of study in the international press was less comprehensive, with account being made of the main newspapers in those countries where PISA has most impact or international projection. These publications include *Le Monde* and *Le Figaro* in France, *The Guardian*, *The Times* and *The Financial Times* in the UK, *Berliner Zeitung*, *Die Welt*, *Frankfurter Allgemeine Zeitung* in Germany, *The Washington Post* and *New York Times* in the USA and *La Repubblica* and *Corriere della Sera* in Italy. We also included a reference to a Swedish English-language newspaper *Skolverket* in order to estimate the impact of PISA in a Scanidnavian country other than Finland. The amount of news items appearing in these newspapers on the PISA report was low, with the exception of Germany, where the impact was higher, taking into account the spread of news in the interval described above. A total of 56 news items directly related to PISA were published in the international press (See table in annex, in the context of 232 items appearing using the search options).

It should also be mentioned that in both the Spanish and the international press news items regarding the PISA programme have continued to appear gradually, centring mainly on the analysis of the data and their repercussions for the educational system and the learning process. Expert analyses, normally by university professors, have also been published, as well as statements by politicians justifying or legitimizing the published data.

We should underline that the visualization of PISA 2009 impact in the written press is merely descriptive and we do not engage here with detailed analysis, which will be the subject of subsequent research. Our aim here is merely to illustrate the contrast between the scientific treatment of the PISA phenomenon and its impact in the daily press.

Main characteristics and viewpoints

We can make the following observations after an initial examination that does not go into detailed analysis of the news items published and their interpretation.

- a) The news items appeared irregularly. There are two key days that concentrate 72% of the items, which are the date of publication (7th December 2010) and the following day. News items concerning PISA have continued to appear spread over the established interval and particularly concerning analysis of the effects (*ABC*, December 7, 2010: “Five keys to understanding the Spanish education system”; *La Vanguardia*, December 7, 2010: “Spain must encourage excellence after its maximum efforts to leave the lower levels”; *El Periódico*, December 7, 2010: “Spain stuck in the middle levels of the PISA report”; *Público*, December 8, 2010: “PISA proves that investment is not enough for school success”; *Le Monde*, 7 December, 2010: “The French school poorly classified and accused of being unfair”; *Frankfurter Allgemeine Zeitung*, 7 December, 2010: “The PISA shock is healthy. Awareness instead of shock”; *Skolverket*, 7 December, 2010: “Reading literacy of 15-year-olds and equity in the school have decreased in Sweden”) and repercussion of the data in clear reference to national educational systems in comparison with those of countries in the upper echelon (*Diario El Mundo*, December 7, 2010: “Five Asian and two Oceanic countries in the international education top 10”; *Financial Times*, December 7, 2010: “Why are Chinese schoolkids so good”; *New York Times*, December 7, 2010: “Top Test Scores From Shanghai Stun Educators” and December 8, 2010: “Western Nations React to Poor Education Results”).

Table 5. Repercussion and incidence of PISA 2009 on the educational system

Newspaper	Headlines
<i>ABC</i>	<p>“La educación española continúa por debajo de la media en el informe PISA”.</p> <ul style="list-style-type: none"> • Cinco claves para entender el sistema educativo español. • Sin motivación, ni expectativas de futuro, perfil del alumno repetidor español. • La educación española continúa por debajo de la media en el informe PISA. • Estancados en la mediocridad.
<i>El Periódico</i>	<p>“España se estanca en los niveles medios en el informe PISA”.</p> <ul style="list-style-type: none"> • Mejora en comprensión lectora pero sigue 13 puntos por debajo de la media de la OCDE. • Un 20% de estudiantes están por debajo del nivel requerido.
<i>Publico</i>	<p>“PISA prueba que la inversión no basta para el éxito escolar”</p> <ul style="list-style-type: none"> • El contexto socioeconómico, el entorno familiar y la formación del profesorado son claves.
<i>Le Monde</i>	<ul style="list-style-type: none"> • L'école française mal classée et jugée injuste. <ul style="list-style-type: none"> ○ La France en chute sur les mathématiques. ○ L'école ne joue plus son rôle d'ascenseur social. • Dix leçons du classement PISA 2009. • La France, pays du grand écart scolaire <ul style="list-style-type: none"> ○ Tous concluent que notre système scolaire ne parvient pas à favoriser la réussite des élèves.
<i>Frankfurter Allgemeine Zeitung</i>	<ul style="list-style-type: none"> • Deutschland ist aufgestiegen <ul style="list-style-type: none"> ○ Mit Erleichterung reagieren Bildungsforscher, Lehrer und Politiker auf die verbesserten Ergebnisse deutscher Schüler beim Pisa-Test. • Heilsamer PISA-Schock. Sensibilität statt Schockstarre. <ul style="list-style-type: none"> ○ Zehn Jahre nach dem ersten Pisa-Test im Jahre 2000 liegt Deutschland beim Lesen im OECD-Schnitt, in Mathematik und Naturwissenschaften deutlich darüber und zählt zu den wenigen Ländern, die sich durchgängig verbessert haben.
<i>Skolverket</i>	<ul style="list-style-type: none"> • Reading literacy of 15-year-olds and equity in the school have decreased. <ul style="list-style-type: none"> ○ Reading literacy and the knowledge and skills of Swedish 15-year-olds in mathematics have decreased over the last decade. In the natural sciences Swedish students today perform below the international average. Equity in the Swedish school has decreased, and the number of students not achieving the lowest reading literacy levels is growing.
<i>Financial Times</i>	<ul style="list-style-type: none"> • Why are Chinese schoolkids so good.
<i>The New York Times</i>	<ul style="list-style-type: none"> • Top Test Scores From Shanghai Stun Educators. • Western Nations React to Poor Education Results. <ul style="list-style-type: none"> ○ Can you imagine the reaction if we told the students of Chicago that the PISA was an important international test and that America's reputation depended on them performing well?" Mr. Schneider said. "That said, China is taking education very seriously. The work ethic is amazingly strong."

b) We can observe that the majority of news items about PISA were published by journalists from different sections in the newspapers.

These items are accompanied by expert analyses, mainly by university professors associated with the educational field (Tomás Recio in *El País*, December 3, 2010: “The PISA report as a weapon against the education system”; Julio Carabaña in *El País*, December 5, 2010: “We carry on where we were”; Eric Charbonier, in charge of PISA in France, *Le Monde*, December 7, 2010: “Good pupils save our education system”; Mariano Fernández Enguita in *El País*, 10 December, 2010: “Aurea Mediocritas”; Antonio Bolívar in *El País*, December 11, 2010: “Mediocre in excellence, outstanding in equity”; Nathalie Mons in *Libération*, 26 December, 2010: “PISA classification: when we like to put education to the test”; Jorge Calero in *Público*, December, 8, 2010: “The bottom-line of PISA” and January 8, 2011: “Isn’t it too much failure?”; Luis Rico and Lorenzo Blanco in *El País*, January 8, 2011: “What to do with the PISA data”, or Gimeno Sacristán in *El País* January 14, 2011: “The magical power of numbers and the PISA report”).

c) Almost all the news items (87%) were published in the *Society and Culture* section, which deals occasionally with educational issues when necessary. The days of highest impact were those closest to publication of the data, when editorial comment appeared evaluating the results and their influence on the country’s education policy. There followed interviews with political figures (ministers, those responsible for education in the various autonomous regions and other political decision-makers), in which they explained the measures to be taken after analysing the results. Then again, some articles merely evaluated the results on the basis of whether they represented an improvement on those of the 2006 report.

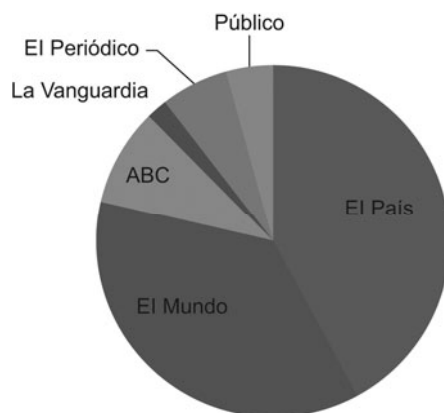


Figure 6. Comments on news items about PISA 2009.

The French Minister for Education Luc Chatel announced a programme designed to strengthen scientific knowledge and “guarantee that throughout their school career, student will be more inspired by science and have more

opportunities to participate in these sectors” (*Le Monde*, 9 December, 2010). In the *Corriere della Sera* (7 December, 2010), the Italian Minister for Education Mariastella Gelmini stated: “We can be proud of this result,” referring to the good results obtained by Italy in the tests. Arne Duncan, Federal Secretary of the Education Department of the United States, said in the *Washington Post* (7 December, 2010): “For me, it’s a massive wake-up call. Our goal should be absolutely to lead the world in education.” In the *New York Times* (7 December, 2010) Andreas Schleicher, director of the PISA programme, said in relation to the rise of Asian countries and their good results: “This is the first time that we have internationally comparable data on learning outcomes in China”. “While that’s important, for me the real significance of these results is that they refute the commonly held hypothesis that China just produces rote learning.” “Large fractions of these students demonstrate their ability to extrapolate from what they know and apply their knowledge very creatively in novel situations.” In contrast to the data obtained in Shanghai (China), the United States results did not leave former members of the North American educational administration indifferent. Chester E. Finn Jr., who worked in the Department of Education under President Reagan, referred in the *New York Times* (7 December, 2010) to the success of the Soviet Sputnik space programme in the late 1950s in relation to the results obtained by China in the PISA report: “Wow, I’m kind of stunned, I’m thinking Sputnik.” “I’ve seen how relentless the Chinese are at accomplishing goals, and if they can do this in Shanghai in 2009, they can do it in 10 cities in 2019, and in 50 cities by 2029.”

- d) In Spain, the different results obtained in the different regional communities with responsibility for education also revealed political differences as regards the presentation and interpretation of results. In *El País* (7 December, 2010) we find: “One Country, Two Schools.” “The communities of the north pass the exam easily; the south fails. Experts attribute this to socio-economic, cultural and management factors.” In *El Mundo* (7 December, 2010): “The two Spains persist. The communities in the northern half are far ahead of those in the south.” This differential context also underlines the divergences and contrasts in educational policies put into place or under consideration (See [Tables 3 and 4](#) in Annex). The ranking view is also operative internationally (See [Figure 1](#) in Annex published by *Die Welt* in Berlin, 7 December, 2010), with consideration of China and Korea and other European countries with better results.
- e) There is noticeable criticism of the pupils’ low levels and of the educational authorities responsibility for the omissions and inefficiency of the measures used. The report data are given in a general fashion, but with little subtlety of statistical analysis, merely providing overall interpretations of ranking. For example, *La Vanguardia* (7 December, 2010): “Repeating students and immigration, main causes of educational stagnation in Spain.” “Spain must encourage excellence after making the maximum effort to leave the lower levels.” and in *ABC* (10 December, 2010): “The PISA report shatters the work of the Junta de Andalucía over the last decades, as it does not focus on

education, but on propaganda”, and again (8 December, 2010): “a society fails.” The Andalusian supplement of *El País* (16 December, 2010) takes a similar line: “Are the Andalusians more stupid?” or in the *Heraldo de Aragón* (7 December, 2010): “Aragonese pupils among the best in the country in all subjects. PISA warns of the high risk of social exclusion for Spanish repeating students.” The international press also commented on the responsibilities of the educational authorities for the overall results obtained. *Le Figaro* (7 December, 2010): “French pedagogical methods devalued.” *New York Times* (7 December, 2010): “The results also appeared to reflect the culture of education there, including greater emphasis on teacher training and more time spent on studying rather than extracurricular activities like sports.”

- f) The surprise and susceptibility about the results from South East Asia (Shanghai and South Korea). Both the Spanish and the international press reflected the optimum results obtained by countries such as China and Korea and openly asked about the reasons for this success. In its edition corresponding to the publication of the results by the OECD, the *Financial Times* asked in one of its headlines “Why are Chinese schoolkids so good?”; similarly, *The Washington Post* specified “Finland is over. Now it’s all about Shanghai and the *New York Times*: “With China’s debut in international standardized testing, students in Shanghai have surprised experts by outscoring their counterparts in dozens of other countries, in reading as well as in math and science, according to the results of a respected exam.” The European and Spanish press also underlined the progress of Asian students, *Le Figaro* pointed out that “The results of the PISA 2009 report put Shanghai and South Korea at the top.”; while *Le Point* remarked: “PISA results: The secrets of the Chinese triumph. Young Chinese students achieve great success in the OECD evaluation.”; while in Spain, *El Mundo* remarked: “Five Asian countries and two from Oceania in the international educational top 10.” These are just a few examples of the many news items where the concern was to establish comparisons with countries with better marks in the report.

FINAL CONSIDERATIONS

Our final considerations are directly related to our original aim, which was to analyse the presence in the scientific literature, on the one hand, and in the daily press, on the other, of such a news-making question as the PISA programme, set in operation by the OECD in the late 1990s. To this end, we used the two most popular scientific databases, *WoS* and *Scopus*, to which we incorporated the visualization of the PISA question through *Google Scholar*, which is a new, open, free search engine of great power. The daily press was consulted for news items on PISA during the period from 5th to 10th December 2010, when the report was published.

We first made a broad consideration of the information provided by these three powerful search instruments by viewing the thematic areas. We attempted to answer the questions: Does PISA correspond to a particular thematic area? What treatment does the scientific literature offer in this respect? It was found that in the main the PISA question corresponds to a very broad thematic area in the field of

the Social Sciences, principally Education, but with presence in other areas such as Economics, Sociology, Psychology, Teaching of Mathematics, History and even Philosophy, thus giving a multidimensional aspect.

In addition, we determined which journals held the highest number of studies published on the subject and the most cited studies and authors using *WoS*, *Scopus* and *GS*, as well as the possibilities of each search device. The German journals *Zeitschrift für Pädagogik* and *Zeitschrift für Erziehungswissenschaft* concentrated most studies and authors, which leads us to an initial consideration of the repercussion the PISA report has had in Germany and the subsequent reforms introduced into its educational system as a whole. Apart from comparisons with other international tests, which occupy a large sector of the publications, we also found treatments at a more specific level regarding the relation of the tests with the “g” intelligence factor, with classroom learning techniques, or with the effects of specific divisive social contexts. *Google Scholar* also allowed us to examine work in different European languages, and we were able to visualize the presence of PISA through its fields of influence in the scientific press.

The daily press produced an extraordinary amount of information in the space of just one week, especially just a few hours and days after publication of the report by the OECD. The information concentrated on highlighting above all the positions of the participating countries and communities in comparison with the highest ranking countries, which gave rise to a plethora of interpretations and evaluations.

In general, we found a broad variety of contextualized responses with a very considerable supply of complementary documentation that will help to interpret many of the key issues yet to be analyzed and explained.

ACKNOWLEDGMENT

We wish to thank Prof. Romuald Normand of the IFE-ENS in Lyon (France) and Prof. Miguel A. Pereyra of the University of Granada (Spain) for their suggestions and comments on the composition of this text.

NOTES

- ¹ In Social Sciences & Humanities - Arts and Humanities- Business, Management and Accounting- Decision Sciences- Economics, Econometrics and Finance- Psychology- Social Sciences- Multidisciplinary Scopus has access to some 5300 titles, in comparison to ISI-WoS, which has around 3500 in this area.
- ² All the searches involved in this study were concluded by the end of February 2011.
- ³ Go to <http://home.arcor.de/p.ulrich/extra/Text1.pdf> (16/06/2010), although using Open Access anyone can get access from their computer through Google Scholar (<http://scholar.google.es/>) and the search equation [PISA and OECD].
- ⁴ This bibliometric factor is calculated on the basis of the distribution of citations made of a researcher’s scientific publications.
- ⁵ An example is the extensive study in Spanish on “PISA en Alemania”, coordinated by Hans-Georg Kotthoff and Miguel A. Pereyra (2009) in the monographic edition of *Profesorado. Revista de Curriculum y Formación del Profesorado*, 13 (2), which can be consulted at <http://www.ugr.es/~recfpro/Rev132.html>.

REFERENCES

- Aburrá, L. (2006). *Pisa 2003: bravi come gli altri. Nuova luce sulle competenze dei quindicenni dal confronto fra regioni italiane ed europee*. Milano: Franco Angeli
- Adams, R. & Wu, M. (2002). *PISA 2000 technical report*. Paris: OECD. Available at: <http://www.oecd.org/dataoecd/53/19/33688233.pdf>
- Acevedo, J.A. (2005). TIMSS y PISA. Dos proyectos internacionales de evaluación del aprendizaje escolar en Ciencias. *Revista Eureka sobre enseñanza y divulgación de las ciencias*, 2 (3), 282–301. Available at: <http://cimm.ucr.ac.cr/ojs/index.php/eudoxus/article/view/433/432>
- Amador, J.C. (2008). La evaluación y el diseño de políticas educativas en México. Centro de Estudios Sociales y de Opinión Pública. Documento de Trabajo núm. 35. Available at: http://www3.diputados.gob.mx/camara/content/download/177008/424704/file/Educacion_docto_35.pdf
- Bar-Ilan, J. (2008). Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271.
- Baudelot, C. & Establet, R. (2009). *L'Élitisme républicain. L'école française à l'épreuve des comparaisons internationales*. Paris: Seuil
- Batagelj, V. & Mrvar, A. (2010). Pajek. Program for analysis and visualization of large networks. Reference manual. Faculty of Social Sciences Ljubljana (Slovenja). Available at: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf>
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich (pp. 159–202). In J. Baumert et al., (Eds.) *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich*. Berlin: Deutsches PISA-Konsortium.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Eds.) (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske Budrich. Available at: <http://home.arcor.de/p.ulrich/extra/Text1.pdf>
- Bodin, A. (2005). Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français. Joint Finnish-French Conference, Teaching mathematics: Beyond the PISA survey (Paris 6 – 8 octobre 2005). Available at: http://www.apmep.asso.fr/img/pdf/com_pisa_ff_francais-2.pdf
- Bosman, J., Mourik, I. van, Rasch, M., Sieverts, E. & Verhoeff, H. (2006). *Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar*. Utrecht: Utrecht University Library. Available at: http://igitur-archive.library.uu.nl/darlin/2006-1220-200432/scopus_doorgeleken_&_vergeleken_-_translated.pdf
- Brunner, J. J. (2007). Los procesos de selección en los países participantes en PISA 2003. In: J. J. Brunner & C. Peña (Eds.) *La reforma al sistema escolar: aportes para el debate*. Santiago de Chile: Universidad Diego Portales.
- Bybee, R., McCrae, B. & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching*, 46(8), 865–883.
- Calero, J. & Escardibul, J. O. (2007). Evaluación de servicios educativos: el rendimiento en los centros públicos y privados medido en PISA-2003. *Hacienda Pública Española-Revista de Economía Pública*, 183, 33–66. Available at: http://cursos.cepcastilleja.org/uploaddata/1/documentos/rendimiento_centros_pisa_2003.pdf
- Dalud-Vincent, M. & Normand, R. (2011). Analyse textuelle et analyse de réseaux: Exemple du traitement d'une base de données bibliographiques à l'aide des logiciels Alceste et Pajek. *Bulletin de Méthodologie Sociologique*, 109, 20–38.
- Damelang, A. & Haas, A. (2006). *Arbeitsmarkteinstieg nach dualer Berufsausbildung – Migranten und Deutsche im Vergleich. Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung (IAB- Forschungsbericht, 17/2006)*. Available at: <http://www.econstor.eu/handle/10419/26696>

- Dei, M. (2006). *Economia e società nella cultura dei giovani. Rappresentazioni e credenze degli studenti medi*. Milano: Franco Angeli
- De Nooy, W., Mrvar, A. & Batagelj, V. (2007). *Exploratory social network analysis with Pajek*. Cambridge and New York: Cambridge University Press
- De Olano, D., Knodel, Ph., Martens, K. & Popp, M. (2010). Das PISA-Echo – Resonanzen und Erklärungsansätze. In Ph. Knodel, K. Martens, D. de Olano & M. Popp (Eds.) *Das PISA-Echo Internationale Reaktionen auf die Bildungsstudie*. Frankfurt/New York: Campus Verlag, (pp.9–25).
- Duit, R. & Treagust, D. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688.
- Ehmke, T. & Baumert, J. (2007). Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006 (pp. 309–339). In PISA-Konsortium Deutschland (Ed.) *PISA 06. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster.
- Fuchs, Th. & Wößmann, L. (2007). What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economics*, 32(2), 433–464.
- Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, 24(1), 23–37.
- Harzing, A.W. (2007a). *Google Scholar a new data source for citation analysis*. Available at: http://www.harzing.com/resources.htm#pop_gs.htm
- Harzing, A.W. (2007b). *Reflections on Google Scholar*. Available at: http://www.harzing.com/resources.htm#pop_gs.htm
- Hirsch, D. (2002). Notre enseignement est-il de bonne qualité? L'enquête du programme PISA. *OECD Observateur* (mars 2002). Paris: OECD. Available at: <http://www.observateurocde.org/>
- Jacsó, P. (2006). Open access scholarly databases—A bird's eye view of the landscape. In *Keynote address of the Asia & Oceania Section of Department of IFLA*, Seoul. Available at: <http://archive.ifla.org/IV/ifla72/papers/157-Jacso-en.pdf>
- Kotthoff, H. G. & Pereyra, M. A. (2009). La experiencia del PISA en Alemania. *Profesorado. Revista de Curriculum y Formación del Profesorado*, 13 (2). Available at: <http://www.ugr.es/~recfpro/Rev132.html>
- Kousha, K. & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*, 74(2), 273–294.
- Lau, K. Ch. (2009). A critical examination of PISA's Assessment on Scientific Literacy. *International Journal of Science and Mathematics Education*, 7(6), 1061–1088.
- Leydesdorff, L., Moya-Anegón, F. de & Guerrero-Bote, V.P. (2010). Journal maps on the basis of Scopus Data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, 61(2), 352–369.
- Lynn, R. & Mikk, J. (2009). National IQs predict educational attainment in math, reading and science across 56 nations. *Intelligence*, 37(3), 305–310.
- Martens, K., Nagel, A. K., Windzio, M. & Weymann, A. (2010). *Transformation of educational policy*. London and New York: Palgrave-Macmillan
- Mayr, Ph. & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online information review*, 31(6), 814–830. Available in: <http://eprints.rclis.org/archive/00011202/>
- Meho, L.I. & Yang, K. (2007). A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125.
- Nardi, E. (2002). *Come leggono i quindicenni. Riflessioni sulla ricerca OCSE-PISA*. Milano: Franco Angeli.
- Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *Libri*, 55, 170–180.
- OECD (2001). *Connaissances et compétences: des atouts pour la vie. Premiers résultats de PISA 2000*. Paris: OECD
- Pauly, D. & Stergiou, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI's Citation Index and Google's Scholar service. *Ethics in Science and Environmental Politics*, 33–35. Available at: <http://www.int-res.com/articles/esep/2005/E65.pdf>

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

- Perelman, S.; Pestieau, P. & Santin, D. (2009). Pourquoi les performances des élèves flamands et francophones sont-elles si différentes? Une analyse par la méthode des frontières stochastique. *Working Papers Series*, 2009/06. Liège, Belgium: CREPP-Centre de Recherche en Economie Publique et de la Population.
Available at: <http://www2.ulg.ac.be/crepp/papers/crepp-wp200906.pdf>
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neutrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Eds.) (2004). *PISA 2003: der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*, PISA-Konsortium Deutschland. Münster: Waxmann Verlag.
Available at: http://pisa_band_i_ergebnisse_1455-6-4.pdf
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (2006). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*, PISA-Konsortium Deutschland. Münster: Waxmann Verlag.
Available at: http://pisa.ipn.uni-kiel.de/zusammenfassung_PISA2006.pdf
- Revista de Educación (2006). PISA. Programa para la evaluación internacional de alumnos. *Número Extraordinario*, 1-511.
- Rico, L. (2006). La competencia matemática en PISA. *PNA* 1(1), 3-22. Available at: <http://cumbia.ath.cx/pna/imgs/Num1TextoLR.pdf#page=5>
- Rindermann, H. (2007). The g-Factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, 21, 667-706.
- Robin, I. & Rocher, Th. (2002). La compétence en lecture des jeunes de 15 ans: une comparaison internationale. *Education et Formation*, 2, 93-101. Available at: http://www.insee.fr/fr/ffc/docs_ffc/donsoc02c.pdf
- Rodrigo, L. (2006). Argentina en el informe PISA 2000. In M. García Lastra et al., (Eds.). *Convergencia con Europa y cambio en la universidad*. Santander (XI Conferencia de Sociología de la Educación). Available at: <http://dialnet.unirioja.es/servlet/articulo?codigo=2377234>
- Rodríguez, R. (2005). México en los resultados PISA 2003. Una interpretación no catastrofista. *Revista Mexicana de Investigación Educativa*, 10(24), 255-266.
- Torres-Salinas, D.; Ruiz-Pérez, R. & Delgado-López-Cózar, E. (2009). Google Scholar como herramienta para la evaluación científica. *El Profesional de la Información*, 18(5), 501-510.
- Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41(4), 455-470.

Antonio Luzón
Faculty of Educational Sciences
University of Granada

Mónica Torres
Faculty of Educational Sciences
University of Granada

**ANNEX 1. HEADLINES IN NATIONAL, REGIONAL AND INTERNATIONAL
NEWSPAPERS WITH REFERENCE TO THE PISA REPORT 2009**

Table 1. Newspapers with highest print run and circulation (2009-2010)

Newspaper	Print Run (daily)	Circulation (daily)
<i>El País</i>	497,597	383,420
<i>El Mundo</i>	390,831	292,608
<i>ABC</i>	331,810	251,337
<i>La Vanguardia</i>	231,281	197,503
<i>El Periódico</i>	174,960	138,454
<i>La Razón</i>	166,006	118,862
<i>Publico</i>	130,294	83,497

Source: Oficina de Justificación de la Difusión (OJD)

Table 2. Regional newspapers with highest circulation (2009-2010)

Newspaper	Print Run (daily)	Circulation (daily)
<i>El Correo</i>	127,508	106,684
<i>La Voz de Galicia</i>	113,899	98,829
<i>Diario Vasco</i>	87,831	74,146
<i>Diario de Navarra</i>	58,773	49,065
<i>Heraldo de Aragón</i>	58,352	48,615
<i>La Verdad de Murcia</i>	40,478	33,463
<i>Diario Montañés</i> (Cantabria)	39,864	33,774
<i>Ideal de Granada</i>	38,675	30,114
<i>Norte de Castilla</i>	37,997	32,386
<i>Diario de las Palmas</i>	28,784	23,669
<i>Diario de Mallorca</i>	24,149	24,143
<i>La Rioja</i>	17,499	14,790
<i>Correo de Andalucía</i>	17,439	13,319
<i>La Voz de Asturias</i>	8,895	6,709

Source: Oficina de Justificación de la Difusión (OJD)

Table 3. Headlines published in the main Spanish newspapers

Newspapers	Headlines
<i>El País</i>	<p>“España recupera en PISA el bajón de 2006, pero sigue a 12 puntos de la media de la OCDE”</p> <ul style="list-style-type: none"> • Un 36% de repetidores lastra las puntuaciones medias.- Solo el 3% de los alumnos está en los niveles más altos de resultado.
<i>El Mundo</i>	<p>“El nuevo estudio PISA consagra el estancamiento de la educación española”</p> <ul style="list-style-type: none"> • España recupera su posición de 2003, pero sigue por debajo de la media. • El informe se centra en esta ocasión en la comprensión lectora de los alumnos

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

	<ul style="list-style-type: none"> • Tanto la OCDE como el Ministerio interpretan nuestro resultado como estable.
<i>ABC</i>	<p>“La educación española continúa por debajo de la media en el informe PISA”</p> <ul style="list-style-type: none"> • Cinco claves para entender el sistema educativo español • Sin motivación, ni expectativas de futuro, perfil del alumno repetidor español • La educación española continúa por debajo de la media en el informe PISA • Estancados en la mediocridad.
<i>La Vanguardia</i>	
<i>El Periódico</i>	<p>“España se estanca en los niveles medios en el informe PISA”</p> <ul style="list-style-type: none"> • Mejora en comprensión lectora pero sigue 13 puntos por debajo de la media de la OCDE. • Un 20% de estudiantes están por debajo del nivel requerido.
<i>La Razón</i>	<p>“España el país de los estudiantes repetidores”</p> <ul style="list-style-type: none"> • Nuestro sistema sigue por debajo de la media en Matemáticas, Ciencia y lectura / Educación se escuda en que uno de cada tres alumnos no pasa de curso.
<i>Publico</i>	<p>“PISA prueba que la inversión no basta para el éxito escolar”</p> <ul style="list-style-type: none"> • El contexto socioeconómico, el entorno familiar y la formación del profesorado son claves.

Source: By the authors

Table 4. Headlines in national, regional and provincial newspapers concerning PISA results in autonomous communities

<i>Newspapers</i>	<i>Headlines</i>
<i>El País</i>	<p>“Déficit de alumnos excelentes”</p> <ul style="list-style-type: none"> • Solo el 2% de los estudiantes andaluces examinados en PISA logra la puntuación máxima en lectura. • El 26% de los chicos está en los niveles más bajos. <p>“Un país, dos escuelas”</p> <ul style="list-style-type: none"> • Las autonomías del norte superan con holgura el examen; el sur suspende. • Los expertos lo atribuyen a factores socioeconómicos, culturales y de gestión. • Los alumnos catalanes superan en lectura, matemáticas y ciencias al conjunto de España y la OCDE. • Se reduce la distancia entre centros públicos y concertados en las pruebas PISA. • Andalucía vuelve a estar a la cola de España en los resultados Pisa.
<i>El Mundo</i>	<p>“Las dos españas se perpetúan”</p> <ul style="list-style-type: none"> • Las comunidades de la mitad norte superan con creces a

- las de la mitad sur
- Castilla y León se consagra como la región española que tiene mejor nivel.
 - Los alumnos madrileños destacan en Lectura y Ciencias, pero flojean en Matemáticas
 - Castilla y León revalida su 'sobresaliente' en Ciencias, Matemáticas y Lectura
 - Los vascos brillan en matemáticas pero sacan un 'aprobado raspado' en ciencias.
- ABC*
- Los alumnos andaluces empeoran en Matemáticas y Ciencia y siguen a la cola.
 - Galicia se lleva todas las calabazas de la zona norte, según el Informe PISA.
 - El nivel de los alumnos catalanes remonta, pero sigue bajo la media en ciencias.
 - Alumnos riojanos, segundos de España en dos categorías del Informe Pisa 2009.
 - El informe PISA echa por tierra la labor de la Junta de Andalucía durante las últimas décadas, que no se centra en la educación sino en la propaganda.
 - La Comunidad de Castilla-León supera en comprensión lectora, Matemáticas y Ciencias a la media de la OCDE.
 - La región de Castilla-León supera a Estados Unidos, Francia y Alemania en educación.
- La Vanguardia*
- La calidad de la educación catalana reacciona y se sitúa a sólo dos puntos de la media de la OCDE en el informe PISA.
- El Periódico*
- Los alumnos catalanes superan la media de los países de la OCDE, según el informe PISA.
- El Correo*
- El informe PISA revela que los alumnos vascos están en los niveles de la OCDE y por encima de la media española.
- Diario Vasco*
- La educación española continúa por debajo de la media en el informe PISA.
 - Los escolares vascos destacan en matemáticas y flaquean en ciencias.
- Heraldo de Aragón*
- Los alumnos aragoneses, entre los mejores del país en todas las materias.
 - PISA alerta del alto riesgo de exclusión social de los repetidores españoles.
- Norte de Castilla*
- Castilla y León supera la media en comprensión lectora, Matemáticas y Ciencia.
 - La comunidad se sitúa, de esta manera, por encima de la tónica dominante en España e incluso en la OCDE.
- Diario de Navarra*
- Los escolares navarros superan el examen PISA mejor que la media nacional. Sin embargo, quedan lejos de países punteros como Finlandia o Corea del Sur.
- La Verdad*
- La Región de Murcia es la cuarta peor en lectura, matemáticas y ciencias.
 - PISA 2009: de la quiebra educativa al colapso económico.

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

<i>El Correo de Andalucía</i>	<ul style="list-style-type: none"> • La escuela andaluza sigue anclada en 2006 pese a la ley. La Junta confió en la LEA para mejorar los malos datos de 2006, publicados al poco de aprobarse la ley. • El informe PISA, que acaba de hacerse público, sitúa a Andalucía por debajo de la media española y de la OCDE en comprensión lectora y rendimiento escolar en general.
<i>Ideal de Granada Diario de las Palmas</i>	<ul style="list-style-type: none"> • La educación española continúa por debajo de la media en el informe PISA. • PISA da un aprobado a los Immigrantes. Los alumnos extranjeros escolarizados en Canarias ofrecen similares resultados en su rendimiento que los alumnos nativos de las Islas en la competencia lectora, científica y matemática, según el informe evaluador.
<i>Diario de Mallorca La Rioja</i>	<ul style="list-style-type: none"> • Los alumnos de Balears, a la cola de España en lectura. • PISA: esforzarse o fracasar. • Los alumnos riojanos, entre los mejores de España y otros 64 países de la OCDE. • El estudio, ¿marca la diferencia?
<i>La Voz de Galicia</i>	<ul style="list-style-type: none"> • Galicia mejora en lectura aunque sigue por debajo de la media de la OCDE. • Los sindicatos, preocupados por las desigualdades educativas entre autonomías.
<i>Diario Montañés</i>	<ul style="list-style-type: none"> • Los escolares cántabros mejoran en Lectura pero se estancan en Matemáticas y Ciencias. • PISA 2009 y Cantabria
<i>La Voz de Asturias</i>	<ul style="list-style-type: none"> • Asturias sube en lectura pero se estanca en ciencias. • El norte y el sur educativo.

Source: By the authors

Table 5. Headlines in International Newspapers

Newspapers	Headlines
F R A N C E	<ul style="list-style-type: none"> • L'école française mal classée et jugée injuste. <ul style="list-style-type: none"> ○ La France en chute sur les mathématiques. ○ L'école ne joue plus son rôle d'ascenseur social. • Dix leçons du classement PISA 2009. • La France, pays du grand écart scolaire <ul style="list-style-type: none"> ○ Tous concluent que notre système scolaire ne parvient pas à favoriser la réussite des élèves.
	<ul style="list-style-type: none"> • Les méthodes pédagogiques françaises mises à mal. <ul style="list-style-type: none"> ○ L'étude de l'OCDE révèle un système de plus en plus inégalitaire. • Systèmes éducatifs: la France moyenne. <ul style="list-style-type: none"> ○ Les élèves français de 15 ans ont des résultats dans la moyenne des pays de l'OCDE mais les inégalités scolaires se sont accrues en France depuis l'an 2000,
	<i>Le Figaro</i>

- selon les résultats de l'enquête "Pisa 2009"
publiée qui place Shanghai et la Corée du
Sud en tête.
- Le Point*
- Les inégalités s'aggravent dans l'école française, dit l'OCDE.
 - Classement Pisa: les secrets du triomphe chinois
 - Les jeunes élèves chinois cassent la baraque à l'évaluation de l'OCDE. Ce qui n'empêche pas Pékin de vouloir mieux faire.
- G
R
E
A
T
- The Guardian*
- UK schools slip down world rankings.
 - OECD study shows that despite comparatively high levels of per-pupil spending, the UK is behind Poland and Norway.
 - World education rankings: which country does best at reading, maths and science?
 - The OECD's comprehensive world education ranking report, PISA, is out.
- B
R
I
T
A
I
N
- The Times*
- UK schools tumble down world table
 - Labour policies blamed as British pupils are outperformed by students across the world, with results in Wales found to be the worst of all.
- Financial Times*
- Why are Chinese schoolkids so good?
- The New York Times*
- Top Test Scores From Shanghai Stun Educators.
 - Western Nations React to Poor Education Results.
 - Can you imagine the reaction if we told the students of Chicago that the PISA was an important international test and that America's reputation depended on them performing well?" Mr. Schneider said. "That said, China is taking education very seriously. The work ethic is amazingly strong."
- U
S
A
- The Washington Post*
- What international test scores really tell us: Lessons buried in PISA report.
 - Unfortunately, federal and state policies do little to adopt these factors that other nations have found so successful.
 - Hysteria over PISA misses the point.
- I
T
A
L
Y
- La Repubblica*
- OCSE, migliorano gli studenti italiani si riduce il divario con gli altri Paesi
 - Per la prima volta dopo dieci anni risultati confortanti per i nostri ragazzi in Lettura, Matematica e Scienze. Nella classifica internazionale recuperano da una a sei posizioni .

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

	<i>Corriere Della Sera</i>	<ul style="list-style-type: none"> • Classifiche Ocse: scuola, i 15enni italiani migliorano, in Lombardia i più bravi. <ul style="list-style-type: none"> ○ Riduzione del divario con gli altri Paesi coinvolti, ma permangono ancora sacche di insufficienza.
G E R M A N Y	<i>Berliner Zeitung</i>	<ul style="list-style-type: none"> • Pisa-Studie. Deutsche Schüler jetzt Mittelmaß <ul style="list-style-type: none"> ○ Bildung in Deutschland - "Zehn Jahre Pisa-Studie haben dem deutschen Bildungssystem gut getan", sagte Bundesbildungsministerin Annette Schavan (CDU) bei der Vorstellung der Studie.
	<i>Die-Welt (Berlin)</i>	<ul style="list-style-type: none"> • Migranten und sozial Schwache steigern Pisa-Werte <ul style="list-style-type: none"> ○ Laut Pisa-Studie ist der Bildungs-Abstand zwischen Angehörigen höherer und niedrigerer sozialer Schichten kleiner geworden.
	<i>Frankfurter Allgemeine Zeitung</i>	<ul style="list-style-type: none"> • Deutschland ist aufgestiegen <ul style="list-style-type: none"> ○ Mit Erleichterung reagieren Bildungsforscher, Lehrer und Politiker auf die verbesserten Ergebnisse deutscher Schüler beim Pisa-Test. • Heilsamer PISA-Schock. Sensibilität statt Schockstarre. <ul style="list-style-type: none"> ○ Zehn Jahre nach dem ersten Pisa-Test im Jahre 2000 liegt Deutschland beim Lesen im OECD-Schnitt, in Mathematik und Naturwissenschaften deutlich darüber und zählt zu den wenigen Ländern, die sich durchgängig verbessert haben.
S W E D E N	<i>Skolverket</i>	<ul style="list-style-type: none"> • Reading literacy of 15-year-olds and equity in the school have decreased. <ul style="list-style-type: none"> ○ Reading literacy and the knowledge and skills of Swedish 15-year-olds in mathematics have decreased over the last decade. In the natural sciences Swedish students today perform below the international average. Equity in the Swedish school has decreased, and the number of students not achieving the lowest reading literacy levels is growing.

Source: By the authors

ANNEX 2. MOST CITED STUDIES ON PISA IN DIFFERENT LANGUAGES OF EUROPEAN UNION COUNTRIES.

Table 6. Most cited studies in GS in German (de), Spanish (sp), French (fr) and English

German: 8230 records in GS (de)

<i>Publications</i>	<i>Year</i>	<i>Citations</i>	<i>Authors</i>
<i>PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich,</i>	2001	654	Baumert, J.; Klieme, E.; Neubrand, M.; Prenzel, M.; Schiefele, U.; Schneider, W.; Stanat, P.; Tillmann, K.-J. and Weiß, M.
“Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich”	2002	504	Jürgen Baumert y Gundel Shümer
<i>PISA 2003: der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs</i>	2004	176	Prenzel, M.; Baumert, J.; Blum, W.; Lehmann, R.; Leutner, D.; Neutrand, M.; Pekrun, R.; Rolf, H.-G.; Rost, J. and Schiefele, U.
<i>Soziale Herkunft und Kompetenzerwerb: Vergleiche zwischen PISA 2000, 2003 und 2006</i>	2007	38	Ehmke, T. and Baumert, J.
<i>PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie</i>	2007	37	Prenzel, M.; Artelt, C.; Baumert, J.; Blum, W.; Hammann, M.; Klieme, E. und Pekrun, R.
“Arbeitsmarkteinstieg nach dualer Berufsausbildung– Migranten und Deutsche im Vergleich”	2006	17	Damelang, A. and Haas, A

Spanish: 2830 records in GS (sp)

<i>Publications</i>	<i>Year</i>	<i>Citations</i>	<i>Authors</i>
<i>Revista de Educación. N° Monográfico. “PISA. Programa para la Evaluación Internacional de los Alumnos”</i>	2006	52	Marchersi, A., Schleicher, A., Turner, D.
<i>Revista Hacienda Pública Española- Revista de Economía Pública, 183</i>	2007	28	Calero, J., Escardibul, J. O.
<i>Revista Pensamiento Numérico (1) “La competencia matemática en PISA”</i>	2006	22	Rico, L.
<i>Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 2</i>	2005	22	Azevedo, J. A.
“Los procesos de selección en los países participantes en PISA 2003”, en Chile	2007	5	Brunner, J.
“Argentina en el Estudio PISA 2000”	2006	4	Rodrigo, L.
“México en los resultados PISA 2003. Una interpretación no catastrofista”	2005	3	Rodríguez, R.

VISUALIZING PISA SCIENTIFIC LITERATURE VERSUS PISA PUBLIC USAGE

“La evaluación y el diseño de políticas educativas en México” 2008 2 Amador, J.C.

French: 1440 records in GS (fr)

<i>Publications</i>	<i>Year</i>	<i>Citations</i>	<i>Authors</i>
<i>PISA 2000 Technical Report</i>	2002	179	Adams, R. and Wu, M.
“Caractéristiques des systèmes éducatifs et compétences des jeunes de 15 ans: l'éclairage des comparaisons entre pays”. IREDU	2004	34	Duru-Bellat, M; Mons, N. et Suchaut, B.
<i>L'Élitisme républicain. L'école française à l'épreuve des comparaisons internationales</i>	2009	7	Baudelot, Ch. et Establet, R.
<i>Connaissances et compétences: des atouts pour la vie. Premiers résultats de PISA 2000</i>	2001	3	OCDE
“La compétence en lecture des jeunes de 15 ans: une comparaison internationale”, <i>Education et Formation, 2</i>	2002	5	Robin, I. et Rocher, Th.
“Notre enseignement est-il de bonne qualité? L'enquête du programme PISA”, <i>OCDE Observateur</i>	2002	3	Hirchs, N.
“Pourquoi les performances des élèves flamands et francophones sont-elles si différentes? Une analyse par la méthode des frontières stochastique”, <i>Working Papers Series, 2009/06</i>	2009	3	Perelman, S., Pestieau, P. et Santin, D.
“Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français”	2005	2	Bodin, A.

English: 47.600 records in GS

<i>Publications</i>	<i>Year</i>	<i>Citations</i>	<i>Authors</i>
“What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data” (CESifo <i>Working Paper No. 1235</i>)	2004	194	Thomas Fuchs and Ludger Woessmann
<i>PISA 2000 Technical Report</i>	2002	179	Adams, R. y Wu, M.
“Cross-Country Efficiency of Secondary Education Provision. A Semi-Parametric Analysis with Non Discretionary Inputs” (Working Paper, N° 494)	2005	109	António Afonso and Miguel St. Aubyn
<i>Reading for change: performance and engagement across countries: results from PISA 2000</i> . OECD	2002	65	I. Kirchs
“Culture, Gender, and Math” (Supporting Online Material)	2008	54	Luigi Guiso; Ferdinando Monte; Paola Sapienza y

			Luigi Zingales
“Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries” (Working Paper N° 1981)	2007	51	Ludger Woessmann
“International Comparisons of Student Attainment: some Issues Arising from the PISA Study” (Assessment in Education, IOE, UK)	2004	50	Harvey Goldstein
“PISA: What Makes the Difference? Explaining the Gap in PISA Test Scores Between Finland and German” (Discussion Paper No. 04-04)	2004	42	Andreas Ammermüller
<i>15-up and counting, reading, writing, reasoning ...: how literate are Australia's students?: the PISA 2000 survey of students' reading, mathematical and scientific literacy skills</i>	2001 2009	34	Lokan, Jan; Greenwood, Lisa; Cresswell, John
<i>Student Achievement in England: Results in Reading, Mathematical and Scientific Literacy among 15-year-olds from OECD PISA 2000 Study.</i> (London: The Stationery Office)	2002	14	Gill, Baljit, Dunn, Mark, Goddard, Eileen
<i>Learning for tomorrow's world. First results from PISA 2003</i>	2004	6	OECD
Science Competencies for Tomorrow's World	2007	6	OECD
<i>Knowledge and Skills for Life—First results from PISA 2000</i>	2001	3	OECD

ANNEX 3. IMAGES ABOUT PISA IN THE INTERNATIONAL PRESS

Figure 1. Ranking of European countries in PISA 2009.



Source: Die Welt (Berlin) (7-12-10). Bildungsstudie

ANNEX II

THE POSTER EXHIBITION

Previous a presentation of their content by their authors in the Teatro Chico, the Poster Exhibition took place at the Casino (Círculo de Instrucción y Recreo) of Santa Cruz de La Palma (located at Anselmo Pérez de Brito Street, 15). The posters exhibited were the followings posters with their abstracts:

A PORTRAIT OF EUROPEAN TOP PERFORMERS IN SCIENCE

Fabio Alivernini
Laura Palmerio
Valeria Tortora
National Institute for the Evaluation of the Education System
Frascati (RM) Italy

E-mail: tortora77@gmail.com

Science proficiency levels in PISA refer to the various different competencies of students in science subjects (OECD, 2007). The students' scores are divided into six proficiency levels, with levels 5 and 6 at the top (meaning that students have those competencies that are essential for the creation of innovative new technology) and Level 1 at the bottom (referring to students who do not possess even the most basic science skills and knowledge). 15 year-old students at levels 5-6 are therefore the top performing students in Europe regarding their future possibilities to participate effectively in real-life situations related to science and technology.

The aim of this project is to identify, from the European PISA 2006 data, the pattern of variables which are distinguishing features of those students performing at the highest science proficiency levels. The theoretical framework of the study is consistent with the PISA 2006 assessment framework (OECD, 2006) that considers a large number of contextual factors which could, hypothetically, influence students' performance in science.

The contextual and individual factors include:

- various aspects of schools, such as the quality of human and material resources, the nature of public or private control and funding and decision-making processes;
- the context of instruction, including institutional types and structures and class size;
- students and their families economic, social and cultural background and resources;
- students' attitudes such as students' self-efficacy, self-regulated learning, types of motivation and goals;
- various aspects of students' lives, such as their attitudes towards learning, their behavior and lifestyle at school.

All contextual variables are central (OECD, 2006) to the analysis of results within the theoretical framework referring to a range of student and school characteristics. Consequently an exploratory approach has been applied, which

ANNEX II

allows the identification of interaction effects which are not completely predictable *a priori*, among a large number of variables of different types (Nominal, Ordinal, and Interval).

The analysis has been based on classification and regression trees (CART) (Williams, Lee, Fisher, & Dickerman, 1999), a full non-parametric method designed to detect and interpret complex interactions that most traditional means of regression and classification analysis might ignore or find difficult to estimate and interpret (Allore, Tinetti, Araujo, Hardy, & Peduzzi, 2005).

This way it has been possible to identify two groups of science top performers, one characterized by a low value of the index of economic social and cultural status (ESCS), the other characterized by a high value of the same index.

The aim of this project is to depict a portrait of these two groups of students, in attempt to identify which features, besides different values of ESCS at school level, describe them better.

REFERENCES

- Allore, Heather, Tinetti, Mary E., Araujo, Katy L. B., Hardy, Susan & Peduzzi, Peter (2005). A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and Social and Productive Activities scores, *Journal of Clinical Epidemiology*, 58, 154–161.
- OECD (2007). *PISA 2006 Science competencies for tomorrow's world*. Paris: OECD
- OECD (2006). *Assessing Scientific, Reading and Mathematical Literacy – A Framework for PISA 2006*. Paris: OECD.
- Williams, Christopher J., Lee, Stephen S., Fisher, Rachel A. & Dickerman, Lois H. (1999). A comparison of statistical methods for prenatal screening for down syndrome. *Applied Stochastic Models in Business and Industry*, 15, 89–101.

THINKING ABOUT PISA AND THE PARTICIPATION OF THE PARENTS

Claudio Almonacid Águila

Ana M^a Jiménez Saldaña

Daniel Ríos Muñoz

University of Santiago de Chile

E-mail: claudio.almonacid@usach.cl; ana.jimenez@usach.cl; daniel.rios@usach.cl

The Chilean educational system is characterized by its quasimarket structure, being one of the few countries of the OECD that has this trait. In this quasi market three types of educational establishment compete: municipal, subsidized and private. The competition between these establishments is based in two principles established in the Constitution of the State: rights to education and freedom of schooling. In accordance with this last principle, on the one hand, parents have the right to choose the type of education they want for their children and, on the other, the State promotes the participation of society to create educational establishments. To achieve these goals, the State has created a financial system based in vouchers. This mechanism generates public - private competition to access public resources of the State. When analyzing the evolution of the number of students and number of

educational establishments, we can appreciate that the subsidized system has been consolidated compared to the municipal system, and this has produced a high social segmentation between both types of establishments.

Because parents could decide on the type of education they wanted for their children, in the '80s the State designed an annual system for measuring the quality of the education through the results of what was learned by the student (SIMCE). This information was distributed publicly as if it was a ranking of schools. The State expects that the parents, with this information, can choose to enroll to their children in a school of their liking. Also, Chile has decided to take part in various international measurements (PISA, TIMMS, LLECE), all of which show the low results obtained by the educational establishments and the inequality inside of the educational system, where the municipal schools have the poorest students and obtain the worst results.

This investigation, funded by the Spanish Agency of International Cooperation (AECI) attempts to analyze the dynamics of family participation as the basis of social cohesion and the strengthening of Chilean civil society. Some specific aims proposed are to analyze the educational politics of family participation in the educational system; examine the utility level of the mechanisms of direct family participation; study the recent changes produced in Chilean civil society as a result of the creation of the School Councils; and to understand the mechanisms and strategies of school election made by the families.

For this, methodologically, parent surveys made in the capital of Chile. The sample was stratified in accordance with the administrative patterns of the school system. These data were compared to a sample of parent surveys from a previous investigation, in order to observe its evolution and execute a national analysis.

The results of this analysis show that parents choose the school for their children based on their personal valuation of a range of characteristics such as infrastructure, pedagogy, social atmosphere and students' selection. Generally, parents tend to value the education that their children receive, even in schools with low results in the test of SIMCE. They do not tend to consider this information in their decision. In this sense, we can suggest that, in spite of the efforts made by State in publishing this information, for parents it is not relevant. Finally, if parents had to change their children's school, a great number of them would shift to the subsidized system. On the other hand, the data show that parents are beginning to take part in the School Councils and they value their participation as a way to improve the quality of education, while accepting the existence of a stratified educational system linked with the social class of students

REFERENCES

- Bellei, Cristian et al., (2008). *La agenda pendiente en educación*. Santiago: UNICEF- Universidad de Chile.
- Collins, Randall (1977). Some comparative principles of educational stratification, *Harvard Educational Review*, vol. 47 (1); 1-27.
- Luengo, Julián (Ed.) (2005). *Paradigmas de gobernanación y de exclusión social en la educación. Fundamentos para el análisis de la discriminación escolar contemporánea*. Barcelona: Ediciones Pomares.

ANNEX II

Marks, Gary Neil (2008). Are Father's or Mother's Socioeconomic Characteristics More Important Influences on Student Performance? Recent International Evidence, *Social Indicators Research*, vol. 85 (2), 293–309.

AN INVESTIGATION OF REASONS FOR FINLAND'S SUCCESS IN PISA

Dr. Jennifer H. Chung
Liverpool Hope University
E-mail: chungj@hope.ac.uk

Based on doctoral research at the University of Oxford

The Programme for International Student Assessment, or PISA, administered by the Organisation for Economic Cooperation and Development, surveyed fifteen-year-olds for the first time in 2000. PISA focuses on mathematics, science, and reading literacy, and intends to undertake a new study every three years. The administration in 2003 added a section that measured problem-solving skills. The results from all three surveys thus far have placed Finland as the highest achieving country in PISA.

Finland's top performances in PISA astonished the educational world. More so than previous cross-national surveys such as TIMSS and PIRLS, PISA has drawn worldwide educational interest towards Finland and its educational system. PISA, unlike TIMSS, does not measure mastery of curriculum, but rather outcomes of education. PISA focuses on real-life applications of knowledge. Triggering global curiosity, PISA has placed Finland on the map for those wishing to discover the influences behind educational success.

This research focuses on Finland's historical, cultural, and social context as a part of the Nordic countries and also its immense success in PISA. This project uncovers some of the factors contributing to Finland's success in education, as indicated by the results in PISA. Finland's history as a part of both Sweden and Russia has intertwined education with the movement for independence. The struggles after independence, including war and recession, have also reinforced the importance of education within Finnish society. The important status of teachers in Finland, in addition to their high quality, has further enhanced the excellence of the Finnish education system. PISA findings, however, have indicated that the Swedish-speaking Finns score lower than the Finnish-speaking Finns, a phenomenon explored within the research. This project investigates the reasons behind this counterintuitive result.

The project incorporates perspectives through interviews with both Swedish-speaking and Finnish-speaking teachers and head teachers from the six sample schools, Finnish education ministers, PISA creators from the OECD, and Finnish educational researchers and professors responsible for executing PISA in Finland. Their insight, from many different points of view, illuminated different perspectives on PISA and education in Finland.

FROM THE STATIC TO THE DYNAMIC: PISA ITEMS AS AN EDUCATIONAL
RESOURCE

José R. Galo Sánchez^{1, 2}

¹ Descartes Project, Institute of Education Technologies
Ministry of Education of Spain

² University of Cordoba, Spain (Department of Mathematics)
E-mail: ma1gasaj@uco.es

Part of the information that enables the elaboration of the educational indicators of PISA comes from the results obtained by the student body on standardized tests comprised of different items and designed to be done in a traditional way with pen and paper. The PISA items are the result of the work of experts carried out through a rigorous process of proposal, correction and selection. In these items the aim is to assure an appropriate conceptual framework which should be put through an appropriate stimulus (a text, a board, a diagram, etc.) and then followed by a certain number of associated exercises and questions with the aim to measure competence.

The publication of the PISA indicators usually creates social and political concern which has generated a growing demand for information on behalf of professionals implicated in the process of education. This demand is mainly concerned with knowing the type of tests used and the theoretical approach they uphold. For this reason, there have been numerous bibliographic publications where some of the items used in the evaluating process are shown. They are the so called “PISA released items”.

Not only are the released items interesting from an informative point of view, but they can also be used as an important educational resource in the competence formation of students. However, this possibility is limited by its static nature which does not allow for a student to reuse a test. Once a resource has been used, it becomes useless since the approach, as well as the answer, is known.

In this study, an alternative to this situation is provided whereby dynamism is introduced into the released items, enabling the same student to reuse the same item for his/her learning several times. To do this, it is necessary that in each access of an item the student observes a different layout in the stimulus and in the questions, that when numerical data appears he/she sees a different quantification each time. The alternative answers offered should change or permute the order of the presentation, in short, to ensure that the display appears different while maintaining the same structure and objective. What takes place is a reworking of the static item into a dynamic item.

The previous approach becomes effective in the ASIPISA project (Spanish acronym for “*Ayuda Sistemática Interactiva para PISA*” Interactive Systematic Assistance for PISA) which is accessible at the website <http://descartes.cnice.mec.es/heda/ASIPISA/>. Here, a collection of learning objects based on released items has been created. By introducing variability through random sources, repetition is avoided and educational potential increased. All this is put in a context of Information and Communication Technologies which can increase motivation

ANNEX II

and improve the competence learning of the student body. The objectives created cover all the PISA competence areas: Reading, Mathematics, Sciences and Problem Solving.

Technically, the material is configured as a set of “hyper-textual” documents (web pages) that comprise a repository of interactive educational resources with the same functional structure and which incorporate the automatic evaluation of the answers given by users. The evaluation is summative as it is in PISA. However, we have now introduced a formative evaluation, which is far more interesting in an educational formative context to which this project is directed.

For the codification of these objects the interactive kernel for educational programmes known as *Descartes* was used <http://descartes.cnice.mec.es>, within the homonymic project promoted by The Ministry of Education of Spain. The development of the ASIPIISA project was partially supported by The Education Board of the *Junta de Andalucía* - Regional Government of Andalusia (Spain).

PISA, FINLAND AND THE AUTONOMOUS COMMUNITY OF MADRID

María José García Ruiz
UNED
E-mail: mjgarcia@edu.uned.es

The current era of globalization is also the era of the prominence of international organisations. At the beginning of the XXI century, agencies such as the OECD have turned into real “political actors” (Rizvi and Lingard, 2009) that seek to influence world education according to their own political, economic and educational assumptions. A great diversity of academics (Altarejos, 2003; Goldstein, 2004; Simola, 2005) have emphasised the risk for the different education systems of arbitrarily implanting the increasingly pressing recommendations of these institutions. The now typical international studies of educational performance assessment contain two important biases that call for cautious reading of their conclusions. One bias is the lack of consideration of these studies for the context in which these studies are produced. Context evidently highly contributes to the generation of such results. The second limitation of these studies is the underestimation in many of these studies of the educational elements coming from tradition (i.e. McKinsey Report, 2007). We think that a global axiom should be that “everything that functions in education must not be touched”.

This comparative study of education in Finland and in the Community of Madrid will try to show, among other issues, that each education system holds its own excellent educational elements, even if they have not been internationally recognised as such. One confirmed hypothesis of this study establishes that, contrary to what could be inferred from a superficial reading of the PISA Report, Finland is not in the vanguard of international educational developments, and does not pretend to be so. This assertion, far from implying an unfavourable diagnosis for this country, constitutes a virtue of this system, for it confirms the solidity and safeguard of the educational policy already developed in this country, which demonstrates its exquisite balance between tradition and reform, and which only introduces

innovative aspects when they do not modify any aspect of proven validity in the system. Finland does not reveal to the international community a scandalous lack of innovation, but a lesson of greater validity: tradition functions, and the elements of tradition which function must not be altered.

Another hypothesis of this study stresses that the high scores of Finnish students in the PISA reports indicate the success of the comprehensive school *in the way that* it has been organised in Finland. The comprehensive school is not a univocal concept or a univocal reality. It is a model which has developed in different countries with different nuances in all its elements (structure, methods, curriculum, and didactics). In Finland this modality of secondary school has been developed by pedagogically conservative teachers who practise very traditional didactical methods.

Among the conclusions of this analysis we can cite the following:

1. It is necessary to have a great deal of caution with the national implantation of the international recommendations issued by international organisations.
2. It is sensible to promote consciousness of the value of educational tradition.
3. The education of the Community of Madrid reveals a particular excellence in three specific domains which are current challenges of Finnish education: infant education, cooperation with parents, and intercultural education.

REFERENCES

- Altarejos, Francisco et al., (2003). *Retos educativos de la globalización: hacia una sociedad solidaria*. Pamplona: Ediciones de la Universidad de Navarra.
- McKinsey & Company (2007). *How the world's best-performing school systems come out on top (McKinsey Report)*. McKinsey & Company (<http://www.mckinsey.com/>). Available online at: http://www.mckinsey.com/App_Media/Reports/SSO/Worlds_School_Systems_Final.pdf
- Goldstein, H. (2004). Education for all: the globalization of learning targets. *Comparative Education*, vol. 40 (1), pp. 7-14.
- Rizvi, F. & Lindgard, B. (2009). The OECD and global shifts in Education Policy. In Cowen, Robert and Kazamias, Andreas M. (Eds.): *International Handbook of Comparative Education*. London: Springer.
- Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education*, vol. 41 (4), 455-470.

DIGITAL SKILLS APPEAR IN PISA REPORT. ARE STUDENTS AND TEACHERS PREPARED?

Ana García-Valcárcel
Francisco Javier Tejedor
Anunciación Quintero
University of Salamanca (Spain)
E-mail: anagv@usal.es

The PISA report, of the Organization for Economic Cooperation and Development (OECD), for the first time, will measure the capacities of students to handle themselves in the digital era through a test of reading in electronic format that will

ANNEX II

be applied in 17 countries, among them Spain. The decision constitutes recognition of the importance of digital competences to handle oneself in the world.

According to the director of the Report, the goal is to measure the necessary resources for “accessing, handling, integrating and evaluating information; to construct new knowledge from electronic texts” (Schleicher, 2009).

The objective of this test extends the concept of reading capacity, although [the emphasis] “is not as much on technologies, but rather on the cognitive competences that they need for the effective use of the technology”. In this situation, we asked if the schools are preparing the students in this direction and if teachers have the necessary foundation and competences for the required digital abilities.

Considering a research based on the study of cases of schools where they develop experiences of innovation with TIC, we present data relative to the present situation of the teachers and students in relation to the integration of ICT in educational practice.

The data obtained show that:

1. Most students perform different functions with the computer (play, seek information, do classroom assignments and communicate).
2. Virtually all students like to use the computer in class, and there are not many differences in learning.
3. Most teachers have been trained in Internet navigation and use of search engines, also an important group of virtual platforms.
4. Collaborative projects and platforms are used least in the training of teachers in internet use.
5. The teachers are competent in the search and selection of technological resources of interest and raising ICT activities for their students.
6. Teachers show less competence in the design of digital learning resources

Findings suggest that teachers have acquired some skills in using ICT but still have major shortcomings for use in the classroom and that students do not find many differences in learning processes when working with ICT, although they like its use in the classroom.

REFERENCE

Schleicher, Andreas (2009). La era digital llega al Informe Pisa, *El País*. (02/09/2009). Available online at: http://www.elpais.com/articulo/educacion/era/digital/llega/Informe/Pisa/elpepusocedu/20090209elpepiedu_1/Tes?print=1

DIFFERENCE OF GENDER IN THE EDUCATIVE PRACTICES AND EXPECTATIONS OF THE STUDENTS

Carmen Nieves Pérez Sánchez
Moisés Betancort
University of La Laguna
E-mail: cperez@ull.es & moibemo@ull.es

Girls make better proficiency from their school stay than boys. They have better suitability levels with respect to age and academic year; they are less likely to leave after the minimum period of mandatory secondary school; they choose longer scholar trajectories, etc.

The study presented here, elaborated in the frame of a larger project “Class, gender, family and education achievement” starts from research during 2006-2007 with a sample of 2247 students at the 4^o course of ESO (*Educación Secundaria Obligatoria* or Compulsory Secondary Education) in Canary Islands. We find remarkable differences between boys and girls regarding the identification patterns, the academic expectations and the attitude at school. All these are the variables which best explained the superior scholarly adaptation of girls.

PISA ON FRONT PAGE

Ariadne Runte-Geidel
Diego Sevilla
University of Granada
E-mail: ariadnegeidel@hotmail.com & sevilla@ugr.es

The importance of information and communication is evident in our globalized world; hence, the influence of mass media on the formation of public opinion is highly relevant. Many studies (Bhugra, 1989; Wolff, 1997) have shown that mass media represent the main information sources for topics such as mental health.

Based on these results, we wondered what could be found in the Spanish news regarding the PISA report and how this information might impact on the social imaginary.

All the articles related to the PISA report included in the main Spanish newspapers between 2006 and 2009 were reviewed.

A total of 58 articles included in *El País*, *El Mundo*, *Público*, *Expansión* and *Cinco Días* (the last two specialized in economy) were reviewed. In all of them we found an implicit unanimity on the importance/relevance of the PISA results, independent of the newspaper’s political ideology. None of them showed any criticism or debate about the fact that PISA is the product of a commercial organization which is beyond the educational field.

The majority of the reviewed articles were focused on reporting the Spanish scores as compared to the average scores obtained in OECD and other countries. Only a minority of articles used these data to analyze more complex aspects of Spanish education.

Almost all the articles’ headlines and content interpreted the Spanish result in PISA in a negative fashion, using expressions such as “Spain fails”, “Education worsens”, “Spain at the very bottom...” or “Spanish students are the worst at sciences”.

Study limitations: according to the Association for the Investigation of Mass Media (Asociación para la Investigación de los Medios de Comunicación, AIMC, 2009), only 40% of the Spanish population reads the newspaper, against 88.5% that

ANNEX II

watch TV and 55.2% listen to the radio. However, we assume that news content doesn't differ significantly across the different media.

In our opinion, the fact that all the reviewed articles have focused on the ranking position offered by PISA would reinforce the myth of good and bad countries in education, strengthening the dichotomy between “winners and losers”.

Could this generic negativism in depicting the Spanish results in PISA generate frustration and discredit feelings? And could this only be useful to discourage the teachers' efforts even further? Could this make families escape to the “best” schools as a way to make up for the failure pointed out by the PISA report? Could these families' attitude contribute to the marginalization of the so called “worse” schools? We are also concerned about the impact of this negative way of understanding on the students' motivation in general: “if I am not up to standards anyway, why make an effort?” Finally, what does society think about all of this? Is society assuming its commitments to teaching its students or is it only blaming someone else?

Nevertheless, the fact that a topic such as education comes to the front page of the newspapers shows an intensification of the debate and a new openness to a more complex and profound analysis of the problems in the Spanish educational system.

REFERENCES

- Wolff, G. (1997). Attitudes of the media and the public. In Leff, Julian (Ed.): *Care in the Community: Illusion or Reality?* New York: John Wiley & Sons.
- Bhugra, Dinesh (1989). Attitudes towards mental illness, *Acta Psychiatrica Scandinavica*, 80 (1), 1–12.
- AIMC (2009). *Estudio General de Medios*. Available online at:
<http://download.aimc.es/aimc/02egm/resumegm209.pdf>

THE EDUCATIONAL AND SOCIOLABOUR GUIDANCE PROGRAM: AN INSTRUMENT TO FACILITATE THE SOCIOLABOUR TRANSITION

Lidia E. Santana Vega
José A. Santana Lorenzo
University of La Laguna, GIOES Group
E-mail: lsantana@ull.es

The socio-labour transition of younger generations is a challenge for all political systems and a greater challenge for the educational system of each country. One of the more important reports of international impact, the PISA report (the result of the tests applied to a statistically significant group of the entire population of 15 year olds from 41 countries), seeks to discover the ability to use knowledge, skills that allows people to face the difficulties of life and the ability to integrate into the real world. The different reports of PISA are not intended to assess content knowledge, but instead work as a study of functional literacy that incorporates mathematical content, scientific and language relevant to their insertion into the adult world (Pajares, 2005). Its purpose is to assess the ability of young people to

integrate into the world of work and further their education; in short, the assessment is most interested in measuring what students can do.

The complexity of contemporary society requires good academic and professional guidance plans to facilitate the transition of students between educational stages and active life. In this sense 1) Telling young people to get a job that gives them the happiness of “work as one likes” is not enough. Standard biographies and predictable itineraries underlie this idea. Now, biographies and work schedules are constructed and reconstructed to accommodate the changing fortunes of time. 2) If we want to prepare young people for the transition to adult and working life, the education system should look to the world of labor. 3) To educate and guide future generations, disciplinary or psycho-pedagogical knowledge is not sufficient. It is necessary to unravel the societal key and to assume that if the times change, so have the professional responsibilities of educators to help the students face the transition process and to maintain them throughout their lives (Santana Vega, 2003).

GIOES research group, at the University of La Laguna has developed a program aimed at developing young people’s capacity to integrate into the world of work from a global view of the proposed secondary curriculum.

REFERENCES

- OECD (2004). *Learning for tomorrow’s world. First results from PISA 2003*. Paris: OECD.
- Pajares, Ramón (2005). *Resultados en España del estudio PISA 2000. Conocimientos y destrezas de los alumnos de 15 años*. Madrid: MEC-INECSE.
- Santana Vega, Lidia E. (2003). *Orientación educativa e intervención psicopedagógica. Cambian los tiempos, cambian las responsabilidades profesionales*. Madrid: Pirámide (2009 3ª ed.).

SIMILARITIES AND DIFFERENCES BETWEEN SPANISH AND FINNISH TEACHERS: ORGANIZATION OF SCHOOL TIME AND SPACE, SELECTION AND SALARY

Begoña M^a Zamora Fortuny
University of La Laguna (School of Education. Department Sociology of Education)
E-mail: bezamora@ull.es

Finland is the country with the best educational outcomes in PISA Reports. For this reason, we look to Finland as an example for others and we try to discover the Keys to its success. The aim of this study is to uncover some of these clues, beyond the main conclusions of PISA Reports, which highlight social origin as the significant variable in the differences in educational outcomes.

Some of the most important characteristics for Finnish success are outlined and related to the Spanish case. There is no doubt that both the educational policy and population, economic and cultural aspects are very important. Organizational issues about school time and environment are decisive questions, as well as specific characteristics of one of the main educative agents: the teachers.

ANNEX II

In this sense, a strict teacher selection and exhaustive training ensure a commitment to education and gain parents' confidence in the educational system and teachers' faculties, giving them a high degree of professionalism and social prestige. The teaching profession is one of the most desirable of the professions. Teachers work independently and with autonomy, and they are highly competent in getting good results in terms of educational commitment. But in comparative terms, it is remarkable that this does not correspond to very high salaries.

In this work, the results obtained in PISA are linked to different aspects, such as school time spent in hours of teaching: lessons that students attend, net hours of teachers in the classroom and number of school days. We also establish relations between PISA results and class size, student/teacher ratio and teachers' salaries. In this last aspect, salaries have been taken into consideration according to level of experience, wage position in different countries, hourly rate, and time it takes to get the maximum wage.

The difference in school time is not a basic indicator of differences in academic performance. Neither do salary differences themselves constitute an element of prime importance to explain academic outcomes. Moreover, countries with the best educational outcomes are not always those with more school time and the highest teachers' salaries, like Finland.

Throughout this work, the main comparison is done between Finland and Spain, but there are references to other countries, especially those which have some outstanding aspects about their teachers. For example, Korea had better results in PISA Reports and is noted because teachers have the highest purchasing power. Luxembourg stands out with the highest teachers' wages at all educational levels. Great Britain and Netherlands also stand out because they have had trouble getting teachers.

Societies and educational systems are complex so it makes any comparison difficult. Nevertheless, there are some clues that can be enlightening on the importance of organizational and teaching questions, even to replace, to offset or to work with socio-cultural differences.

THE NEW CHALLENGES POSED BY PISA FOR A GRADUATE PROGRAMME IN UPPER SECONDARY EDUCATION TEACHING AT THE NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO

Dr. Juan F. Zorrilla
Programa de Maestría en Docencia para la Educación Media Superior
UNAM, México
E-mail: madems@posgrado.unam.mx

In the last fifteen years, Mexico has gone through major political and economic changes. These changes have modified the structural conditions that underwrite the functioning of the education system, especially for Upper Secondary Education (USE). In 2001 the OECD published the first results of the *Programme for International Student Assessment* (OECD: 2000). These results showed that 44% of the total Mexican student population surveyed found themselves in level 1 or under in

reading. Nonetheless, 96 percent of Mexican students reported exceeding minimum academic requirements in reading-related subjects. This contrast between school evaluation criteria and PISA's results continues to have a distinct effect on public opinion and in academic circles.

Once the gap between PISA and the criteria in effect in Mexican education had been established, it became obvious that the latter unduly favoured compliance rather than learning. The resulting lack of congruency put current Mexican educational and evaluation practices under a rather critical light. The ensuing consequences account now for some of the major priorities in educational policy for upper secondary schooling. At the same time, public opinion has shown a considerable interest, for the first time, in teaching, learning and evaluation topics.

In 2008 the Mexican Ministry of Education enacted the Comprehensive Reform of Upper Secondary Education (*Reforma Integral de la Educación Media Superior*). In Mexico USE offers two types of diplomas: one for general education and one comprising both general education and vocational education. This reform provides a Common Curriculum Framework for all of Upper Secondary Education. The Common Framework consists of the required competences that USE graduates should attain. These competences comprise both basic generic competences and specific competences for the four major curriculum disciplines: Math, Experimental Sciences, Social Sciences and Humanities and Spanish.

Discipline competences are statements that express specific discipline content standards, as well as skills and attitudes to be attained in each subject at the end of USE. These minimum competences are set in accordance with what is considered necessary for life long learning and application. Generic competences pertain to knowledge, skills and attitudes that underlie and permeate life in modern societies.

Given their basic nature, discipline competences can be developed in different ways, with different pedagogical methodologies, different partial standards for each subject and different teaching and learning didactic models. Disciplines are organized in four fields in the Common Framework: Mathematics, Experimental Sciences, Social Sciences and Humanities, and Mother Tongue and Communication. They express the capabilities considered necessary for all students, independent of whether students enter the work force after completion or decide to continue into higher education. These competences are expressed in an USE graduate's academic profile.

Along with these reforms of 2008, in the last five years, the National Institute of Educational Assessment has built an independent indicator intended for the identification of literacy type performance of pupils in language and mathematics. National sample tests have been carried out for Kindergarten pupils, for Third grade pupils, for Sixth grade pupils, and for Ninth grade pupils. Results are similar but also contain additional useful information for educational planning. One important aspect of this indicator is that it identifies the proportion of students with major deficits in knowledge and skills at each grade. This indicator was built on the basis of a wide consensus among a large group of academic consultants and faculty (INEE, 2009, *Panorama Educativo de México. Indicadores del Sistema Educativo Nacional. 2008*, pp. 228.)

ANNEX II

These changes offer immense challenges for a graduate teacher training programme like ours at the National Autonomous University of Mexico. Posters are intended to illustrate some of these challenges and responses.

ANNEX III

**PISA A EXAMEN, CAMBIAR EL CONOCIMIENTO,
CAMBIAR LAS PRUEBAS Y CAMBIAR LAS
ESCUELAS**

ESCUELA

PERIÓDICO PROFESIONAL ESCUELA ESPAÑOLA • FUNDADO EN 1941 www.periodicoescuela.es NÚMERO 3.848 • 10 DE DICIEMBRE DE 2009 • PRECIO 4 EUROS

JESÚS ROMERO, ANTONIO LUZÓN AND MÓNICA TORRES

PISA A EXAMEN, CAMBIAR EL CONOCIMIENTO, CAMBIAR LAS PRUEBAS Y CAMBIAR LAS ESCUELAS



Bajo este título, la CESE (*Comparative Education Society in Europe*) congregó durante cuatro días en la isla de la Palma a expertos nacionales e internacionales para debatir sobre PISA. Tras la bienvenida a cargo de la presidenta del Cabildo de la Palma, Guadalupe González Taño, y de Pilar Teresa y Germán González, de la Consejería de Educación del Gobierno de Canarias y de Rayas (Museo de Historia de la Educación de La Palma), que han patrocinado -junto con el Ministerio de Ciencia e Innovación- este Simposio, y tras una presentación de Miguel A. Pereyra, como presidente de la CESE y coordinador del evento, el programa PISA se contextualizó dentro de la nueva gobernación mundial de la educación. El Simposio se inauguró con sendas conferencias pronunciadas por Ulf P. Lundgren y Thomas S. Popkewitz. El primero, catedrático en la Universidad sueca de Uppsala ha sido experto del Consejo de Europa, la OCDE, UNESCO y el Banco Mundial,

ofreció al auditorio un análisis histórico de los discontinuos antecedentes de las “evaluaciones internacionales de los rendimientos escolares”, atento a la “temporalización” de tal idea, esto es, al modo en que ha ido modificándose su significado al compás de las cambiantes circunstancias socio-económicas, ideológicas y geopolíticas concurrentes. El segundo ponente, catedrático de la Universidad norteamericana de Wisconsin-Madison, se sumó al imprescindible esfuerzo de pensar históricamente el presente. En esa línea, “deconstruyó” la “receta PISA” para hacer aflorar algunos “ingredientes” poco visibles, aunque no por ello menos responsables del sabor resultante. Hablamos, por ejemplo, de algunas presuposiciones heredadas acerca de la posibilidad de redimir a la sociedad futura mediante la “formación” científicamente inspirada o de los usos políticos de la ciencia y la estadística para producir una “verdad” de apariencia objetiva y neutra. En el caso de las pruebas PISA, las reclamaciones de objetividad y neutralidad reposan en la mística de los números y en la pretensión de ser una evaluación de competencias libre de contenido, como si tal cosa fuese factible, cuando en realidad los tres tipos de “alfabetización” medidos (lectora, matemática y científica) no pueden entenderse al margen de esa peculiar “alquimia” que produce un “conocimiento escolar” entreverado de “tesis culturales” sobre el individuo, el ciudadano y la gobernanza colectiva. Estas dos conferencias inaugurales sirvieron de pórtico a unas densas jornadas organizadas en sesiones monográficas de mañana y tarde. Durante la primera, dedicada a ‘Los desafíos comparativos del Programa PISA de la OCDE’, Clara Morgan, profesora de la Universidad canadiense de Carleton, examinó la creciente centralidad de la educación en la agenda de la OCDE desde la década de 1990, e interpretó PISA como un reflejo de la progresiva “cientifización de la política”. La segunda sesión se ocupó de ‘PISA y el conocimiento escolar’, y estuvo alimentada por las ponencias de David C. Berliner (Universidad del Estado de Arizona) y David Scott (Instituto de Educación de la Universidad de Londres). En su celebrada intervención, Berliner quiso utilizar la experiencia de Estados Unidos como aviso para navegantes no vacunados contra el “fetichismo de los tests”, esa suerte de idolatría que acepta acríticamente los rendimientos computados como imagen válida de lo que ocurre y/o debería ocurrir en la educación. Sus propias investigaciones han desvelado varios efectos perversos: para aumentar las puntuaciones, muchos colegios han estrechado su currículo para incrementar la carga horaria de las materias examinables de lengua y matemáticas. Y como de esos resultados dependen en bastantes estados el salario de los profesores, la financiación de los centros e, incluso, su supervivencia, han brotado prácticas corruptas. David Scott desgranó las falsas creencias que subyacen en los *ranking* y resaltó su responsabilidad en el encajonamiento del currículo inglés, sobre el que acaba de alertar el *Alexander Report*.

La siguiente sesión versó sobre ‘Las evaluaciones PISA, la eficacia de las escuelas y la dimensión sociocultural’. Bajo ese paraguas, Katharina Maag Merk, de la Universidad de Zurich, expuso las conclusiones de un estudio empírico destinado a aquilatar la potencial mejora en los logros educativos derivada de la implementación de exámenes estandarizados en algunos Länder alemanes. Por su parte, Gerry Mac Ruairc, inspector de educación y, en la actualidad, profesor de la Universidad de Dublín, compartió sus indagaciones sobre el modo en que los

alumnos irlandeses se han enfrentado a las pruebas PISA. Del tema ‘PISA y la cuestión del alumnado Immigrante’ se ocuparon Petra Stanat, de la Universidad Libre de Berlín, y Julio Carabaña, de la Complutense de Madrid. Stanat afirmó que ya se sabía con anterioridad que el alumnado Immigrante sufría desventajas en las escuelas germanas, pero gracias a PISA ha sido posible visualizar el alcance del problema y confirmar que su causa remitiría fundamentalmente a las habilidades con la lengua alemana.



Carabaña coincidió asimismo en la riqueza empírica de PISA pues, entre otras virtudes, permite comparar los resultados de los estudiantes Immigrantes con los de sus connacionales en los países de origen. Partiendo del hecho de que esos resultados son relativamente similares, y dando por sentado que con sus tests libres de contenido, PISA cuantifica en realidad la inteligencia general de los examinados. Además, aventuró la controvertida hipótesis de que la raíz de tales similitudes no podía estar en las habituales explicaciones sociológicas, sino que habría que buscarla en las capacidades cognitivas.

Una ulterior sesión nos proporcionó algunas visiones procedentes de dos países que han vivido la experiencia PISA de un modo harto dispar: Finlandia y Alemania. Hannu Simola, de la Universidad de Helsinki, sometió a una aguda revisión el mito del “milagro” finés como corolario lógico de una política deliberada y cuidadosamente planificada, y abogó por manejar la idea de “contingencia” como instrumento teórico en las comparaciones. En su opinión, la elevada confianza en el sistema educativo, el fortísimo apoyo a la escuela comprensiva y la alta consideración social de la profesión docente, que estarían en la base de la preeminente posición de Finlandia en el *ranking*, no serían el fruto de ningún destino manifiesto, sino de la peculiar y accidental conjunción de varios procesos históricos e institucionales. Daniel Tröhler, de la Universidad de Luxemburgo, recordó que ningún país ha reaccionado de una forma tan furibunda como Alemania a su pobre escalafón en

J. ROMERO, A. LUZÓN AND M. TORRES

PISA, e hizo hincapié en la consiguiente confusión motivada por la colisión de dos ideales formativos distintos: la tradición alemana de la *Bildung* y el enfoque de las competencias promovido por la OCDE. Aunque para sus defensores las competencias no serían otra cosa que la movilización de los conocimientos y destrezas propios de la *Bildung*, Tröhler dudó de semejante matrimonio, habida cuenta que PISA supondría la cristalización de una discutible ideología, pretendidamente no ideológica, gestada inicialmente en el marco de la Guerra Fría y que ha conseguido hacerse dominante cuarenta años después.



A propósito de “La dimensión económica de PISA y el desafío de las competencias”, Javier Salinas y Daniel Santín, de la Universidad Complutense de Madrid, sopesaron los beneficios y las limitaciones de los datos aportados por PISA desde el punto de vista de los estudios económicos de la eficiencia escolar. Desde una perspectiva antropológica, los profesores de la Universidad de Roma *Tor Vergata*, Donatella Palomba y Anselmo Paolone, mostraron el contenido de una investigación en Italia, de alcance mucho más amplio, sobre la incidencia de PISA en la actitud de profesores e instituciones escolares hacia estudiantes que han permanecido largos períodos de estudio en el extranjero. En los estudios de casos analizados, se muestra la influencia del modelo PISA en las pruebas que realizan los estudiantes, como en las competencias adquiridas, constituyendo un modelo de referencia para la evaluación de las citadas competencias.

Posteriormente, el catedrático de la Universidad de Granada Antonio Bolívar acercó al auditorio a la influencia de PISA en el mundo de habla española. Habló de cómo el conocimiento orientado se transforma en un dispositivo de regulación que articula la manera de pensar, actuar e interactuar de los actores en cuanto a la elaboración de las políticas educativas. Así, PISA es una herramienta “multiuso”

con una gran relevancia simbólica, con una gran capacidad de adaptabilidad para convertirse en “filtro nacional”. En este contexto, y dada la influencia de los medios de comunicación tendente a dibujar una determinada imagen de PISA, el ponente mostró la influencia notable de PISA en el ámbito iberoamericano. Una presencia dispar, pero creciente y cada vez más relevante en la política educativa. Por último, y como clausura del Simposio, la conferencia de Robert Cowen, catedrático emérito del Instituto de Educación de Londres, fue sin lugar a dudas un broche de oro a unas jornadas de intenso debate, de las que el ponente se hizo eco en su alocución. Con especial sensibilidad académica, Cowen propuso un recorrido por los diferentes discursos que se habían sucedido a lo largo de estos cuatro días de Simposio para mostrar cómo desde los diferentes ámbitos del conocimiento, PISA está presente, emulando el mito de Cassandra. Alimentado su discurso con una rica y variada simbología metafórica, mostró con solvencia y claridad la deconstrucción de PISA en el marco de las metamorfosis de las instituciones y los procesos sociales, mostrando incluso cómo se ha convertido también en generador de nuevas formas de gobernanza transnacional, además de la movilidad de las ideas y los sistemas de transferencia. Cowen, de forma magistral y con grandes dosis de ironía, llegó a reivindicar la necesidad de teorizar, de realizar una *metateoría* o un *metanálisis* sobre PISA, sobre su origen, su nacimiento y su devenir, denominada por el autor como *big science* o gran ciencia.

NOTES ON CONTRIBUTORS

David C. Berliner is emeritus professor of Education at Arizona State University and on the summer faculty of Simon Fraser University in Vancouver, Canada. He is a member of the National Academy of Education, and a past president of both the American Educational Research Association (AERA) and the Division of Educational Psychology of the American Psychological Association (APA). He is co-author (with B. J. Biddle) of the best seller *The Manufactured Crisis* (Reading, Mass.: Addison-Wesley, 1995), of *Putting Research to Work* (with Ursula Casanova) (Arlington Heights, Ill.: IRI/Skylight Training and Pub., 1996.), and of the textbook *Educational Psychology* (with N. L. Gage), now in its 6th edition (Boston: Houghton Mifflin, 1998). His newest co-authored book, with Sharon Nichols, is *Collateral Damage: How High-Stakes Testing Corrupts American Education* (Cambridge, Mass.: Harvard Education Press, 2007).

Antonio Bolívar is full professor of Curriculum at the Faculty of Education in the University of Granada. His research focused on curriculum, innovation and educational change, organizational and professional development on which he has published over twenty books and more than one hundred articles. He is director of the electronic journal *Profesorado. Revista de Currículum y Formación del Profesorado* (Teachers. Journal of Curriculum and Teacher Education, <http://www.ugr.es/~recfpro/>), and a member of various editorial committees or advisory boards of journals. He has published, among others, *Cómo mejorar los centros educativos* (Madrid: Síntesis, 1999); *Los centros educativos como organizaciones que aprenden* (Madrid: La Muralla, 2000), and (with J.L. Rodríguez Diéguez), *Reformas y retórica: la reforma educativa de la LOGSE* (Málaga: Ediciones Aljibe, 2002).

Julio Carabaña is full professor of Sociology of Education at the Complutense University of Madrid. He was the coordinator of the Spanish version of the *Consciousness and Class Biography Survey* (Estudio Internacional sobre Estructura, Conciencia y Biografía de Clase, 1991). He is the author of major works in his field in Spain, such as *Educación, ocupación e ingresos en la España del siglo XX* (Madrid: Ministerio de Educación y Ciencia, 1983), “La inmigración y la escuela” (*Economistas*, 2004 99: 62-73), “Desigualdades desiguales: buenas y malas, crecientes y menguantes” (in J. Monreal, C. Díaz & J. J. Escribano, Eds., *Viejas Sociedades, Nueva Sociología*, Madrid: Centro de Investigaciones Sociológicas, 2005), he has been recently publishing on PISA: “Desigualtats explicables y desigualtats inexplicables a l'informe PISA” (*Quaderns d'avaluació*, Consell d'Avaluació-Generalitat de Catalunya, 2008, 7:35-61), and *Las desigualdades entre países y regiones en las pruebas PISA*, 2008 (www.colegiodeemeritos.es).

Robert Cowen is emeritus professor of Comparative Education at the Institute of Education of the University of London, and also a Senior Research Fellow in

INDEX

Oxford University. He was President of CESE from 2004-2008. His research interests include higher education systems, and especially the 'quality' control of universities. Among his most recent publications are: "Extreme Political Systems, Deductive Rationalities, and Comparative Education: Education as Politics" (in D. Halpin & P. Walsh, Eds., *Educational Commonplaces: Essays to Honour Denis Lawton*, London: Institute of Education University of London, 2005), "Acting Comparatively upon the Educational World: Puzzles and Possibilities" (*Oxford Review of Education*, 2006 32: 561-573), and "Comparing and Transferring: Visions, Politics and Universities", in *Higher Education and National Development: Universities and Societies in Transition*, D. Bridges et al., Eds. (London and New York: Routledge, 2007). He has edited with A. Kazamias the *International Handbook of Comparative Education*, 2 vols. (Dordrecht: Springer, 2009).

Marie Duru-Bellat is full professor of Sociology of Education at Sciences-Po (Institute of Political Sciences, Paris), and researcher in the OSC (Observatory of Sociological Change, Paris) and the IREDU-CNRS, in Dijon. She leads research upon gender and social inequalities in school, assessing their importance and their evolution. She analyses also some school processes, such as the impact of tracking or of the social and academic mix of a school or class. Last, she puts reforms (such as the openness of the educational system, or the process of decentralization) in an evaluative and comparative perspective. Among her publications are: "The Democratization of Education in France: Controversy over a Topical Issue" (*Population*, 2001 13: 189-218, with A. Kieffer); *Social Inequalities at School and Educational Policies* (UNESCO, 2004), *Education and Equity: International Perspectives on Theory and Policy* (3 vols., edited with R. Teese & S. Lamb) (Dordrecht: Springer, 2007), and "Recent Trends in Social Reproduction in France: Should the Political Promises of Education Be Revisited?" (*Journal of Education Policy*, 2008 23: 81-95).

Aileen Edele holds a MA in Psychology (Humboldt Universität Berlin), and is a research scientist in Educational Research in the Department of Education and Psychology at the Freie Universität Berlin. Her research interests are the effect of proficiency in the first language on the educational success of children and adolescents with migration background, and the impact of identity on educational careers of children and adolescents with migration background.

Hans-Georg Kotthoff is full professor of Comparative Education at the Pädagogische Hochschule (University of Education) in Freiburg, Germany. Prior to this, he taught in secondary schools in England and Germany. His research interests include comparative analyses of European education systems, European educational policy, educational governance, and school evaluation and teacher education. He is currently vice-president of CESE. Among his latest publications is a joint publication with M. A. Pereyra on the German PISA experience appeared in the electronic journal *Profesorado. Revista de currículum y formación del profesorado* (vol. 13-2, 2009: <http://www.ugr.es/~recfpro/?p=435>), and "Between Excellence and Equity: the Case of the German Education System", in *Revista*

Española de Educación Comparada (vol. 18, 2001, 27-60, retrievable at <http://e-spacio.uned.es/fez/view.php?id=bibliuned:reec-2011-18-5020>).

Ulf P. Lundgren is holding a personal chair in Education at the University of Uppsala in the field of Educational Policy and Educational Philosophy. From 1991-2000 he was the Director General of the National Agency for Education in Sweden, and from 2004-2008 Secretary General of the Swedish Research Council. An expert at the European Council, OECD, UNESCO, World Bank and different European universities, he has been producing educational theory for more than three decades within the field of curriculum theory, educational history, educational sociology, and evaluation. In 1977 produced his influential *Model Analysis of Pedagogical Processes* (Lund: CWR Gleerup), and *Between Educational Schooling: Outlines of a Diachronic Curriculum Theory* (Geelong: Deakin University Press, 1991) has been translated to different languages. At the present he is editor of the *Studies in Educational Policy and Educational Philosophy* (<http://www.upi.artisan.se/>).

Antonio Luzón, is associate professor of Comparative Education at the University of Granada, teaches international and comparative education with a focus on European educational systems. As member of the Research Group “Educational Policies and Reforms” of the regional government of Andalusia (Spain), from which he has participated in national and international projects of research, linked to the educational policies and reforms. Among his most recent works are: “Verso una democrazia creativa. Il significato di democrazia e educazione in John Dewey”, in Gramigna, A. (Ed.), *Democrazia dell’educazione. Saggi su Chomsky, Dewey, Foucault, Freire, Gadamer, Habermas, Levinas, Ricoeur, Savater, Maturana e Varela* (Milan: Unicopli, 2010); with M. Pereyra, J.C. González Faraco & M. Torres, “Social Change and Configurations of Rhetoric: Schooling and Social Exclusion-Inclusion in Educational Reform in Contemporary Spain”, in R. Cowen & A. Kazamias (Eds.), *International Handbook of Comparative Education* (Dordrecht: Springer, 2009); and with J.C. González Faraco, “Éducation et autonomie regionale en Espagne”, *Revue Internationale d’Éducation* (2010, 51: 41-48).

Katharina Maag Merki is full professor of “Theorie und Empirie schulischer Bildungsprozesse” (Theoretical and empirical analysis of learning in school) at the Institute of Education of the University of Zürich, Switzerland. Her research interests are related to school effectiveness and school improvement, educational governance, empirical educational research. She is Vice-president of the Swiss Society for Research in Education. Her most recent publications have appeared in “Unterrichtsgestaltung unter den Bedingungen zentraler Abiturprüfungen. Eine Analyse auf Schulebene mittels Latent Class Analysen” (*Zeitschrift für Pädagogik*, 2008 54: 791-809), “Introduction of Educational Standards in German-speaking countries” (in J. Hartig et al., Eds, *Assessment of competencies in educational contexts*, Göttingen: Hogrefe, 2008), *Lernort Gymnasium. Individuelle Entwicklungsverläufe und Schulerfahrungen* (Bern: Haupt, 2006), and *Kooperation*

INDEX

und Netzwerkbildung Strategien zur Qualitätsentwicklung in Einzelschulen (Seelze: Friedrich Verlag, 2009).

Gerry Mac Ruairc was a primary school teacher for many years. He was also a member of the Inspectorate of the Department of Education and Science of Ireland for a number of years. He moved in 2005 to University College Dublin to take up a position as lecturer in education with specific responsibility for developing the work of the school in the area of school leadership and management. Since his appointment to the staff in the School of Education in UCD he has initiated and been centrally involved in a number of research awards to the School of Education, primarily in the areas of educational leadership and educational disadvantage. He is deeply committed to attempting to provide a better insight into the reproduced patterns of inequality that persist in the education system, and in doing this he has been recently researching on PISA issues and the impact of this model of testing on students. His most recent publication on PISA is: “Dip, dip sky blue, who’s it? NOT YOU: Children’s experiences of standardised testing: A socio-cultural analysis” (*Irish Educational Studies*, 28: 47-66).

Clara Morgan holds a Ph.D. in Public Policy from Carleton University’s School of Public Policy and an MA in Education from the University of Toronto’s Ontario Institute for the Studies in Education. Her doctoral research examined the practice of international student assessments with a focus on the OECD and the IEA. She has published on the topic of post-secondary education and, more recently, presented conference papers on transnational educational governance and the history of international student assessments. She teaches at Carleton University’s Department of Political Science and works as an independent researcher (www.cmorganconsulting.com).

Donatella Palomba is full professor of General Education (*Pedagogia Generale*) in the Faculty of *Lettere e Filosofia*, Department of Philosophy, University of Rome “Tor Vergata”. She is currently President of SICESE – the Italian Society of Comparative Education. From 2000 to 2004 she has been President of CESE. For more than two decades, she has acted in international and comparative studies in education. Her main research interests are European Education, Citizenship and Intercultural Education, Higher Education policies, Distance Education. On many occasions she has acted as a scientific consultant for the Italian Ministry of Education in relation with European and international projects, and the OECD Country Report. Amongst her more recent publications, the edited volume *Changing Universities in Europe and the “Bologna Process”* (Rome: Aracné, 2008), and *Oltre l’e-learning? Università aperta e nuovi modelli di formazione* (Roma-Pisa: Fabrizio Serra Editore, 2008).

Anselmo Roberto Paolone attended the *École des Hautes Études en Sciences Sociales* in Paris where he defended a *mémoire* de D.E.A in Anthropology, and then entered the *European University Institute* in Florence, where he defended his Ph.D. on the birth of school ethnography in Britain. He has been *Visiting Global Fellow* at New York University in 1999-2000 and *Visiting Researcher* at the *London School of Economics* in 2001. He has been researching at the University of Udine,

for a research on qualitative methodologies in the study of schools where he holds a tenured research and lecturer position. He has also been lecturing in comparative education and school ethnography at the University of Rome “*Tor Vergata*”. One of his most recent publication is a book entitled: *Educazione Comparata e Etnografia, tra Globalizzazione e Postmodernità* (Roma: Monolite, 2009).

Miguel A. Pereyra is full professor of Comparative Education at the University of Granada, and President of CESE. Trained both as an educationist and an historian, concluded his tertiary studies on history & comparative education at Columbia University and its Teachers College, and his research and publications are focused on comparative & cultural history of education, and educational reforms and educational policies. His last publications in English are the edition of *Changing Knowledge and Education (Communities, Mobilities and New Policies in Global Societies* (Frankfurt: Peter Lang Verlag, 2008), “Social Change and Configurations of Rhetoric: Schooling and Social Exclusion-Inclusion in Educational Reform in Contemporary Spain” (with J.C. González Faraco, A. Luzón & M. Torres), in *International Handbook of Comparative Education*, edited by R. Cowen and A. Kazamias (Dordrecht: Springer, 2009), and the edition of a volume on the German debate about PISA (with H.-G. Kotthoff): *La experiencia del PISA en Alemania* appeared in the electronic journal *Profesorado. Revista de currículum y formación del profesorado* (vol. 13-2, 2009: <http://www.ugr.es/~recfpro/?p=435>).

Thomas S. Popkewitz is full professor of Curriculum and Instruction at the University of Wisconsin-Madison, and holds a Ph.D.h.c. in four European leading universities. His research is about politics of knowledge as its relates with the systems of reason that govern educational policy and research related to pedagogy and the issues of social inclusion and exclusion in educational research, focusing on research paradigms, the changing terrains of knowledge about teaching and teacher education, and reforms in school subjects. He has also conducted national and international studies of historical, ethnographic and comparative studies of national educational reforms and policies in Asia, Europe, Latin America, Southern Africa, and the U.S. Some of his main books are: *A Political Sociology of Educational Reform* (New York: Teachers College Press, 1991), *Struggling for the soul. The Politics of Schooling and the Construction of the Teacher* (New York: Teachers College Press, 1998), *Cosmopolitanism and the Age of School Reform. Science, Education, and Making Society for Making the Child* (New York: Routledge, 2008), and *The Child, the Citizen, and the Promised Land: Comparative Visions in the Development of Schooling in the Long 19th Century*, edited in collaboration with D. Tröhler & D.F. Labaree (New York: Routledge, 2011).

Risto Rinne is full professor of Education and director in the Department of Education at the University of Turku. He is also the director of the Centre for Research on Lifelong Learning and Education (CELE) in the University of Turku as well as the director of the Finnish Graduate School in Education and Learning (FiGSEL). His main interests and publications include sociology of education, international comparative education, educational policy, citizenship and learning in knowledge society and history of education. Among his many publications are

INDEX

“Changing the Tide of Education Policy in Finland: From Nordic to EU Educational Policy Model” in D. Hill (Ed.) *The Rich World and the Impoverishment of Education – Diminshing Democracy, Equity and Workers’ Rights* (London: Routledge, 2008) (together with J. Kivirauma & P. Seppänen), “Quality Assurance and Evaluation (QAE) in Finnish Compulsory Schooling: A National Model or Just Unintended Effects of Radical Decentralization” (*Journal of Education Policy*, 2009 4: 163-178) (together with H. Simola & J. Varjo), and “National Policy Brokering and the Construction of the European Education Space in England, Sweden, Finland and Scotland” (*Comparative Education*, 2009 45: 5-21 (together with S. Grek, M. Lawn & B. Lingard).

Javier Salinas obtained his M.Sc. in Economic and Social Policy Analysis at the University of York (UK) and his Ph.D. in Economics at the University of de Extremadura (Spain). He is professor of Public Economics at the School of Law, Complutense University of Madrid. He has been deputy Director of the *Instituto de Estudios Fiscales* (Spain) and his main research interests include: efficiency assessment in the public sector; public sector management; economics of education and regional economics. With J.M. Cordero y F. Chaparro, he has published their researches “Eficiencia en Educación Secundaria e inputs no controlables: sensibilidad de los resultados ante modelos alternativos” (*Hacienda Pública Española/Revista de Economía Pública*, 2006 173: 61-83) and “Measuring Efficiency in Education: An Analysis of Different Approaches for Incorporating Non-discretionary Inputs” (*Applied Economics*, 2008 40: 1323-1339), and, with Daniel Santín, *Análisis económico de los efectos de la inmigración en el sistema educativo español* (Madrid: Fundación Alternativas, 2009).

Daniel Santín holds a M.Sc. in Multivariate Data Analysis for Social Research and a Ph.D. in Economics from the Complutense University of Madrid (Spain). He is professor of Public Economics at the Faculty of Economics, the Complutense University of Madrid. In 2003, he was awarded by the Spanish Ministry of Education the Second Prize in Educational Research. His research interests include economics of education, the measurement of efficiency and productivity in the public sector, fiscal federalism, multivariate data analysis and operational research. Recently he has published “La medición de la eficiencia en las escuelas: una revisión crítica” (*Hacienda Pública Española/Revista de Economía Pública*, 177: 57-83), with Javier Salinas, *Análisis económico de los efectos de la inmigración en el sistema educativo español* (Madrid: Fundación Alternativas, 2009), and “Measuring Educational Efficiency at Student Level with Parametric Stochastic Distance Functions: An Application to Spanish PISA Results (with S. Perelman in *Education Economics*, 2011 19: 29-49).

David Scott has extensive experience of teaching in schools and universities. His present post is full professor of Curriculum, Pedagogy and Assessment at the Institute of Education, University of London. Previously he was Acting Dean of Teaching and Learning and Director of the International Institute for Educational Leadership and Learning, University of Lincoln. Among his most recent funded research projects as principal investigator are: *India Capacity Building to the*

Elementary Education Programme (in conjunction with Cambridge Education), and *Work-based Learning for Education Professionals Centre for Excellence*. Among his most recent publications are: *Education, Epistemology and Critical Realism* (London and New York: Routledge, 2008), *Critical Essays on Major Curriculum Theorists* (London and New York: RoutledgeFalmer, 2007), *Curriculum Studies - Volumes 1-4: Curriculum Knowledge; Curriculum Forms; Pedagogy; Subjects, Assessment and Evaluation* (London: Routledge, 2003), and *Assessment and the Curriculum* (New York: Greenwood Press, 2000).

Hannu Simola is full professor of Sociology and Politics of Education at the Department of Education of the University of Helsinki. His research interest has been moving from the socio-historical construction of schooling and teaching towards education policy and politics. Recently he is becoming fascinated with problems of education policy transfer, and how the trans-national meets the national and the local in policy and politics of education. Head of the Research Group focusing on New Policy, Politics and Governance of Education (KUPOLI), he has been a partner of different leading European research projects. His most recent publications include: “The Finnish Miracle of PISA: Historical and Sociological Remarks on Teaching and Teacher Education” (*Comparative Education*, 41: 455–470), “Researching the Political Effects of Quality Assurance and Evaluation (QAE) in education – Reflections on Some Comparative Issues in Sociology and the Politics of Education in the Audit Society” (with R. Rinne in M. A. Pereyra, Ed., *Changing Knowledge and Education. Communities, Mobilities and New Policies in Global Societies*. Frankfurt am Main: Peter Lang, 2008), and “Trans-national Technologies, National Techniques and Local Mechanisms in Finnish University Governance: A Journey Through the Layers” (*Nordic Educational Research* 2009 29: 6-17).

Petra Stanat holds a position as full professor for Educational Science at the Freie Universität Berlin. She worked as a research scientist at the Max Planck Institute for Human Development in Berlin where she managed the international part of the first cycle of the OECD’s Programme for International Student Assessment (PISA 2000) for Germany. One of her current research focuses is on conditions of immigrant students’ educational success. In cooperation with Gayle Christensen, she has written an OECD report on the situation of immigrant students in countries participating in PISA. She is a member of the German PISA 2009 Consortium, the Board of Directors of the Institute for School Quality Berlin-Brandenburg (ISQ), the German Council for Social and Economic Data (RatSWD), and the Advisory Board of the German Federal Office for Migration and Refugees (BaMF). She has published widely on questions of educational equity associated with students’ immigrant status and gender, on second-language teaching and learning, on school success in international comparison, and on PISA her last book (co-authored with Jürgen Baumer & R. Watermann) is *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit: Vertiefende Analysen im Rahmen von PISA 2000* (Wiesbaden: VS Verlag, 2006).

INDEX

Mónica Torres is an assistant professor in the Department of Pedagogy at University of Granada. She teaches international and comparative education, with a focus on European educational systems. As member of the Research Group “Educational Policies and Reforms”, of the regional government of Andalusia (Spain), has participated in national and international projects of research related to educational policies and reforms. Among her most recent works are: “From Centralism to Decentralization. The Recent Transformation of the Spanish Education System” (*European Education*, 2005, 37 1); with M. Pereyra, J.C. González Faraco & A. Luzón, “Social Change and Configurations of Rhetoric: Schooling and Social Exclusion-Inclusion in Educational Reform in Contemporary Spain”, in *International Handbook of Comparative Education*, edited by R. Cowen and A. Kazamias (Dordrecht: Springer, 2009); with Antonio Luzón, “La Educación Internacional, un nuevo escenario entre naciones desde una política global”, in *Educación Internacional*, (Valencia: Tirant Lo Blanch, 2009), and with J. Luengo & A. Luzón, “Las reformas educativas basadas en el enfoque por competencias: una visión comparada”, in *Profesorado. Revista de Currículum y Formación del Profesorado* (2008, vol. 12-3: <http://www.ugr.es/~recfpro/rev123ed.pdf>).

Daniel Tröhler is full professor of Education at the University of Luxembourg. Formerly he was Professor the Zurich University of Teacher Education and Director of the Pestalozzianum Research Institute for the History of Education. His research interests centre on the comparative analysis of the paradigmatic languages of education, on the history of educational ideas and institutions in the US, England, France, Germany, Luxembourg, and Switzerland. His recent publications include *Republikanismus und Pädagogik. Pestalozzi im historischen Kontext*” (Bad Heilbrunn: Klinkhard, 2006), *Democracy and the Intersection of Religion and Traditions. The Reading of John Dewey's Understanding of Democracy and Education* (edited together with R. Bruno-Jofré, J. Scott Johnston & G. Jover) (Montreal: McGill-Queen’s University Press, 2010), *Pragmatism and Modernities*. (edited together with T. Schlag and F. Osterwalder) (Rotterdam: Sense Publishers, 2010), *Languages of Education. Protestant Legacies, National Identities, and Global Aspirations* (New York: Routledge, 2011) and *The Child, the Citizen, and the Promised Land: Comparative Visions in the Development of Schooling in the Long 19th Century* (edited in collaboration with T.S. Popkewitz & D.F. Labaree) (New York: Routledge, 2011).