



Annika Wilmers
Sieglinde Jornitz (eds.)

International Perspectives on School Settings, Education Policy and Digital Strategies

A Transatlantic Discourse
in Education Research

Verlag Barbara Budrich



International Perspectives on School Settings, Education Policy and Digital Strategies

Annika Wilmers
Sieglinde Jornitz (eds.)

International Perspectives on School Settings, Education Policy and Digital Strategies

A Transatlantic Discourse in Education Research

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2021

© 2021 This work is licensed under the Creative Commons Attribution-ShareAlike 4.0. (CC-BY-SA 4.0)

It permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you share under the same license,

give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

To view a copy of this license, visit <https://creativecommons.org/licenses/by-sa/4.0/>

© 2021 Dieses Werk ist beim Verlag Barbara Budrich GmbH erschienen und steht unter der Creative Commons Lizenz Attribution-ShareAlike 4.0 International (CC BY-SA 4.0):

<https://creativecommons.org/licenses/by-sa/4.0/>

Diese Lizenz erlaubt die Verbreitung, Speicherung, Vervielfältigung und Bearbeitung bei Verwendung der gleichen CC-BY-SA 4.0-Lizenz und unter Angabe der UrheberInnen, Rechte, Änderungen und verwendeten Lizenz.



This book is available as a free download from www.barbara-budrich.net (<https://doi.org/10.3224/84742299>). A paperback version is available at a charge. The page numbers of the open access edition correspond with the paperback edition.

© 2021 by Verlag Barbara Budrich GmbH, Opladen, Berlin & Toronto
www.budrich.eu

ISBN 978-3-8474-2299-0 (Paperback)

eISBN 978-3-8474-1660-9

DOI 10.3224/84742299

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Ein Titeldatensatz für die Publikation ist bei der Deutschen Bibliothek erhältlich.

Verlag Barbara Budrich GmbH

Stauffenbergstr. 7. D-51379 Leverkusen Opladen, Germany

86 Delma Drive. Toronto, ON M8W 4P6 Canada

www.barbara-budrich.net

Jacket illustration by Bettina Lehfeldt, Kleinmachnow, Germany – www.lehfeldtgraphic.de

Picture credits: www.istock.com

Technical editing by Anja Borkam, Jena, Germany – kontakt@lektorat-borkam.de

Printed in Europe on acid-free paper by paper & tinta, Warsaw, Poland

Contents

Sieglinde Jornitz, Annika Wilmers

Transatlantic Encounters: Placing Education Research Interests in an International Context.....	9
---	---

I. School Leadership and School Development

Stefan Brauckmann-Sajkiewicz, Petros Pashiardis and Ellen Goldring

Comparing School Leadership Practices in Germany and the United States: Contexts, Constructs and Constraints	31
--	----

Esther Dominique Klein, Michelle D. Young and Susanne Böse

Successful Leadership in Schools Serving Disadvantaged Communities in Germany and the USA	47
---	----

Pierre Tulowitzki, Marcus Pietsch and James Spillane

Leadership for Learning in Germany and the US: Commonalities and Differences	62
--	----

Barbara Muslic, Jonathan Supovitz and Harm Kuper

Distributed Leadership in Schools: German and American Perspectives.....	80
--	----

II. Migration, Refugees, and Public Education

Lisa Damaschke-Deitrick and Alexander W. Wiseman

Migration, Refugees, and Education: Challenges and Opportunities	95
--	----

Johanna Fleckenstein, Débora B. Maehler, Steffen Pötzschke,

Howard Ramos and Paul Pritchard

Language as a Predictor and an Outcome of Acculturation: A Review of Research on Refugee Children and Youth	110
---	-----

Michael Filsecker and Hermann Josef Abs

Attitudes Towards Refugees. A Case Study on the Unfolding Approach to Scale Construction	120
--	-----

Ericka Galegher

Refugee Experiences in Higher Education: Female Perspectives from Egypt.....	137
--	-----

III. International Large-Scale Assessments and Education Policy

Nina Jude and Janna Teltemann

International Large-Scale Assessments – (How) Do They Influence Educational Policies and Practices?	155
---	-----

Kerstin Martens and Dennis Niemann

Lost in Translation? Local Governance and Policy Responses to International Large-Scale Assessments	171
---	-----

Lluís Parcerisa, Clara Fontdevila and Antoni Verger

Understanding the PISA Influence on National Education Policies: A Focus on Policy Transfer Mechanisms	182
--	-----

David C. Miller and Frank T. Fonseca

An Average Is Just an Average: What Do We Know About Countries' Low- and High-Performing Students in Mathematics?	195
---	-----

IV. The Management and Use of Digital Data in Education

Sieglinde Jornitz and Laura C. Engel

The Management and Use of Data in Education and Education Policy: Introductory Remarks.....	223
---	-----

Sigrid Hartong

Digital Education Governance and the Productive Relationalities of School Monitoring Infrastructures	242
--	-----

Steven Lewis

Data, Diagnosis and Prescription: Governing Schooling through the OECD's PISA for Schools	253
---	-----

Bernard Veldkamp, Kim Schildkamp, Merel Keijsers, Adrie Visscher and Ton de Jong

Big Data Analytics in Education: Big Challenges and Big Opportunities...	266
--	-----

Elmar Souvignier, Natalie Förster, Karin Hebbecke and Birgit Schütze

Using Digital Data to Support Teaching Practice – quop: An Effective Web-Based Approach to Monitor Student Learning Progress in Reading and Mathematics in Entire Classrooms	283
--	-----

V. Economization of Education

Marcelo Parreira do Amaral and Paul R. Fossum

Education Gone Global: Economization, Commodification, Privatization and Standardization.....	301
---	-----

Sabine Hornberg

Agents of Privatization: International Baccalaureate Schools as
Transnational Educational Spaces in National Education Systems 310

Alexandra Ioannidou and Annabel Jenner

Regulation in a Contested Space: Economization and Standardization in
Adult and Continuing Education..... 321

VI. Challenges of Translation in Educational Research

Norm Friesen

The Necessity of Translation in Education: Theory and Practice 337

Kathrin Berdelmann

When Dictionaries are not Enough: Translational Challenges of
Conceptual Historical German Terms in Educational Research 352

Inés Dussel

Translating Research: Tensions and Challenges of Moving Between and
Through Research Practices 360

Britta Upsing and Musab Hayatli

The Challenges of Test Translation 373

Index 389

Transatlantic Encounters: Placing Education Research Interests in an International Context

Sieglinde Jornitz¹ and Annika Wilmers²

1. Introduction

In recent decades, education science has increasingly become networked internationally. In Germany for example, prior to the year 2000, the discipline was rather focused on national discourse whereas an interest in educational policy or pedagogical matters across other countries was only shown in individual cases. A new impetus came from international student assessments run by the International Association for the Evaluation of Educational Achievement (IEA) and the OECD. This trend was supported by manifold funding research programs which did not only target European and international conference activities but specifically attempted to foster research co-operations among scientists from the discipline (Berg et al. 2004; Jornitz/Wilmers 2018).

From a German perspective, the term “international” often implies collaborations with scientists based in the USA. At least two reasons can be assigned with respect to this particular interest. On the one hand, the English language has made it fairly easy to follow up on the discourse in the US while on the other hand, the US have been and still are leading in the development of all types of student achievement tests and assessment procedures (Jornitz 2018; Aljets 2014). The (recurrent) growth of assessment studies in Germany made it necessary to co-operate and pressure from the science community in the USA complementarily also evoked a desire to learn more about education science in Europe, including Germany, and many other countries throughout the world. The increased participation of German scientists in the annual meeting of the American Educational Research Association (AERA) reflects this development, to which we have given shape by conceptualizing and launching a series of international sessions in this context. The format has not only proven successful but it has also led to diverse research co-operations on both sides of the Atlantic. The annual event has moreover facilitated stability in the initiation of contacts, which many of the participants were pleased to

1 Sieglinde Jornitz is Senior Researcher at the DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main. Email: jornitz@dipf.de

2 Annika Wilmers is Senior Researcher at the DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main. Email: wilmers@dipf.de

take on. The thematic diversity and depth of the international discourse over the years are reflected in this volume, which illustrates that the focus has never been on a mere comparison of developments in Germany and the USA. Rather, such developments are comprehended as being located in a diverse international context to which colleagues from other countries and other discourses have contributed.

In this introductory chapter, we will outline some of the central characteristics of the school systems in the US and Germany. This will be followed by an exploration of some of the discourses on school reforms that both countries participated in over the past 150 years. A third section of this chapter examines the development and concepts of comparative and international education research in Germany and the US before the last section introduces this volume and gives an overview on the international activities it is based upon.

2. Historical pathways of the German and American school systems

The school systems of Germany and the US have often provided a starting point for many thematically diverse networks in education research. In both countries the school systems are federal, but they show some significant differences in structures and organization due to their different historical and political developments. In Germany and the USA the national government and the Ministry of Education have no legally binding access to the education system as a whole. In both countries, the states or *Länder* are politically and thus also legally in charge of the school education system. Whilst the US are constituted of 50 partially federal (autonomous) states, the Federal Republic of Germany consists of 16 federal states. Differences can be found with regard to the stronger local influence on schools in the US states on the one hand and some efforts of national coordination through the implementation of the Standing Conference of the Ministers of Education and Cultural Affairs (KMK, founded in 1948) in Germany on the other hand. Two structural aspects are important for Germany. First, as one of the German particularities, students leave a comprehensive primary school after four years, generally at the age of ten. Depending on their achievement profile, they are then allocated to a secondary school in a three-track (*Hauptschule* (5 years), *Realschule* (6 years) or *Gymnasium* (8-9 years)) or, more recently, a two-track (*Realschule* or *Gymnasium*) system. Hans Döbert, a German expert on school systems, points out: “During the course of the nineteenth century, a three-track school system came into existence, whose role was essentially to

cater to and stabilize the social interests of the three-class society of Germany.” (Döbert 2015: 306). The leading education minister of the state of Prussia in the 19th century, Wilhelm von Humboldt, developed a three-tier school system intended to reflect a segregation of society into three parts. Humboldt believed that the respective school should equip students with a type of general education that qualified them for working in skilled labor, administrative and academic professions.

This secondary school system persisted after World War II and was forcefully defended in the 1970s in the former Federal Republic, when proponents of a comprehensive school system were accused of wanting to introduce a uniform, socialist or communist school system, similar to the one existing in the former GDR or the Soviet Union. “Thus, from 1949, the education system of West Germany and its federal structure were diametrically opposed to the centralized structure of East Germany.” (Döbert 2015: 308). After the fall of the Berlin Wall in 1989 the West German school system was implemented in the newly founded East German federal states. In this regard, the *Gymnasium* does not only stand for the opportunity to obtain an academic qualification but it also symbolizes an opposition to a comprehensive school system. The achievement-based allocation of students to three (or two) school types is thus meant to create homogenous learner groups.

Secondly, the German school system is centered around a commitment to science disciplines that are represented by school subjects and adapted according to student age. Topics and school subjects defined by the curriculum largely correspond to science disciplines. In the case of Germany, “a remarkable consistency in subjects” (Döbert 2015: 323) over the centuries can be observed. Arguing from the school perspective and the demands society links to school, Dietmar Waterkamp, a German scholar and expert in comparative education, characterized the German school as a “hasty school” (“*eilige Schule*”) (Waterkamp 2012: 97-109). Hence, a large number of subjects are taught at schools in Germany. Exercises and revision units are usually assigned as homework and thus relocated to extracurricular afternoon sessions. At the same time, students are held responsible for ensuring that they have understood the subject. Waterkamp asserts that “public classroom discourse” (Waterkamp 2012: 98) is characteristic for the way in which teachers design their lessons. Based on an interrogative dialogue between teacher and class, an individual student’s contribution to a topic is assumed to be relevant for all the others.

Following Germany’s participation in international large-scale assessment studies like TIMSS and PISA, a paradigm shift has taken place. Whereas state control formerly focused on the curriculum and followed a so-called input-oriented model of state control and monitoring from 2000 onwards, the model has shifted towards an output-oriented one (see Döbert

2015: 315). To measure learning outcomes, national achievement tests were implemented. This instrument, including its specific items and scaling practices, was as new for German students and their parents as for teachers. This shift in education policy also brought Germany's education research into closer alignment with the international discourse and development of evaluation instruments. "Today, comprehensive educational monitoring which now embraces standardized tests and comparative work, national and international studies of school achievement, and educational reports, is part of the fixed repertoire of control functions in education." (Döbert 2015: 315).

Schools in the United States are rooted in a different tradition. They are characterized by the idea of one school for all. All students are taught in the same type of school, which differentiates by age and courses. There is no early tracking via school types and students are grouped in courses regarding interest and learning level. Testing is a typical instrument in American schools. These data are used for steering educational practice and policy (see section 4 of this book). Both characteristics – course tracking within one school type and students' testing – are rooted in the history of the American school system, which Paul Fossum divides into four educational historical periods or "movements" (Fossum 2021, forthcoming; see also Rury 2014). The first period took place in the mid-1800s and was centered on the question of a common school. Its leading figure was Horace Mann (1796-1859) who fought for the establishment of a public school system and broadened the availability of education in the US.

This was followed by the progressive education movement that lasted from the late 19th century until the mid-1900s. John Dewey was its well-known supporter and protagonist. Progressive education puts the learner and his or her needs at the forefront of pedagogical thinking and practice. For the US, in contrast to Europe, it was also the time "intensive testing of students [began] as a means of gauging their intelligence and of enabling their sorting and channeling into instructional emphases" (Fossum 2021, forthcoming). Concerning the progressive schools in the 1920s, Ellen Lagemann states that these schools "were increasingly giving up traditional subject-focused curricula in favor of problem- or project-focused activities." (Lagemann 2000: 100). With an ongoing school enrollment, students' testing and the establishment of a course system in school became widespread. The idea of a "uniform academic core" (Lagemann 200: 101) for the school curriculum was more or less turned down until nearly 100 years later, when it emerged again vehemently with the controversy on the Common Core Standard in 2010.

A school day in the United States largely follows a course structure. Subjects are thus less aligned to a science discipline structure and students have more freedom to choose their courses according to their aptitudes and interests. In his comparative study, Waterkamp describes the US school as a "school of alteration or variety" ("*Schule der Abwechslung*") (Waterkamp

2012: 139-153). Courses, instead of subjects, are taught and these courses span a broad range of topics. This can be explained by large immigration movements in the 19th century which brought people from many different countries to the US. Hence, the US school system had to serve people from diverse cultures, languages and biographies. Joel Spring writes in his classic work on the American school: “The idea of using education to solve social problems and build a political community became an essential concept in the common school movement.” (Spring 2018: 91). Therefore, establishing a nation-wide school system is closely linked to the concept of becoming an American citizen and forming a new nation (Rury 2014).

According to Fossum, the third educational period spanning the 1960s and 1970s concentrated on fighting against the ongoing segregation in schools, and expanded its focus on anti-discrimination activities from race to gender, ethnicity and religious belief (Fossum 2021, forthcoming). It was a time when the education system was challenged with integrating every child into its system and offering him or her the best education available.

When in 1983 the controversially discussed report “A Nation at Risk” was published (see: Fossum 2021, forthcoming; Spring 2018: 478ff.), with the main result that schools were not able to reach their goals, it led to an *Accountability Movement* that is still in place today. This fourth period (Fossum 2021, forthcoming) started two important reform activities, one on standardization of curriculum and one on school choice. Both are topics of an ongoing debate (Ravitch 2010; Schneider 2016). Assessment and the expansion of different test structures are central elements of American schools, while in Germany and Europe, this instrument of measuring student achievement is rarely used, or implemented only on special occasions. Nevertheless, criticism of achievement studies has been growing in the US. In 2002, the No Child Left Behind Act was introduced (passed in 2001; signed in 2002) sparking a development that Urban, Wagoner and Gaither describe as a process of “reinforcing a steady diet of high-stakes standardized testing” (Urban/Wagoner/Gaither 2019: 344).

A comparison of the two school systems points to both similar and different traditions and thematic priorities. However, the set-up of the two public education systems was accompanied by an ongoing transatlantic exchange on education reforms and policies.

3. School reform in a transatlantic exchange

Over time, similar topics were addressed in both the United States and Germany, as can be seen from the discourse on particular educational reforms, the set-up and expansion of education systems or the debates on

quality in education. Still, this does not imply that discourses have taken place at the same time nor that debates are grounded in the same conceptions across countries. But the similar foci of interest are striking in both the US and Germany, and so are returning references to the respective other country in attempts at education system reform over the past centuries. For example, in many cases Germany served as a role model for the American education system in the early stages of its development. A lively intellectual exchange on educationally relevant topics can be found throughout the 19th century, and following the Second World War, American re-educating activities took place in West Germany. From a historical perspective, two episodes stand out in the continuing transatlantic educational discourse: First, the interest in education systems in German states, particularly universities, during the establishment of a higher education system in the US in the 18th and 19th centuries and, second, activities linked to the goal of (re)democratization of the German education system and the so-called re-education measures after 1945 (for information on the history and development of transatlantic exchange in education, cf. Overhoff/Overbeck 2017; Uljens/Ylimaki 2017).

“Re-education” was not merely an isolated objective after 1945, as Thomas Koinzer demonstrates in his work on experiences and appraisals of German pedagogues who travelled to America as part of a German “Educators’ Mission” between 1960 and 1971 (Koinzer 2011). Following the re-occurrence of anti-Semitic incidents in Germany, the American Jewish Committee and the study office for political education at the Institute for Social Research in Frankfurt am Main (*Institut für Sozialforschung*) had organized the program to enable German pedagogues to experience the American education and school system, which was perceived as taking a leading role on the path to a democratic school model. The participants’ experiences and observations focused on concepts of teaching and realizing democracy at school as well as concepts of implementing and running empirically-oriented research in the social sciences (Koinzer 2011). The group of *Amerikafahrer* (America-goers) was heterogeneous and came from all over West-Germany. It was comprised of German pedagogues from the areas of practice, policy-making and research who were particularly interested in practice-related, applied pedagogy. In the assessment of the American system, the German educators painted a diverse and ambivalent picture, fed by claims for a democratic school on the one hand, and perceived political and social problems on the other, e.g. race segregation and violence in American society or foreign political developments, such as the Vietnam War in the 1960s and 1970s. Nevertheless, it affected the education reform measures in West Germany in different ways (Koinzer 2011: 12-13).

Ewald Terhart has identified an Anglo-American influence on German educational reform discourse in particular for the period spanning 1965 to 1975, concerning educational science concepts and methods (Terhart 2017).

According to Terhart, at this time the educational discourse in Germany became more susceptible to influences of empirical educational science, psychological research on learning and teaching, programmed instruction and curriculum research. These were meant to help overcome a standstill in the reform process in Germany as well as to foster a new orientation within education science. These reform efforts came to an end in the late 1970s and in the 1980s, when new economic crises (see e.g. the high unemployment rate among teachers) and (inner-)political crises (e.g. the Red Army Faction activities) arose and other developments, such as the rise of new social movements, evoked a shift of educational political interests (Terhart 2017: 166-170).

In this context, Terhart refers to the relationship between taking up and adapting American concepts in German studies and relevant translations of important American educational works by German scientists, and the dissemination of American theories in Germany. At the time, the translation efforts were essential to studying Anglo-American methods to this extent in Germany (Terhart 2017: 164).³ The need to translate English language publications into German has rapidly declined since the 1990s, because since then knowledge of English has increasingly become a standard in German and international education science. However, this transfer is by no means a completed task, which becomes clear when looking conversely at ways to discuss German research internationally and at continuing challenges in the field of translating non-English studies from humanities research, as will be discussed later in this volume (see section 6 of this book).

For endeavors at familiarizing an American readership with German research, it is interesting to take a look at the German pedagogue Erich Hylla (1887-1976), who had been able to do research in the US in 1926/27 and who had been a visiting professor at Columbia University and Cornell University in the second half of the 1930s. After World War II, he served as advisor in education questions to the US High Commissioner in Germany and was involved in the German-American plans for a new research institute for international pedagogical research in Germany, which eventually led to the founding of the DIPF – today the “Leibniz Institute for Research and Information in Education” – in 1951. In his book, “Education in Germany. An Introduction for Foreigners”, published in 1954, Hylla explains the German education system to an English-speaking readership.⁴ An earlier volume had already been published in 1928, called “Die Schule der Demokratie. Ein

3 A list of exemplary translations from the reform age in the 1960s and 1970s can be found in Terhart 2017.

4 In 1929 Hylla translated Dewey’s “Democracy and Education” and this work was reedited in 1949 and in 1964 followed by a new edition from Oelkers in 1993 (Hylla 1949; Oelkers 1993). Regarding the reception of Dewey in Germany after the turn of the millennium see Bellmann 2017.

Aufriss des Bildungswesens der Vereinigten Staaten” (“School of Democracy. An outline of the education system of the United States”, Hylla 1928), wherein Hylla exhaustively described the American education system to German readers. Both books aim to inform the respective counterpart with the underlying assumption that new foreign phenomena can only be understood within the context of the system one is familiar with, as Hylla points out in his preface of “Education in Germany”: „Since any given educational system can be really understood only as a part of the cultural and socioeconomic texture in which it has developed, an attempt was made to indicate this frame of reference in the extended explanatory passages [...] accompanying the discussion of various aspects of German education. Thus the foreign reader should be enabled to find the common denominator for corresponding phenomena of education in his own country and in Germany.” (Hylla 1954: 3)

The globalization of educationally relevant topics and a growing interest in international comparisons, which is evident from large-scale international assessments, prominently placed international exchange on educational topics on the agenda in the past three decades (see section 3 of this book). The idea that, in a globalized society, education is a determinant factor, also given global competition, is not new, as the “Sputnik Shock” after 1957 and the American debate following the “A Nation at Risk” Report in 1983 showed. The Sputnik shockwaves did extend to West Germany, yet it was the later “PISA shock” in 2000 that alerted the German population profoundly and persistently with regard to education, whilst comparatively little attention was paid to the results of the first PISA study in the US (Martens 2010). Attention only rose when China ranked higher than the US in the PISA cycle of 2009 (see Parcerisa, Fontdevila and Verger in this volume). The examples illustrate the wide scope when positioning educational topics on a country’s agenda, ranging from national education aspirations to international (education) competition. Educational topics are simultaneously placed on a transnational agenda as well as developing highly national and even regional trajectories and dynamics. In this regard, the issue of international transferability and its relation to country-specific education concepts are debated under the slogan of “educational borrowing and lending” on both sides of the Atlantic. These refer to a complex construct of international settings and national adaptations (cf. Steiner-Khamsi/Waldow 2012; Phillips/Ochs 2010).

4. Comparative and international education research – pathways and concepts in Germany and the USA

In recent decades international exchange in education research has been centered around a comparative perspective. But with growing international cooperation, comparative research has nearly lost its former reference point in the battle of political systems. Prior to the fall of the Berlin Wall and the collapse of the Soviet Union, capitalist and socialist societies and education systems challenged each other, in a tug of war for better performance. Now, the systems seem to be competing globally against each other for the best performance as an education system, understood as an expression of economic power. But the different histories of comparative education research in Germany and the USA are still virulent in international co-operation. They are worth shedding light on. A highly data-based tradition of comparative research can be found in the Anglo-American context as opposed to the rather philosophic, hermeneutical access common in Europe.⁵ Both are briefly outlined in the following paragraphs.

The establishment of comparative education in Germany has often been linked to the French Revolution and a respective rise of science disciplines with Marc-Antoine Jullien de Paris (1775-1848) as its founder (see Allemann-Ghionda 2004; Waterkamp 2006). In his text “*Esquisse et vues préliminaires d’un ouvrage sur l’éducation comparée*”, published in 1817, he suggested collecting data on different education systems in a standardized manner. This marks a beginning in placing the knowledge of education systems in analogy to the natural sciences. The aim was to collect data to gain scientific – i.e. positivist – insights into education systems from different countries. These data would build up an extended knowledge base for one’s own pragmatic actions. Accordingly, not only data but also country reports were fundamental to such studies.

Moreover, the French Revolution was in line with an idea to conceive science and also educational science in terms of finding relevant valid natural laws. For a comparison of education systems, this would in consequence have meant that one valid form of education system would fit societies anywhere in the world. Ideally, this system’s structure ought to enable a student to optimally acquire skills and knowledge. Students would thus be inspired to develop autonomous, free minds. Such an intended system would gain validity from reason. And because reason is perceived as culturally indifferent, such an educational science would lead to a valid education system that might be set in place worldwide (cf. Koneffke 1988/2018).

5 As an example of narrowing the German and American discourse on comparative education, see: Suter, Larry E./Smith, Emma/Denman, Brian D. (eds.) (2019): *The SAGE Handbook of Comparative Studies in Education*. London et al.: SAGE.

However, this position did not win ground. Instead, differences in education systems were in many cases understood as being idiosyncrasies – which today might be conceived in terms of cultural specificities. In comparative education research, the discourse on national character became dominant and was evident in specific attitudes towards individuals and societies at large, or in educational objectives and ideals (Waterkamp 2006: 20). At this point, it should be noted that the term “nation” bears different connotations in German and English. In German, the idea of a nation is traditionally tied to a mother tongue. By contrast, the English language refers to belonging to a state, adhering to its citizens as a whole – in a more abstract sense (see Waterkamp 2006: 28-31).

For Germany, comparative education science after 1945 is largely determined by area studies, i.e. systematic descriptions of education systems. These descriptions served as a basis for comparison (see Allemann-Ghionda 2006: 25; 29-30). Until the late 1980s, the countries from the so-called Eastern Block were at the center of interest, not least because of the two divided German states. Moreover, countries in Africa and Asia were studied, which had just set out to become democratized and industrialized. This type of comparative research was always highly linked to a philosophical-hermeneutic tradition of education science in Germany.

The development of comparative educational science has taken a different path in the USA, where James E. Russell prepared the ground in 1900. Michael E. Sadler (1861-1943) built on this foundation, which gained further shape by the work of Isaac Leon Kandel (1881-1965). In 1933, Kandel designed his “Studies in Comparative Education”, where, instead of describing individual systems, a comparison of both was actually conducted. The comparison was based on a socio-historical approach and an education system was perceived to be an impression of a given national character. However, such national character was not taken for granted but it could be deduced from a historically grown, socio-economic and political state structure.

This comparison was determined by the principal orientation of education science in the US which had been understood as an empirical science with a clear reference to psychology and its quantitative measurement methods between 1890 and 1920 (Lagemann 2000: 16; 23). Accordingly, there is a principal understanding that comparative education science should be data-based and that such data should be collected for education systems in other countries, too. Since the 1930s, comparative education shifted from a traditionally descriptive science to a discipline that works with sociological and mostly quantitative methods.

From the 1960s onwards, the US increasingly began working with the United Nations. Comparative education science at the time centered on the question of how education systems help nurture and strengthen democratic

structures. Owing to this thematic orientation, countries in Asia and Africa were the focus of their research – many of which were in the process of regaining independence after decolonization and building new governance structures. Regarding education systems, the UNESCO emerged as a central organization to support these developments. Comparative education scientists from the USA were required to apply their insights in the respective countries and to offer counselling. In contrast, Germany took increasing interest in these countries later on in the 1980s, and worked together with UNESCO (see above).

In parallel, measurement instruments were developed for the comparison of education systems. The Association for the Evaluation of Educational Achievement (IEA) was consequently founded in 1958. Torsten Husén, Neville Postlethwaite and Richard Wolfe were among the initiators. Even today, their contributions to an educational psychometry remain at the core of comparative studies and as such of international assessments of student achievement and diverse outcomes of education systems. This groundbreaking work was used and effectively presented to the public by the OECD and its PISA studies – which could not have been successfully designed if the IEA had not prepared the groundwork in a methodological sense.⁶ Looking back at comparative education in the US, Martin Carnoy states that “international testing [...] is by its very nature internationally comparative, it has become the dominant force in shaping comparative education research” (Carnoy 2019: 197).

It looks like comparative education lost its political dimension with regard to opponent society models in East and West; such a “classical” comparison became obsolete with the collapse of the countries belonging to the Eastern Block. Developments within the IEA and the OECD in the 1990s and 2000s in the field of comparative analyses of education systems could easily fill the gap. Especially for Germany it is true that leading researchers in this field were not rooted in comparative studies, but in quantitative psychology and psychometry – disciplines that had always oriented their methods toward the Anglo-American discourse.

Ultimately, a situation emerged that reshaped the landscape of education research to date. In the 2000s many traditional comparative research chairs were no longer upheld by universities in Germany. Instead, a new area of comparative research of education systems has internationalized education science as a whole, and the discipline became largely oriented toward

6 In his historical account of comparative education at Stanford University, Martin Carnoy (Carnoy 2019: 16-21) shows that these test methods have also changed the orientation of comparative education as is directed toward co-operation with developing countries. Without an opportunity to systematically collect data on education systems or test-based assessments, UNESCO would probably not have launched the Education for All initiative from 2000 onwards.

quantitative approaches (Waldow 2015). In the course of this development, many scholars in education showed an interest in comparative methods and linked themselves to an international respectively transatlantic exchange.

This development led to two epiphenomena. First, for Germany, a nuance of the discipline became less visible. In research qualitative-hermeneutic analyses of pedagogical practices are widespread and a characteristic approach for education science. This approach is deeply rooted in the history of education research in Germany, and led to extensive work on analyses of educational practice with hermeneutical methods. It is this area that remains to be discovered by comparative research. Second, Martin Carnoy raises another aspect that is nearly lost in the research discourse on other education systems. He critically underlines that despite significant progress in methods made in comparative education, researchers are increasingly losing interest in theory-building. Making comparisons of educational systems “was never expected or intended to substitute for deeper analysis of differences among educational delivery systems and explanations for how and why differences exist.” (Carnoy 2019: 197). An answer to such questions might only be found by a theory that is substantiated with data.

Both aspects are worth keeping in mind as a stimulus to carry on with an international exchange of research methods, results and theories. For international as for comparative education science, the respective differences and commonalities offer manifold incentives, which were thematically taken on at the international sessions (see below) and this volume aims to provide some impressions.

5. Introduction to this volume and its different sections

Contributions in this book stem from a series of international seminars which were organized by the office “International Cooperation in Education” (ice) located at the DIPF | Leibniz Institute for Research and Information in Education in Frankfurt and took place as affiliated group meetings at the Annual Meetings of the American Educational Research Association (AERA). The staff at ice started these international activities in 2013 in San Francisco. For two years we organized poster presentations about research projects and topics that were of interest on both sides of the Atlantic, research infrastructures and discussion rounds at the DIPF booth and a panel discussion on the implementation of and national debates about education standards in mathematics in Germany and the US. From 2015 on, these international events were organized as seminars with panels and roundtables providing for an intensive and lively exchange on research projects and common research interests. While participants primarily came from Germany

or German-speaking countries, the US and Canada, this setting gradually moved from a German-American exchange to a broader perspective including researchers from other American and European countries as well as from other continents, such as Australia or Asia. By adding this additional international perspective to the discussions, the intention was not to provide more single country studies, but to add a wider variety of perceptions of internationally relevant common issues, such as large-scale assessments or digitization in education to the discussion.

Contributions in this book represent a selection of topics that were discussed and further developed between 2016 and 2019 at the AERA annual meetings in Washington D.C., San Antonio, New York and Toronto. Our international sessions during these years were oriented around the annual meeting theme in each year. In 2016 participants discussed “International Perspectives on School Governance” at a panel discussion on “Data-driven School Improvement – the Role of Data for Teaching and Learning” as well as at three roundtables with presentations dealing with monitoring and school leadership, computer-assisted progress monitoring and the potentials and boundaries of digitization in education research. The exchange was supplemented by a poster session introducing several American and German research programs, centers and initiatives.⁷ The 2017 international session shed light on “Societal Challenges and Educational Research” with a panel on current challenges in education, such as the influence of neoliberal politics on education, the tasks related to the integration of school children with a migrant background into the educational systems and the role of multilingualism in this context. Six roundtables took up these topics from different perspectives analyzing questions related to instructional school leadership, migrants and refugees in educational systems, the use of data from large-scale assessment in educational policy as well as digital education policies and practices. In addition, one group discussed methodological questions in a workshop setting. In 2018, our international sessions focused on “International Perspectives on Public School Systems”, starting with a panel on “Raising Standards and Educating for Democracy: Contradiction or Interdependency in Public Education?” The panel was followed by six roundtables on school leadership and public school development; migration, refugees and public education; international perspectives on data-driven education; the economization of education and trends towards a global education industry as well as methodological questions around the challenges of translation in transnational education research.

7 Among the presenters were the National Educational Panel Study in Germany (NEPS at IIfBi), the US National Center for Education Statistics and the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), the German Center for International Student Assessment (ZIB) and the Leibniz Education Research Network (LERN) as well as the American College Board.

The question of how public education is understood by different actors in the field including students' perspectives was also taken up at a second panel discussion jointly organized by the office ice, the University Alliance Ruhr and the German Center for Research and Innovation. This event did not only address education researchers, but explicitly invited interested New Yorker citizens – teachers, journalists, publishers or parents of school children, among others, to discuss crossroads in public education at the beginning of the 21st century. The German Center for Research and Innovation had also already kindly sponsored our international events at the annual meetings of the AERA in 2014 and 2016. The series of international sessions was continued in 2019 around the annual meeting theme of leveraging education research in a “post-truth” era and the meaning of democratizing evidence. In 2020 another session was going to deal with the topic of education in a digital world but could not take place due to the Covid-19 pandemic.

Discussions at the previous sessions were open to researchers at all stages of their academic career and from different organizational backgrounds. The sessions included both research perspectives that already involved a comparative analysis and projects that highlighted single country examinations, but were then placed in an international context during the discussion at the roundtables. The selection of topics in this book represents research questions that were discussed continuously and further developed over several years. These topics were specifically relevant within the German and US context, but also with regard to a broader international research perspective. Reflecting the setting of the roundtable discussions, this volume therefore also includes additional country perspectives from participants of the international sessions.

The first section on school leadership (section editors: Stefan Brauckmann-Sajkiewicz, Petros Pashiardis and Ellen Goldring) explores different facets of school leadership practices in Germany and the US while taking into account the different school and policy contexts as well as differences in governance structures and school management traditions, for instance by contrasting the American picture of school governance with the German model that used to place more emphasis on the teaching than the managing process. By so doing, the authors also point to the different research traditions in the two countries and to the changing roles school leaders are identified with. The section is structured by the themes “leadership in challenging environments” and the question of how school leaders can contribute to the success of schools serving disadvantaged communities (Esther Dominique Klein, Michelle Young, Susanne Böse), “leadership for learning” (Pierre Tulowitzki, Markus Pietsch, James Spillane) as a comprehensive theoretical model and “distributed leadership” (Barbara Muslic, Jonathan Supovitz, Harm Kuper) as a concept that stands for a democratic and cooperative leadership style. While exploring different

settings and expectations, the section also suggests common frameworks for future international research on school leadership.

The second section focuses on the worldwide urgent topic of migration and education (section editors: Lisa Damaschke-Deitrick and Alexander W. Wiseman). Education is seen as an option to facilitate the transition of migrant and refugee youth and their families into the new countries and communities. Schooling is one part of education only; education works much more as a mechanism for social integration. After describing the current situation of migration and seeking refuge worldwide, Damaschke-Deitrick and Wiseman emphasize the aspects of trauma, identity and language as characteristics of the refugees' experiences. In this respect the following three chapters contribute to this topic in specific ways. Johanna Fleckenstein, Débora B. Maehler, Howard Ramos, and Paul Pritchard examine language as a predictor and an outcome of acculturation. In their literature review, the author team presents empirical findings and highlights research gaps with regard to the topic of language skills and refugee children and youth. Michael Filsecker and Hermann Josef Abs present an item set how to measure attitudes towards refugees. Their development is connected to the International Civic and Citizenship Study (ICCS) and its German extension, and addresses the challenges and limitations of the test instrument. In the last chapter, Ericka Galegher examines female refugees' experiences in Egyptian higher education. She interviewed female refugee students from Syria and Yemen and analyzed cultural and linguistic implications of this forced transition. By presenting different perspectives and national contexts, the section on migration, refugees and education shows how important education is for these societal challenges.

The third section of this book (section editors: Nina Jude and Janna Teltemann) analyzes the interplay of large-scale assessments (LSA) and education policy in the US, Europe and other countries around the globe. Considering the examples of PISA and TIMSS, the authors point to several aspects of this relation and examine the effects of large-scale assessments on different political levels from the national one to federal and local policies. In the first chapter, Nina Jude and Janna Teltemann discuss whether for Germany an impact of education policies that resulted from the PISA shock can be found in the PISA outcomes of later assessment cycles. This is followed by Kerstin Martens' and Dennis Niemann's analysis of policy reactions to LSA results and the effects of such educational reforms on the classroom level in one of the German states. Lluís Parcerisa, Clara Fondevila and Antoni Verger focus on transfer processes between LSA cycles and education policies in different European countries whereas David C. Miller and Frank T. Fonseca examine changes in TIMSS results pointing to ways to identify achievement gaps over time.

The fourth section on the management and use of digital data in education (section editors: Sieglinde Jorntz and Laura Engel) unfolds the topic in two directions: first with respect to education governance structures and institutions and second to educational school practice. An increasing amount of data leads to the development of instruments that reshape education governance institutions and schools. Research points to the potentials and risks that lie in this usage for democratic societies and the education of children and adolescents. By taking national and supranational context into account, the four chapters show different implications for education governance and practice. Sigrid Hartong's research is framed by a comparison between Germany and the US about data usage in school administration agencies. She presents insights into the American context and highlights how school monitoring is linked to an extensive graphical way of relation making. Steven Lewis' context is the supranational institution of the OCED and its program *PISA for Schools*. Though the OECD gathers data from single schools in this program, they report back schematized results and do not give the context from which the data was taken. The third chapter written by Bernard Veldkamp, Kim Schildkamp, Merel Keijsers, Adrie Visscher and Ton de Jong presents results from a study carried out in the Netherlands. Its aim was to explore the potentials and challenges for big data usage from primary to higher education. Looking deeper into the classroom, the author team Elmar Souvignier, Natalie Förster, Karin Hebbecker, and Birgit Schütze presents the web-based monitoring system *quop* that was developed in Germany to provide teachers with a tool to measure learning progress within the classroom. The section spans the analyses from a supranational via national to local context in which digital data is used.

The fifth section of this book (section editors: Marcelo Parreira do Amaral and Paul Fossum) considers education from global perspectives and analyses factors that influence global education trends. In their introduction, Marcelo Parreira do Amaral and Paul Fossum examine facets of the "Global Education Industry" and the current trends of economization, commodification, privatization and standardization that are shaping education worldwide. With this perception in mind, the following papers of this section span over the different educational sectors from school settings to the field of lifelong learning and adult education. Sabine Hornberg takes a closer look at the role of the International Baccalaureate Organization for internationally generated standardization in education and the expansion – not only within the private sector, but also in the public education sector – of the International Baccalaureate Certificate, as a way of providing internationally regulated access to universities and thus transgressing national education systems. Alexandra Ioannidou and Annabel Jenner turn their attention to the non-regulative character of Adult and Continuing Education. This constellation opens space for non-public and international actors to exert influence on the

market, which for example becomes clear when looking at the International Organization for Standardization.

The sixth section is related to the overarching, but seldom discussed issue of translation in an international research context (section editors: Norm Friesen and Rose Ylimaki). Relating to German conceptions of education, Norm Friesen shows how important and limited translation aspects are for an international understanding within the discipline. This indicates how deeply implemented in the cultural and theoretical context each terminology is. In this sense, Kathrin Berdelmann takes a deeper look at how German terms of education history are translated into English. She expands this perspective to the French language and discusses the re-interpretation of educational terms. Inés Dussel broadens the scope to a global historical perspective. She argues that translation has been a central part of research since it became a scholarly practice and is part of every research action that has ever taken place. In this respect, the dominant usage of English in research contexts may lead to a narrowing of concepts. Finally, Britta Upsing and Musab Hayatli unfold how assessment studies deal with the translation issue in practice. They explore the process of translation as well as strategies to approach this process.

This section on translation processes closes the publication and builds – in a certain manner – a basis for all other contributions. International exchange in education research has to keep in mind that most scholars and researchers are linked to their national communities and the context of the discipline. In this regard, international cooperation is well-advised not to adjust these differences to one standard or scale, but to broaden and welcome multiple concepts of method, theory and thought.

References

- Aljets, Enno (2014): *Der Aufstieg der Empirischen Bildungsforschung. Ein Beitrag zur institutionalistischen Wissenschaftssoziologie*. Wiesbaden: Springer VS.
- Allemann-Ghionda, Cristina (2004): *Einführung in die Vergleichende Erziehungswissenschaft*. Weinheim and Basel: Beltz.
- Bellmann, Johannes (2017): *The Reception of Dewey in Germany after PISA. On the Language of Progressivism and its Adaptability*. In: Overhoff, Jürgen/Overbeck, Anne (eds.): *New Perspectives on German-American Educational History. Topics, Trends, Fields of Research*. Bad Heilbrunn: Verlag Julius Klinkhardt, pp. 175-192.
- Berg, Christa/Herrlitz, Hans-Georg/Horn, Klaus-Peter (2004): *Kleine Geschichte der Deutschen Gesellschaft für Erziehungswissenschaft. Eine Fachgesellschaft zwischen Wissenschaft und Politik*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Carnoy, Martin (2019): *Transforming Comparative Education. Fifty Years of Theory Building at Stanford*. Stanford: Stanford University Press.
- Dewey, John (1949): *Demokratie und Erziehung. Eine Einleitung in die philosophische Pädagogik*. Dt. von Erich Hylla. Braunschweig: Westermann.
- Dewey, John (1993): *Demokratie und Erziehung. Eine Einleitung in die philosophische Pädagogik*. Translated by Erich Hylla, edited by Jürgen Oelkers. Weinheim: Beltz.
- Döbert, Hans (2015): Germany. In: Hörner, Wolfgang et al. (eds.): *The Education Systems of Europe*. Cham, Heidelberg, Dordrecht, New York: Springer International, pp. 305-333. doi: 10.1007/978-3-319-07473-3_19
- Fossum, Paul (2021, forthcoming): US Education: Restless Federalism in Historical and Contemporary Context. In: Jornitz, Sieglinde/Parreira do Amaral, Marcelo (eds.): *Education Systems of the Americas*. New York: Springer International.
- Hylla, Erich (1954): *Education in Germany. An Introduction for Foreigners*, Frankfurt am Main: Hochschule für Internationale Pädagogische Forschung.
- Hylla, Erich (1928): *Die Schule der Demokratie. Ein Aufriss des Bildungswesens der Vereinigten Staaten*. Berlin and Leipzig: Julius Beltz.
- Jornitz, Sieglinde (2018): Bildungsforschung zwischen Politik und Praxis. In: Kleeberg-Niepage, Andrea/Rademacher, Sandra (eds.): *Kindheits- und Jugendforschung in der Kritik. (Inter-)Disziplinäre Perspektiven auf zentrale Begriffe und Konzepte*. Wiesbaden: Springer VS, pp. 321-352.
- Jornitz, Sieglinde/Wilmers, Annika (2018): Transfer zwischen Bildungsforschung und Bildungspolitik. Die europäische Dimension und Diskussion. In: Drossel, Kerstin/Eickelmann, Birgit (eds.): *Does „What works“ work? Bildungspolitik, Bildungsadministration und Bildungsforschung im Dialog*. Münster/New York: Waxmann, pp. 191-205.
- Koinzer, Thomas (2011): *Auf der Suche nach der demokratischen Schule. Amerikafahrer, Kulturtransfer und Schulreform in der Bildungsreformära der Bundesrepublik Deutschland*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Koneffke, Gernot (1988/2018): Zum Konstitutionsproblem der Vergleichenden Pädagogik. In: Koneffke, Gernot/Bierbaum, Harald/Herrmann, Katharina (eds.): *Widersprüche bürgerlicher Mündigkeit. Materialistische Bildungstheorie als politische Explikation der Pädagogik*. Vol. II. Baltmannsweiler: Schneider Hohengehren, pp. 41-53.
- Lagemann, Ellen Condiliffe (2000): *An elusive science. The troubling history of education research*. Chicago/London: University of Chicago Press.
- Martens, Kerstin (2010): USA – Wie man PISA auch ignorieren kann. In: Knodel, Philipp/Martens, Kerstin/Olano, Daniel de/Popp, Marie (eds.): *Das PISA-Echo. Internationale Reaktionen auf die Bildungsstudie*. Frankfurt, Main: Campus Verlag, pp. 235-250.
- Overhoff, Jürgen/Overbeck, Anne (2017) (eds.): *New Perspectives on German-American Educational History. Topics, Trends, Fields of Research*. Bad Heilbrunn: Verlag Julius Klinkhardt.
- Phillips, David/Ochs, Kimberley (2010): Processes of Policy Borrowing in Education: some explanatory and analytical devices. In: *Comparative Education*, 39, 4, pp. 451-461. doi: 10.1080/0305006032000162020
- Ravitch, Diane (2010): *The death and life of the great American school system. How testing and choice are undermining education*. New York: Basic Books.

- Rury, John L. (2014): *Turning Points: Critical Periods in Educating the United States*. In: Mulcahy, D.E./Mulcahy, D.G./Saul, Roger (eds.): *Education in North America*. London et al.: Bloomsbury Academic.
- Schneider, Mercedes K. (2016): *School choice – the end of public education?* New York and London: Teacher College Press.
- Spring, Joel (2018): *The American School. From the Puritans to the Trump Era*. 10th edition. New York/London: Routledge.
- Steiner-Khamsi, Gita/Waldow, Florian (2012): *World Yearbook of Education 2012. Policy Borrowing and Lending in Education*. London and New York: Routledge.
- Suter, Larry E./Smith, Emma/Denman, Brian D. (eds.) (2019): *The SAGE Handbook of Comparative Studies in Education*. London et al.: SAGE.
- Terhart, Ewald (2017): "Research on Teaching" in the USA and "Didaktik" in (West-)Germany: Influences since 1945. In: Overhoff, Jürgen/Overbeck, Anne (eds): *New Perspectives on German-American Educational History. Topics, Trends, Fields of Research*. Bad Heilbrunn: Verlag Julius Klinkhardt, pp. 159-174.
- The National Commission on Excellence in Education (1983): *A Nation at Risk: The Imperative for Educational Reform. A Report to the Nation and the Secretary of Education United States Department of Education*.
- Uljets, Michael/Ylimaki, Rose M. (2017) (eds.): *Bridging Educational Leadership, Curriculum Theory and Didaktik*. Cham: Springer. doi: 10.1007/978-3-319-58650-2.
- Urban, Wayne J./Wagoner, Jr., Jennings L./Gaither, Milton (2019): *American Education. A History*. 6th edition. London/New York: Routledge.
- Waldow, Florian (2015): Hat sich die Vergleichende Erziehungswissenschaft zu Tode gesiegt? Wozu brauchen wir sie noch? In: *Tertium Comparationis*, 21, 1, pp. 130-139.
- Waterkamp, Dietmar (2006): *Vergleichende Erziehungswissenschaft. Ein Lehrbuch*. Münster et al.: Waxmann.
- Waterkamp, Dietmar (2012): *Pädagogische Charakteristik der Schulwesen in zehn Nationen. Dresdner Vorlesungen 2012*. Dresden: TUDpress.

I. School Leadership and School Development

Section Editors:

Stefan Brauckmann-Sajkiewicz, University of Klagenfurt

Petros Pashiardis, Open University of Cyprus

Ellen Goldring, Vanderbilt University

Comparing School Leadership Practices in Germany and the United States: Contexts, Constructs and Constraints

Stefan Brauckmann-Sajkiewicz¹, Petros Pashiardis² and Ellen Goldring³

1. Comparing policy contexts underlying school leadership practices in the US and Germany

Educational policy contexts differ because of the need for more democratic participation and more efficient public management as well as the concern to improve the quality of education (Wößmann/Lüdemann/Schütz/West 2007). However, it seems that this has, so far, resulted mainly in transferring more responsibility and decision-making authority to schools. For instance, in both countries the states/*Länder* have overall responsibility for the education of young citizens, and the federal government has only limited authority for educational policy making. Moreover, school leadership preparation is highly developed and required as a prerequisite for the advancement to the principalship in the USA; on the other hand, in Germany, the professionalization efforts concerning the new roles and functions of school leaders have been intensified in the last few years. This is due to the fact that the principal's tasks have been extended in connection with new public management ideas, and his or her status has changed fundamentally. Traditional teacher training does not provide sufficient training for the tasks associated with the leadership of a school as an organization. Consequently, the states qualify teachers and school principals for the new tasks, and have created corresponding regulations and offers. For example, in Brandenburg, an additional qualification in school management can be acquired. Hessen, Lower Saxony and North Rhine-Westphalia have introduced staged procedures for qualification to take on positions in schools. In Berlin and Hamburg, participation in corresponding qualification programs is mandatory. As far as the terminology is concerned in the German-speaking world, "school leadership" and "school principalship" are not always clearly

-
- 1 Stefan Brauckmann-Sajkiewicz is Professor for Quality Development and Quality Assurance in Education, Institute of Instructional and School Development, at the Alpen-Adria-University Klagenfurt. Email: Stefan.Brauckmann@aau.at
 - 2 Petros Pashiardis is Professor of Educational Studies, Educational Leadership and Policy at the Open University of Cyprus. Email: p.pashiardis@ouc.ac.cy
 - 3 Ellen Goldring is Professor of Education and Leadership at the Peabody College, Vanderbilt University, Nashville. Email: ellen.goldring@vanderbilt.edu

distinguished from each other in the areas of responsibility examined. Neither is the term “leadership function” standardized, nor are the schools equipped with comparable functional leadership positions such as in the US. For instance, the Brandenburg School Act, the Hessian School Act and the North Rhine-Westphalian School Act distinguish the tasks of the broader leadership team from those of the school principal. On the other hand, Hamburg concentrates the leadership responsibility exclusively on the school principal. In other *Länder* (federal states), in addition to the overall responsibility of the school principal, cooperation within the broader leadership team is emphasized (cf. Hanßen 2013).

As we can see, due to deeply rooted cultures both at national and local levels underlying the concept of leadership policies and leadership practices, these aforementioned new public management drivers of change can be interpreted in various ways. The national and local traditions often affect practices to a larger extent than global trends (Dimmock/Walker 2000). This picture becomes even more complex when considering that policy and the organizational structure impact principals’ prerequisites as well as the expectations on principals’ actions.

School-based management and leadership are crucial aspects of any reform strategy in which change and responsibility are involved (De Grauwe 2004), and therefore their relationship merits further study. At the same time, the extent to which the legal and organizational framework affects principals’ professionalism in terms of adherence, coherence, and consistency between expectations and formal regulations has so far not been studied to a sufficient degree. In fact, few countries have explicit policies on the professional development of principals that are linked to a wider reform agenda, even where major programs of decentralization and delegation of authority are underway. Questions emerge regarding the effectiveness and success of school leaders’ actions as well as the responsibility for the school development process and its design. This has happened because the scope of leadership tasks has been broadened, and individual schools are facing higher demands regarding self-organization and responsibility of their operations. A reorganization of individual school processes has thus been initiated, clearly referring to role models from the domain of economics, as is evident from the emphasis on management and organization, as well as explicit reference to topics from organizational theory and development and new forms of coordination (Pont/Nusche/Moorman 2008).

Moreover, school leaders need to keep a balance between the external and the internal operations of the school, by looking both outside and inside the school, as they are responsible for the school in its entirety. According to the concept of New Public Management, school leaders are assumed to primarily possess pedagogical leadership potential, but also to be fully committed to and held responsible for a high-quality development of the

organization and its staff. Leaders need to take into perspective the increasing accountability and the consequences of public education systems being granted more autonomy for decision making at the school level. Thus, the school leader holds the key to quality-oriented school development, owing to increased autonomy and decision-making power but also to the increase in accountability concerning educational administrators and the school maintaining body.

Bearing the situations outlined above in mind, at their inception, the AERA Leadership round tables (organized by the team of “International Cooperation in Education” at the Leibniz Institute for Research and Information in Education in 2015, 2016, 2017, 2018, 2019) were dedicated to observing the spectrum of leadership practices and highlighting interesting developments from the German-speaking and American research areas, as well as gaining inspiration for reform efforts from the ensuing discussions. During these lively debates researchers from the US and Germany (and increasingly other parts of the world) could rely on each other’s valuable, empirically grounded educational research expertise and find out more about the particularities of the respective education systems. New perspectives and critique emerged on the quality and quantity of comparative educational research in the field of school leadership. In fact, we were able to discuss methodological challenges of comparative research in education when it comes to presenting and contrasting findings on school leadership styles from Germany and the US. It has been argued that systematic international comparisons in the field of school leadership should take more into consideration the specific contextual antecedents which might contribute to the structural as well as cultural shape of education systems. Additionally, authors have argued in favor of a more context-oriented comparative approach, which combines context information with empirical (quantitative and qualitative) analyses (Döbert/Sroka 2004). For instance, an increase in school autonomy has led to a change in the proportion of organizational and administrative tasks imposed on school leaders. The expanded workload, as a result of charging new tasks onto principals, coupled with increased demands for effectiveness and efficiency, is viewed in a critical way not only by organizations representing school leaders, but also by educational researchers. Particularly, educational researchers have called into question whether traditional leadership qualification measures as well as actual leadership practices still hold up to the extended school leadership tasks.

More specifically, the organization model of schools envisioned in the context of new public management approaches needs to react upon the speed, complexity and visibility of changes in a school environment. Thus, the realization that schools need to become more flexible, innovative and accommodating in order to fulfill their mission seems to be an inevitable task. The Holistic Leadership Framework (Brauckmann/Pashiardis 2011; Pashiardis

2014; Pashiardis/Brauckmann 2014) represents one of the most recent attempts to research the above in Europe. The specific framework was developed and validated in seven European countries initially (England, Norway, Germany, Slovenia, Hungary, Italy, and The Netherlands), within the context of the EU-funded LISA (Leadership Improvement for Student Achievement) project.

According to this framework, school principals' behaviors and actions are operationalized in terms of five leadership styles: the instructional, participative, structuring, entrepreneurial, and personnel development styles. Consequently, it has been suggested that school leaders can balance the outside with the inside worlds of schools in a more effective and successful way through two main styles of the Pashiardis-Brauckmann Holistic Leadership Framework (Brauckmann/Pashiardis 2011): the Entrepreneurial and the Pedagogical leadership styles. This means that these new leaders should be scanning their environment strategically; providing a good diagnosis about the readiness and ability of their personnel to act; being flexible enough to utilize a variety of leadership styles as well as their hybrids; and influencing both the outside as well as the inside school environment. These actions are intentional in order to stimulate the school improvement process, through a closer collaboration between the various stakeholders at the school level, who operate both in the school's periphery as well as the school's internal environment; in the end, this is realized through the improvement of the teaching and learning processes (Pashiardis/Brauckmann 2018).

A comparison of two Western educational landscapes by analyzing policy documents and searching for principals' rights, responsibilities, and support systems will verify the importance of acknowledging national cultural similarities and differences when discussing schools and their leadership, especially when referring to international trends and movements. By using the same theoretical and/or methodological framework, we can reveal the various prerequisites and expectations principals have in different settings. At the same time, if the global conversation about trends and movements is not supported by empirical data, including the national and local cultural and structural contexts, there is a risk that the findings and conversations become so general that we miss important insights. Moreover, as Peter Ribbins, Peter Gronn and Petros Pashiardis suggested in a jointly edited volume of the *Journal International Studies in Educational Administration* (ISEA) in 2003, consideration of two wider implications of policy-copying attendant on heightened global awareness of different cultural practices should be discussed:

- whether traditional patterns of Anglo-American hegemonic diffusion in educational leadership will perpetuate themselves and,

- whether models of leader formation are more likely to diverge in the interests of cultural particularism or to converge around a norm of cultural universalism.

The concern with cultural diffusion in the areas of educational management and leadership was summarized by Dimmock and Walker (1998: 564) as: “These [Western] paradigms tend to be adopted uncritically and unquestioningly by academics and practitioners in societies and cultures that bear little similarity to those in which the theories originated.” A concern with Western hegemony may simply be because it happens to be the West – with its imperial and colonial past, its wealth, its military power, its liberal-capitalist values and so on – that is dominant, as opposed to another region of the globe with a different set of cultural values. On the other hand, the unease about Western ways and means may be nothing more than an attempt to maintain cultural purity.

In the end, it is important that the school’s goals link back to student achievement; it is even more important for the schools to understand that with great power comes great responsibility, implying that autonomy and accountability are two sides of the same coin called “school quality assurance and development”.

In order to determine the right mix, it is necessary to elaborate on the relationship of accountability and autonomy from a school leadership perspective. There is not an easy straightforward answer. In some cases, there is a need for system leadership and to align organizational/school objectives with personal goals and needs (Goldring/Huff/May/Camburn 2008). From another perspective, there is also a need for distribution, where leaders at all levels cooperate and integrate their professions toward the same goals. This requires professional and skilled actors. To understand, analyze, and lead schools, engaged and highly qualified individuals are needed who can make the right decisions, even if the culture and the national trends point toward another direction. We can, as researchers, contribute to schools’ development by revealing the variation in aspects that many take for granted.

In short, the ongoing debate is characterized by finding the most fruitful balance between contextual challenges and leadership practices, so that we do not overburden leaders, teachers and students along the way. In order to do just that, we need the right “dosage” of external interferences by educational systems and their monitoring mechanisms, towards the schools’ ability to organize itself in flexible ways so that it can accomplish its mission of teaching and learning with the least interference possible.

2. The context-sensitive approach to leadership styles

Therefore, more system-related background information is needed in order to make substantiated judgements regarding the structural and cultural embedment of leadership policies and leadership practices in the US as well as in Germany (Döbert/Klieme/Sroka 2004). In that regard, we developed a framework of guiding questions for the contributors to this section of the book. Those guiding questions were structured according to prominent effective leadership styles which could be identified in all education systems. Our chosen guideline questions had to allow for the complexity of the subject matter, and should be maneuverable enough to ensure the optimal recording of empirical and descriptive findings, the specific sets of conditions as well as intra-national variances.

Furthermore, we formed binational teams of leadership researchers from the US and Germany. In particular, the chosen guideline questions had to ensure that the empirical research on school leadership in the US and Germany could elicit adequate empirically-founded conclusions regarding the chosen effective leadership styles. Of course, the findings from those guideline questions can only claim to be a rather qualitative interpretation and integration of selected facts and reflections provided by the binational expert teams (Döbert/Sroka 2004). We are aware of the limitations of comparative studies in our field which will never provide fully fledged explanation patterns with regard to observed differences in the impact of leadership styles on measurable educational outcomes (Hallinger/Liu/Piyaman 2019; Marfan/Pascal 2018). It would be even less possible to offer recipes for the most effective blend of leadership styles (how the quality of schools and instruction might be dramatically improved) that can be easily imported or exported from country to country (Hallinger 2018). In fact, it might be argued that this kind of export/import process from one country context to another is highly dependent on school leaders' (1) personality characteristics, (2) education and training in school leadership, coupled with (3) experience and common sense. However, the leadership actions that follow will also be dependent on (a) the level of success (or lack thereof) that the school is functioning at, as well as (b) the risk-averse or risk-prone personality of the school leader (Pashiardis/Brauckmann 2019: 493). Based on the last point made, we are tempted to speculate that more successful schools will be risk-averse and not so willing to try out new ideas, as the sentiment will probably be that "we are already doing well" and that there is no need to place our school at risk. On the other hand, the opposite could be true as well, i.e., that the school can be more risk-prone, as it can survive the possibility of failure with little or no damage. Either course of action will depend on how risk-prone or risk-averse the school leader is (Tversky/Kahneman 1974).

The approach introduced in this section can serve to provide better insights regarding the quality and quantity of the leadership challenges experienced in the US and in Germany. Thus, we can identify problems that might be sometimes, despite the varying conditions and the contexts, more identical on a structural or a functional level between the US and Germany than within the two countries. Prior to publication, the chapters were critically reviewed by the editors of this section and revised by the authors. The editorial team concluded that there seems to be a general agreement that school as an institution faces problems of a pedagogical and didactic nature, as well as social and communicative problems and (finally) structural ones. School leaders in particular are challenged to find effective strategies for action and problem-solving in increasingly complex environments. Therefore, at a minimum, clarity on the following issues is needed for the chapters included in this “School Leadership” section of the book:

- What is legally (de jure) expected and required of school principals in terms of leadership in challenging environments, leadership for learning and distributed/shared leadership?
- Does any evidence exist on what school leaders are doing with respect to all of the above (de facto)?
- What kind(s) of conclusions can be drawn as a result of the juxtaposition of the de facto and the de jure description, as per the above (inter)national commonalities, national/local particularities with regard to the German and the US perspective?

Against this background, the leadership section of this book aims at discussing and analyzing these questions and, at the same time, addressing the examination of explicit responsibilities, mandates, and support, as regards schools and school leaders in the US and Germany. Authors will investigate different aspects from their perspectives and country backgrounds, and may refer to the potential of drawing initial comparisons. The juxtaposition of discernible differences and similarities stemming from new educational governance-related goals will inform about the contextual conditions under which school principals operate, and the governance patterns and objectives that they have in mind when implementing their own leadership cocktail (Brauckmann/Pashiardis 2011). Sound evidence-based knowledge of the differences and commonalities of educational systems within these countries will enable a discussion on the benefits and detriments of a transnational model of educational leadership as is often envisioned in the internationally-oriented leadership community. It must be stressed that comparative studies on school leadership so far provide little information on the national contexts underlying school principals’ actions (Brauckmann/Geissler/Feldhoff/Pashiardis 2016).

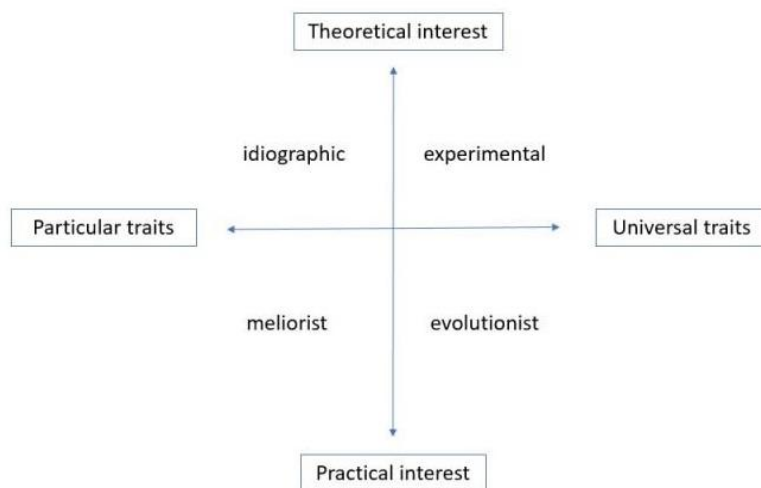
On the other hand, one such recent approach was attempted by Pashiardis, Pashiardi & Johansson, (2016), who examined successful school leadership around the world. It became evident that there is a tendency to compare and find out the “best” practices and interventions and to identify common features that help us build success in the different regions around the world. At the same time, it became clearer that what is valued as best education and what is valued within education is politically motivated and values-driven. Moreover, education systems are micro-political systems, and in this regard, they represent the culture and values of real people. Thus, what is successful and effective in one part of the world maybe “good enough” in another part, because, depending on the level of development of a society and an educational system, what is successful and what is effective suddenly becomes very relative.

Regardless of contextual differences, some common features seem to be identifiable. First of all, the interplay between challenging contexts and the various actors at the school level is an important factor that must be dealt with from a school principal perspective (Miller 2018; Johnson/Dempster 2016; Bottery 2006). Second, the role of the principal as leader of leaders is a prominent one and enhances, mostly indirectly, students’ performance. This leadership role, which seems to be evident in most studies in the international literature, called *instructional leadership* within the USA context or *pedagogical leadership* in other contexts, has an impact on the quality of teaching and learning that takes place at the school level. Third, *distributed leadership* seems to be another common feature irrespective of context (Goldring/Huff/Spillane/Barnes 2009). Fourth, school leaders exhibit an entrepreneurial style of leadership, which is an essential component of the leadership cocktail mix irrespective of context. Finally, school leaders seem to be value-driven and especially trust-driven. Successful principals have in common the fact that they are guided by a set of values consisting of professional, social and political components that they convey to others. This can be seen as their personal philosophy, and it is a common characteristic of leaders in different parts of the world (Pashiardis/Pashiardi/Johansson 2016).

Thus, in a number of contexts across the globe in an effort to “glocalize” our learning, the intention is not only to provide a description of the very different systems, but also the varied ontological and epistemological discourses of differing approaches and lines of thought. As the world becomes smaller – indeed an ecumenical village – the topic of international or comparative perspectives in what constitutes success and effectiveness has attracted more attention. The evidence given in the educational leadership literature stems from the fact that, over the past few years, the ideas and the language of theory and practice – in what constitutes success and effectiveness of school leadership – have become increasingly debated and explored in an international and comparative context. Moreover, in comparative

education, four main points of divergence have been distinguished according to the criteria of practical against theoretical interest on the one hand, and an interest in universal as opposed to particular traits on the other (Hörner 1997: 70f.; cf. Hörner/Döbert 2008: 1-10). These are the idiographic, the meliorist, the evolutionist, and the experimental functions of a comparative approach, as can be seen in Figure 1.

Figure 1. The four functions of a comparative approach (based on Hörner 1997)



For our purposes, we will concentrate on the idiographic aspect of the above comparative paradigm. The purpose of the idiographic function is to work out the particularities and the unique traits of individual phenomena in an education system (Hörner 1997). Comparative research is therefore interested in aspects that render one educational system distinct from all others. This search for particularities is complemented by the search for common features. The ideographic function is of primary importance here, as the country analyses are meant to offer reliable knowledge about particular traits of legally bound rights, duties, and responsibilities. The identified specifics and similarities of national configurations can serve as important context knowledge for judging whether structural analogies allow for the transfer of best practices. Based on the above contextual setting as well as the three groups of questions, the following chapters are included in our section:

3. Sequence of contributions in this section

3.1 Team 1: Leadership in challenging environments

Esther Dominique Klein, Michelle Young, Susanne Böse

The increasing interest in school effectiveness and school improvement within challenging contextual boundaries brings school leadership to the forefront, as Pashiardis (1996) points out; this can be seen from the fact that educational mandates, communities, parents and legislators show a growing interest in the leadership of schools aiming to achieve greater participation in the educational process (Pashiardis/Brauckmann 2018; Pashiardis/Brauckmann/Kafa 2018). The main idea is how we can resolve the paradox, and even better, convince that it is possible to have schools located in unfavorable teaching and learning conditions and yet, producing high student academic results. Thus, this chapter explores the high achievement of the students coupled with the paradox of a school's operating conditions, as is revealed from the description of the challenging backgrounds and the equally challenging contextual characteristics of both students and the school. The chapter further presents an authentic school improvement process, as is evident from the various school factors, which do not copy any school improvement model from somewhere else as their educational policy loan. In conclusion, it is stressed that school leadership in challenging circumstances is not a “one off quick fix activity”. It is a continuous process that requires determination from the people involved. Furthermore, leadership at all levels in the school community may ensure sustainable improvement in increasingly complex, dynamic and challenging environments.

3.2 Team 2: Leadership for learning

Pierre Tulowitzki, Marcus Pietsch, James Spillane

Over the last two decades, Leadership for Learning (LFL) has emerged as a concept that integrates various educational leadership theories and concepts into one comprehensive theoretical model, i.e. instructional leadership, transformational leadership and shared leadership. As the authors of this chapter argue, in contrast to the concept of instructional leadership, where leadership is seen to reside with holders of a formal position, leaders within the Leadership for Learning framework are understood as emergent leaders, irrespective of whether they have been appointed to an official position or not. This – at its core – can be seen as a distributed perspective. Thus, the contribution offers another perspective concerning the role of culture and

context as factors that shape educational leadership before delving into expectations and requirements as well as actual practices of school principals in terms of Leadership for Learning in Germany and the US. Culturally bound as well as more generalized conclusions are drawn as to how Leadership for Learning can be conceptualized and institutionalized. In essence, with this chapter the authors are trying to further illuminate the discussion about how contextual forces at the macro and micro levels help shape important terms, such as: instructional, learning-centered and pedagogical leadership.

3.3 Team 3: Distributed/shared leadership

Barbara Muslic, Jonathan Supovitz, Harm Kuper

Distributed leadership is used as a synonym for cooperative, shared or democratic leadership (i.e. Leithwood/Seashore Louis/Anderson/Wahlstrom 2004; Woods 2004; Harris 2008; Marks/Printy 2003). It has caught the attention of researchers and policy-makers. A distributed view of leadership incorporates the activities of multiple individuals in a school (Spillane/Halverson/Diamond 2004). The basic idea is to have a broad distribution of tasks and, at the same time, provide bounded empowerment to followers and members of an organization such as a school. In essence, it requires the correct “dosage” of distribution and division of duties, responsibilities and powers in order to fulfil the organization’s goals and objectives (Harris 2004, 2008; MacBeath/Oduro/Waterhouse 2004; Huber 2008; Bonsen 2009; Harris/Chapman 2002; Camburn/Rowan/Taylor 2003; Spillane/Diamond 2007). Thus, it is essential to view the school as a professional organization where mechanisms of shared decision-making (Spillane 2006) are put in place. Distributed leadership can also be seen as a presumption of an indirect leadership effect on school quality development and student achievement from mainly school effectiveness research (i.e. Leitner 1994; Creemers/Reezigt 1996; Hill 1998; Bryk/Sebring/Allensworth/Luppescu/Easton 2010; Supovitz/Sirinides/May 2010).

Taking this background into account, the section aims to demonstrate how educational monitoring data can support school leaders in (strategically) realigning their work tasks, and thus also in adjusting their management into a more systemic direction in synch with the readiness and expertise of the school personnel to assume a greater number of responsibilities and authority. On the other hand, it would be useful to present different ways in which school leaders (can) use data from large-scale assessments (e.g., VERA 8) for the evaluation of their schools as well as classroom improvement as part of school monitoring in a distributed function.

4. Concluding remarks

Our binational teams of authors cannot describe the “right” school system and structures which have to be put in place in order to get the results that are needed by society. Instead, they illustrate the level of uncertainty about the creation of the processes and putting the school systems into place, which can create something like a jumping board from which everybody can leap into effectiveness. But a context sensitive division of leadership responsibilities would probably be fairer and more justified by stressing the fact that a successful leader is one that institutionalizes the right processes in order to achieve desired objectives and thus become (in the long run) effective (Pashiardis/Pashiardi/Johansson 2016); a mix of school development and school effectiveness driven measures.

In light of the above, it can be argued that there is a need to explore in a more systematic way “situational components of governance and leadership” with regard to whether these two terms are antagonistic or complementary in an effort to reposition the ongoing discussion of whether a new mix of leadership styles is needed (Brauckmann/Pashiardis 2011). In fact, school leaders around the world are increasingly being asked to do more with less, and do it better with regard to student outcomes by aligning the inner and outer worlds of schools, thus (re)creating a new leadership mix.

As a consequence, it remains to be seen whether school leaders of the 21st century need to embark on more Entrepreneurial leadership, which means: partnering with parents and other external actors in the school’s everyday activities; acquiring more resources for their schools; building strategic coalitions with external agents; and implementing a market orientation to leadership for their schools (Pashiardis 2012). Furthermore, school leaders still might need to employ more of a Pedagogical leadership style, which means: Defining and enabling the achievement of the pedagogical objectives; setting high expectations for self, staff, students (3Ss); monitoring and evaluating students and teachers; stimulating pedagogical innovation and risk taking, and participating in everyday pedagogical dialogues.

This could pave the way for a new generation of “edupreneurial” leaders in schools, thus bringing responsibility for pedagogical purposes and the entrepreneurial sense of risk-taking together (Pashiardis/Brauckmann 2018). Irrespective, it is more than evident that these different successful and effective leadership approaches and the diversity of non-standardized contexts within which they suddenly emerge and fade away (as mentioned in the three contributions) complicate matters even more and, indeed indicate the many differences in the world when attempting to harmonize educational issues internationally.

References

- Brauckmann, Stefan/Geißler, Gert/Feldhoff, Tobias/Pashiardis, Petros (2016): Instructional Leadership in Germany: An Evolutionary Perspective. *International Studies in Educational Administration (ISEA)*, 44, 2, pp. 5–20.
- Brauckmann, Stefan/Pashiardis, Petros (2011): A validation study of the leadership styles of a holistic leadership theoretical framework. In: *International Journal of Educational Management*, 25, pp. 11-32.
- Bonsen, Martin (2009): Führung, Delegation und 'distributed leadership'. Unterrichtswirksame Schulleitung in Zeiten der datengestützten Schulentwicklung. In: Bohl, Thorsten/Kiper, Hanna (eds.): *Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren*. Bad Heilbrunn: Klinkhardt, pp. 267-282.
- Bottery, Mike (2006): Educational leaders in a globalising world: a new set of priorities? In: *School Leadership & Management*, 26, 1, pp. 5-22.
- Bryk, Anthony S./Sebring, Penny A./Allensworth, Elaine/Lupescu, Stuart/Easton, John Q. (2010): *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Camburn, Eric/Rowan, Brian/Taylor, James E. (2003): Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. In: *Educational evaluation and policy analysis*, 25, 4, pp. 347-373.
- Creemers, Bert P.M./Reezigt, Gerry J. (1996): School Level Conditions Affecting the Effectiveness of Instruction. In: *School Effectiveness and School Improvement*, 7, 3, pp. 197-228.
- Dimmock, Clive/Walker, Allan (1998): Comparative educational administration: Developing a cross-cultural conceptual framework. In: *Educational Administration Quarterly*, 34, 4, pp. 558-595.
- Dimmock, Clive/Walker Allan (2000): Cross-Cultural Values and Leadership. In: *Management in Education*, 14, 3, pp. 21-24.
- Döbert, Hans/Klieme, Eckhard/Sroka, Wendelin (eds.) (2004): Conditions of school performance in seven countries. A quest for understanding the international variation of PISA results (*Studien zur international und interkulturell vergleichenden Erziehungswissenschaft*, vol. 3). Münster: Waxmann.
- Döbert, Hans/Sroka, Wendelin (eds.) (2004): Features of successful school systems. A comparison of schooling in six countries (*Studien zur international und interkulturell vergleichenden Erziehungswissenschaft*, vol. 4). Münster: Waxmann.
- Goldring, Ellen/Huff, Jason/Spillane, James P./Barnes, Carol (2009): Measuring the Learning-Centered Leadership Expertise of School Principals. In: *Leadership and Policy in Schools*, 8, 2, pp. 197-228.
- Goldring, Ellen/Huff, Jason/May, Henry/Camburn, Eric (2008): School context and individual characteristics: what influences principal practice? In: *Journal of Educational Administration*, 46, 3, pp. 332-352.
<https://doi.org/10.1108/09578230810869275>
- De Grauwe, Anton (2004): School-based management (SBM). Does it improve quality? Paper commissioned for the EFA Global Monitoring Report. UNESCO.

- Hallinger, Philip/Liu, Shengnan/Piyaman, Patnaree (2019): Does Principal Leadership Make a Difference in Teacher Professional Learning? A Comparative Study China and Thailand. In: *A Journal of Comparative and International Education*, 49, 3, pp. 341-357.
- Hallinger, Philip (2018): Bringing Context out of the Shadows of Leadership. In: *Educational Management Administration & Leadership*, 46, 1, pp. 5-24.
- Harris, Alma (2008): Distributed leadership: according to the evidence. In: *Journal of Educational Administration*, 46, 2, pp. 172-188.
<https://doi.org/10.1108/09578230810863253>
- Harris, Alma/Chapman, Christopher (2002): Democratic leadership for school improvement in challenging contexts. In: *International Electronic Journal for Leadership in Learning*, 6, 9.
- Harris, Alma (2004): Teacher leadership and distributed leadership. In: *Leading & Management*, 10, 2, pp. 1-9.
- Hill, L. Brooks (1998): Leadership. In: Christ, William G. (ed.): *Leadership in times of change*. Lawrence Erlbaum Associates, pp. 199-224.
- Hanßen, Klaus (2013): Rechtliche Regelungen zu Tätigkeitsfeldern von Schulleiterinnen und Schulleitern bei erweiterter Eigenverantwortung von Schule. Eine Untersuchung der Rechtslage in den Ländern Berlin und Niedersachsen (Ergänzung zu der Untersuchung der Rechtslage in den Ländern Bayern, Hessen und Nordrhein-Westfalen sowie in den Ländern Brandenburg und Hamburg). https://www.dipf.de/de/forschung/aktuelle-projekte/pdf/steubis/projekt-sharp-pdf/Endfassung_BE_NI.pdf
- Hörner, Wolfgang (1997): 'Europa' als Herausforderung für die Vergleichende Erziehungswissenschaft – Reflexionen über die politische Funktion einer Vergleichenden Disziplin. In: Kodron, Christoph/von Kopp, Botho/Lauterbach, Uwe (eds.): *Vergleichende Erziehungswissenschaft: Herausforderung, Vermittlung, Praxis*. Festschrift für Wolfgang Mitter zum 70. Geburtstag. Köln: Böhlau, pp. 65-80.
- Huber, Stephan Gerhard (2008): Steuerungshandeln schulischer Führungskräfte aus Sicht der Schulleitungsforschung. In: Langer, Roman (ed.): „Warum tun die das?“. Wiesbaden: VS Verlag, pp. 95-126.
- International Cooperation in Education (ice) (2019): Leveraging education research in a “post-truth” era: Multimodal narratives to democratize evidence. 2019 AERA Annual Meeting – DIPF. April 5-9. Toronto.
- International Cooperation in Education (ice) (2018): International Perspectives on Public School Systems. The Dreams, Possibilities and Necessity of Public Education. 2018 AERA Annual Meeting – DIPF. April 13-17. New York City.
- International Cooperation in Education (ice) (2017): Societal Challenges and Educational Research. Knowledge to Action Achieving to Promise of Equal Educational Opportunity. 2017 AERA Annual Meeting – DIPF. April 27-May 1. San Antonio, Texas.
- International Cooperation in Education (ice) (2016): Public scholarship to educate diverse democracies. 2016 AREA Annual Meeting – DIPF. April 8-12. Washington, DC.
- International Cooperation in Education (ice) (2015): Toward justice. Culture, language, and heritage in education research and praxis 2015 AREA Annual Meeting – DIPF. April 16-20. Chicago, Illinois.

- Johnson, Geer/Dempster, Neil/(eds.) (2016): *Leadership in Diverse Learning Contexts* (Studies in educational leadership, vol. 22). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-28302-9>.
- Leithwood, Kenneth A./Louis, Karen Seashore /Anderson, Stephen/Wahlstrom, Kyla (2004): *How Leadership Influences Student Learning*. Ontario: Center for Applied Research and Educational Improvement & Ontario Institute for Studies in Education.
- Leitner, David (1994): Do principals affect student outcomes? An Organizational Perspective. In: *School Effectiveness and School Improvement*, 5, 3, pp. 219-238.
- MacBeath, John/Oduro, George T.T./Waterhouse, Joanne (2004): *Distributed Leadership in Action: A Study of Current Practice in Schools*. National College for School Leadership, Nottingham.
- Marfan, Javiera/Pascual, Javier (2018): Comparative study of school principals' leadership practices: Lessons for Chile from a cross-country analysis. In: *Administration & Leadership*, 46, 2, pp. 279-300.
- Marks, Hellen M./Printy, Susan M. (2003): Principal Leadership and School Performance: An Integration of Transformational and Instructional Leadership. In: *Educational Administration Quarterly*, 39, pp. 370-397.
- Miller, Paul W. (2018): *The Nature of School Leadership: Global Practice Perspectives*. Intercultural Studies in Education. Cham: Springer International Publishing. <http://dx.doi.org/10.1007/978-3-319-70105-9>.
- OECD. (2008): *Improving School Leadership, Volume 1: Policy and Practice*. Paris: OECD.
- Pashiardis, Petros/Johansson, Olof (2016): *Successful school leadership. International perspectives*. London: Bloomsbury.
- Pashiardis, Petros/Brauckmann, Stefan/Kafa, Antonios (2018): Let the context become your ally: school principalship in two cases from low performing schools in Cyprus. In: *School Leadership and Management*, 38, 5, pp. 478-495.
- Pashiardis, Petros/Brauckmann, Stefan (2019): New Public Management in Education: A Call for the Edupreneurial Leader? In: *Leadership and Policy in Schools*, 18, 3, pp. 485-499.
- Pashiardis, Petros/Pashiardi, Georgia/Johansson, Olof (2016): Understanding The Impact Of Successful And Effective School Leadership As Practiced. In: Pashiardis, Petros/Johansson, Olof (eds.): *Successful school leadership. International perspectives*. London: Bloomsbury, pp. 195-209.
- Pashiardis, Petros/Brauckmann, Stefan (2014): Leadership Styles and School Climate Variables of the Pashiardis-Brauckmann Holistic Leadership Framework: An intimate relationship? In: Pashiardis, Petros (ed.): *Modeling School Leadership across Europe: In Search of New Frontiers*. Dordrecht: Springer, pp. 89-106.
- Pashiardis, Petros (2012): School leadership still matters. In: Jones-Parry, Rupert (ed.): *Commonwealth Education Partnerships 2012/13*, pp. 91-92.
- Pashiardis, Petros (ed.) (2014): *Modeling School Leadership Across Europe: In Search of New Frontiers*. Dordrecht: Springer.
- Pashiardis, Petros (1996): Environmental scanning in educational organizations: uses, approaches, sources and methodologies. In: *International Journal of Educational Management*, 10, 3, pp. 5-9.

- Pont, Beatriz/Nusche, Deborah/Moorman Hunter (2008): Improving School Leadership. Volume 1: Policy and Practice. Paris: OECD.
- Spillane, James P./Diamond, John B. (2007): Distributed leadership in practice. Critical issues in Educational Leadership Series. New York: Teachers College, Columbia University.
- Spillane, James P. (2006): Distributed leadership. New York: John Wiley & Sons.
- Spillane, James P./Halverson, Richard/Diamond, John B. (2004): Towards a theory of leadership practice: A distributed perspective. In: Journal of Curriculum Studies, 36, 1, pp. 3-34.
- Supovitz, Jonathan/Sirinides, Philip/May, Henry (2010): How Principals and Peers Influence Teaching and Learning. In: Educational Administration Quarterly, 46, 1, pp. 3-56.
- Tversky, Amos/Kahnemann, Daniel (1974): Judgment under Uncertainty: Heuristics and Biases. In: Science, 185, 4157, pp. 1124-1131.
- Woods, Philip A. (2004): Democratic leadership: drawing distinctions with distributed leadership. In: International Journal of Leadership in Education, 7, 1, pp. 3-26.
- Wößmann, Ludger/Lüdemann, Elke/Schütz, Gabriela/West, Martin R. (2007): School Accountability, Autonomy, Choice, and the Level of Student Achievement. International Evidence from PISA 2003. EDU/WKP

Successful Leadership in Schools Serving Disadvantaged¹ Communities in Germany and the USA

Esther Dominique Klein², Michelle D. Young³ and Susanne Böse⁴

1. Introduction

Schools that serve communities with a high proportion of residents who receive low wages or are dependent on social welfare, and of ethnic minorities and people who are learners of the language of instruction in schools, often care for students that are less well equipped to meet performance requirements of the school system, and there often is a mismatch between the habitus of the students and that of their mostly middle class teachers (Steins 2016). As a result, they are often struggling to attain their (organizational or educational) goals, and thus are in need of improvement. Research shows that “leadership effects are usually largest where and when they are needed most” (Leithwood et al. 2004: 3), and that leadership is of particular importance for the improvement of these schools (e.g., Potter et al. 2002).

Schools serving disadvantaged communities have not only drawn the interest of school improvement research (Bryk et al. 2015; Pashiardis/Brauckmann/Kafa 2018), but have also become the focus of educational policy efforts in both Germany and the United States of America (USA). However, while improving these schools has been a focus of scholars and politicians since the beginning of school effectiveness research in the USA in the late 1970s (Mintrop/Klein 2017), schools serving disadvantaged communities did not receive much attention in Germany before the 2000s,

-
- 1 The term disadvantaged communities is used in reference to schools that serve students “whose family, social or economic circumstances hinder their ability to learn in school” (RAND, 2019). We believe that using labels that emphasize the (perceived or real) challenges of the schools rather than the students can encourage teachers and principals to externalize reasons for poor performance. In using this language, we hope to assist leaders in making an honest assessment of their own contribution to the students’ success or the lack of the same.
 - 2 Esther Dominique Klein is Professor for School Improvement Research at the Philipps-University Marburg. Email: dominique.klein@uibk.ac.at
 - 3 Michelle D. Young is Professor for Educational Leadership and Policy at the University of Virginia. Email: mdy8n@virginia.edu
 - 4 Susanne Böse is Research Assistant at the DIPF | Leibniz Institute for Research and Information in Education, Frankfurt. Email: boese@dipf.de

and the earliest research studies in Germany were carried out in the 2010s (e.g., Böse et al. 2017; Racherbäumer et al. 2013). Accordingly, there is a wealth of studies on improving struggling schools in the USA, but very limited research from Germany, and this is especially true for research that explores the role of leadership in these schools. Not surprisingly, German scholars often refer to findings from the USA; however, these referrals generally fail to consider, to a significant extent, the institutional and contextual conditions that shape the chances of and barriers to school leadership in the two countries (e.g., Mintrop 2015). In our chapter, we therefore first differentiate the contextual conditions of school leadership and principals in the USA and Germany, before we describe the conditions of schools serving disadvantaged communities (henceforth: SSDC) and summarize research findings from both countries concerning successful leadership for SSDCs.

2. Defining expectations towards principals in SSDCs

When comparing “successful” school leadership in two different countries, we must take into account that different institutional contexts define the expectations for principals’ practice in general, and in SSDCs in particular. We do this by first looking at the defining principles of education in the two countries, and then describing the requirements this entails for the role of school principals.

2.1 USA

The modern education system in the USA was established at the beginning of the 20th century, when the governments of most states endeavored to gain some control over thousands of autonomous school districts. Even today, although state governments are *de jure* responsible for making decisions about the work of schools, the majority of decisions affecting the day to day work of schools are devolved to local school districts (Briffault 2005).

According to Tyack (1974), the goal of state governments in gaining and maintaining control over the education system, at least initially, was to professionalize and to improve education, to enhance its results, and to use research to develop the “one best system” (Tyack 1974) of education. As a result, schools were more or less organized like businesses that were run by managers who worked by the rules of efficiency (Marzano/Frontier/Livingston 2011). At the core of this was the idea that teaching could be “rationalized” (see Mintrop/Klein 2017). As a result, American schools at the

turn of the 20th century can be described as a hybrid of a professional and a managed organization (Mintrop 2015).

Since that time, thinking regarding the field of education and the role of educational leaders has evolved in the USA. Although in some contexts, principals continue to function as managers of the school, functioning separately from the faculty of the school (Brewer/Smith 2006), this model is becoming increasingly uncommon. In its place has emerged a notion of educational or instructional leadership. This notion of “educational leadership” is reflected within national leadership standards in the USA (ISSLC 1996, 2008; PSEL 2015).

The most recent Professional Standards for Educational Leaders (PSEL), “are student-centric, outlining foundational principles of leadership to guide the practice of educational leaders so they can move the needle on student learning and achieve more equitable outcomes” (NPBEA 2015: 1). For example, the PSEL standards place significant emphasis on students and student learning, operating from the understanding that leaders “must approach every teacher evaluation, every interaction with the central office, every analysis of data with one question always in mind: How will this help our students excel as learners?” (NPBEA 2015: 3).

As instructional leaders, principals in the USA are responsible for the school’s improvement (e.g., Sebastian/Camburn/Spillane 2018). Given their role in school and instructional improvement, many principals and their leadership teams have significant influence on and decision-making power over instruction. Principals supervise, evaluate, and seek to improve teacher performance, and are also evaluated by their own supervisors in the district, with the intention that leaders would refine their own skills and expertise.

This normative role is fundamental for defining the expectations directed at principals in SSDCs. Although leadership standards in the USA have advocated a human-centered approach to school improvement and fostering student learning, in practice, expectations continue to reflect the logic of the “business” model, wherein the quality of the school cannot be left to those who do the instruction, but must be managed from the top. Accordingly, principals are responsible for making sure that schools make progress, and they are the ones who are held accountable if adequate yearly progress is not made.

2.2 Germany

In contrast to the American system, the German education system, which was already established in the 18th century, was not built on the principles of business and the quest for “one best system”, but traditionally had a bureaucratic administration based on organized hierarchies and the

enforcement of rules, and focused on consistency and functionality rather than effectiveness and improvement (Brüsemeister 2012). In addition, the view of school education has been shaped by the understanding that teaching and learning processes must be designed case-based and individually by the teacher (e.g., Luhmann/Schorr 1982), which is why teachers have to be very well-trained and must be able to act autonomously. The authority over the teaching process is therefore entirely in the hands of the teachers.

As a result, the German school system is traditionally characterized by a high level of input regulation in the form of standardized teacher education, curricula and school law, but little external control over the process and output quality of schools, which was regulated by “professional accountability” entirely. In this system of bureaucratic and professional control, school principals were primarily “teachers with additional administrative tasks” who had to make sure that the school was able to operate according to the rules, but were not responsible for its effectiveness or improvement (Wiesner et al. 2015) and had no power over the teachers.

Since the late 1990s and the so-called “second empirical turn”, the bureaucratic structures were supplemented with elements of managerial structures. Today, school processes and results are supposed to be focused on effectiveness and improvement, by establishing a results-based quality management (Jann 2005). This essentially involves a “contract management” between the regional authorities and the school, which entails increased autonomy on the one hand (Rürup 2007), but also a partial delegation of the responsibility for the results to schools or, more precisely, principals (Brüsemeister 2012). The local and regional authorities have, in turn, withdrawn from “implementing improvement by rules”, and instead focused on evaluating (and, in theory, counselling) schools, which necessitated the implementation of state testing and school inspections in the 2000s (Dederig/Müller 2011).

Today, there are hardly any publications in German school improvement research that do *not* emphasize the importance of leadership by the principal. However, although principals are responsible for school improvement, they still have no real power over the teachers, remain a part of the teaching staff, and cannot make any substantial decisions without consulting with the faculty and, depending on the area, the authorities. Also, while they are *de jure* responsible for the improvement of their school and, for instance, conduct negotiations with authorities after inspections, they are not accountable for it. Furthermore, principals are not evaluated or required to participate in specific professional development focused on leadership (Klein/Tulowitzki 2020). Finally, while American principals of SSDCs are usually tightly guided by the district and/or state authorities, there is no such superstructure for German schools (Klein/Bremm 2020). As a result, if German principals do not seek external guidance, they are on their own. As Mintrop (2015) accurately

summarizes, the German states have tried to implement a public management reform, but forgot to put managers in the schools and in the local and regional school authorities.

3. Research about “successful” leadership in schools serving disadvantaged communities

Because of oppressive structures in school and society⁵, the distinct economic, cultural, and social capital of students from low socioeconomic status (SES) families, and issues of fit between the lives of the low SES students and the norms of institutionalized education, schools serving disadvantaged communities (SSDCs) often have lower academic outcomes and an increased level of discipline problems (Klein 2017).

3.1 Challenges for SSDCs

As a result, SSDCs are often identified as in need of improvement (e.g., Potter/Reynolds/Chapman 2002), and therefore receive particular attention – and often additional funding – from education policy. In a report for the Wübben Foundation, Klein (2017) points out that there are some differences in the systemic context that SSDCs operate in between Germany and the USA. In the USA, SSDCs are defined by the SES of their student population, which is generally determined by the proportion of students who are entitled to free or reduced school meals, and school districts and state governments usually collect precise data on the students attending each school with regard to SES and a variety of other dimensions, such as race, ethnicity, and family language. Disadvantaged schools are generally identified as those in which the average SES of students is below the national average. With a wealth of data available, many districts, states, and the federal government are able to allocate funding for schools that is tailored to their specific needs, as is the case, for instance, in the *Local Control and Accountability Plan* (LCAP) in California.⁶

5 Khalifa (2018) discusses the educational, occupational, housing, and legal (e.g., police brutality) inequities impacting low income communities, particularly racially and ethnically diverse communities. In Germany, these phenomena have, for instance, been discussed in the context of institutional discrimination (regarding institutional discrimination in education, see Gomolla/Radtke 2002).

6 California Department of Education: Local Control and Accountability Plan. <https://www.cde.ca.gov/re/lc/> [Download on 4 November 2019].

Klein (2017) notes that in Germany, identifying SSDCs is not as easy because neither schools nor the government collect precise data on the socioeconomic background of students, their ethnicity, or their German language learner status in most states (*Bundesländer*). Only a handful of states have implemented structures to provide SSDCs with more resources, and only a few states use individual student data to identify these schools (Weishaupt 2016). Berlin, for instance, has used data on the immigration background and SES of students to determine the allocation of teachers since the 1990s, and SSDCs receive additional funds that they use as they see fit (e.g., *Senatsverwaltung für Bildung, Jugend und Wissenschaft* 2013).

Research shows that the reasons for lower performance are multi-layered and can be traced back to the disadvantaged background of the students, to systemic barriers, but also to less adaptive and disadvantageous instructional and organizational factors *in the schools* influenced by the individual and collective beliefs of people in the school (Khalifa 2018). Too often teachers, leaders and other educational professionals believe that the reasons for the lower achievements of their students are first and foremost a result of their individual family resources and upbringing (Fölker et al. 2016; Nelson/Guerra 2014). At the same time, they underestimate their own influence on the education process of their students; this is due partly to the fact that traditional teaching practices and tools do not work equally well or at all for low income students, institutional norms often emphasize the individual deficits of the students (Valencia 2010), and many educators learned specific narratives when they started teaching (Khalifa 2018). As shared norms and values, deficit thinking can become part of the organizational culture of schools and create low expectations, dysfunctional relationships, and a lack of responsibility.

Central publications on successful SSDCs point out that SSDCs that are able to help their students attain educational goals and be successful are characterized by a success-oriented vision, positive school climate, teacher collaboration, a focus on teaching and learning, strong attention to social justice, equity and inclusive practices, an improved physical infrastructure of the school, clear rules, and leadership that is focused on these aspects (e.g. Capper/Young 2014; Khalifa 2018; Klein 2018a). Studies on school turnaround point out that in order to improve, schools need clear signaling that change is needed, the use of data, an ability to engage in improvement processes, engagement of all people involved, systematic professional development, school autonomy, and support for students (Bryk et al. 2015; Herman 2012).

Principals assume a mediating position between the individual school and the system level, and have an impact on teachers and parents (Böse et al. 2018b). Thus, the acceptance of reform measures by principals is of central

importance and provides a foundation for their efforts to build a success-oriented vision (Böse et al. 2018a, 2019).

In addition to the importance of a vision, clear goals and sensible organizational structures, Hemmings (2012) points out that the core problem of dysfunctional schools, especially those schools whose biographies contain experiences of “failure,” is the *school culture*. Often, low-performing schools are not only characterized by a lack of vision and dysfunctional structures, but also by “widespread resentment, disrespect, apathy, and a pervasive inability [...] to solve problems together” (Hemmings 2012: 200). Hemmings therefore suggests that strategies of “re-envisioning” and “restructuring” should be accompanied by a “re-culturation” and “re-moralization” of the school, meaning that schools must identify and address deficit thinking and create a culture that enables all participants to act ethically, assume responsibility, identify with the school, and support each other.

3.2 Improvement of SSDCs

An essential feature of effective leadership of SSDCs is the ability to lead, advocate for, and implement a mission, vision and strategic plan that focuses on social justice, equity and inclusive practices, and on nurturing the potential and abilities of the students rather than remedying their “deficits” (e.g., Khalifa 2018; Klein 2018a), and supports school effectiveness and continuous school improvement (e.g., Robinson/Lloyd/Rowe 2008; Young/Anderson/Nash 2017). Research indicates that this vision should be developed collaboratively with key stakeholders (Penuel et al. 2010), and should be informed by data (Halverson 2010). It is important that the school leader ensures the school’s mission, vision, and goals are aligned with a set of core values which emphasize important aspects of the school’s culture such as equity, social justice, inclusiveness, community, responsibility, and trust (Capper/Young 2014).

For educators, both teachers and leaders, to be able to adopt the goal and core value of social justice, they often first need to “learn” that their own behavior as well as systemic organizational dimensions have a significant effect on the learning and performance outcomes of their students, and that many students often have potential that educators may not be aware of (Drucks et al. 2020; Khalifa 2018). In a literature review for the Wübben Foundation, Klein (2018a) summarizes that principals must offer their school staff members opportunities to learn about institutional and organizational structures that support the reproduction of social inequalities, question their own presumptions, and reflect about their own deficit thinking. In a study from Germany, Drucks et al. (2020) describe how a school was able to address its deficit thinking using data on the students’ cognitive abilities,

which were significantly higher than the teachers had expected them to be. There also are various examples from the USA where principals and district leaders used data from schools in a similar situation to illustrate that students from disadvantaged backgrounds can be very successful (Doyle/Thomas/Childress 2009; Klein 2018b).

Research from a study in California points out that principals were usually more successful when they were visible in classrooms, providing professional development, but also took a strong stand and clarified that excuses would not be acceptable (Klein 2016, 2018b). Given the high level of autonomy of teachers and the largely egalitarian staff in German schools, Klein (2018a) points out that it is doubtful that teachers would accept such a strong position of the principal; instead, principals would have to include the teachers in all decisions regarding their work, and studies suggest that principals must exert more participation-oriented leadership in Germany (Racherbäumer et al. 2013), at least initially.

Another characteristic of successful principals in SSDCs was leadership that focused on building commonly accepted and sustainable organizational structures that fostered equity and collectivity, and allowed teachers to collegially improve their skills and competences. In the literature review for the Wübben Foundation, Klein (2018a) points out that successful schools succeeded in doing so even under unfavorable conditions (e.g., a lack of resources, poor school climate; Ylimaki and Jacobson 2011). Research indicates that school leaders must be able to lead change by working with staff and school community to implement and evaluate a continuous, responsive, sustainable school improvement process focused on improving learning opportunities (Duke/Salmonowicz 2010; Klar/Brewer 2013). The improvement process should be done collaboratively, as demonstrated by Huggins, Scheurich and Morgan (2011), who reported on principals bringing teachers together in activities of mutual classroom visits, mentoring and tandem structures, as well as general collaboration structures. This involves not only changing the structures, but also *promoting* collaboration among teachers (Huggins et al. 2011; Klein 2018b) and effective two-way communication (Young et al. 2017).

Another important characteristic of more successful SSDCs is that they often have a data-rich environment that helps them refocus their goals and strategies. However, when the use of performance data is not accompanied, teachers might interpret the data as proof of the low skills of students, and thus reinforce their deficit thinking instead of encouraging them to reflect on their own practice (e.g., Jimerson 2014). Thus, successful principals modelled effective behavior with regard to data use and helped their teachers and other staff members focus on student learning rather than performance, determining teaching goals and developing skills (Park 2018). Research from Germany shows that principals generally have a less central position when it comes to

data use (Muslic 2017; Kronsfoth et al. 2018), even though recent findings from a German research project emphasize how important leadership is for dealing with data especially in SSDCs (Drucks et al. 2020).

While there are a variety of studies that focus on how principals can change the goals and structures in schools, there are very few studies that look at how schools can be re-cultured and re-moralized. This is particularly remarkable, because there are several studies that show how important appreciative and trusting relationships are for school collaboration and school improvement (e.g., Bryk/Schneider 2003; Tschannen-Moran 2009). In addition to promoting professional learning for teachers, principals must create a professional environment that empowers teachers and other school staff members with collective responsibility for working collaboratively to achieve the school's shared vision (e.g., Robinson et al. 2008; Tschannen-Moran 2009).

Klein (2018a) summarizes that several studies from the USA point out how principals who had successfully changed their school first created a positive working and learning climate with clear rules and structures that allowed teachers to focus on their teaching. The studies indicate that principals placed priority on taking other burdens off their teachers and took great care to be visible in their schools (Jacobson et al. 2007), built positive relationships with the students, and made a point in recognizing their lived-in world and experiences (Khalifa 2018; Klein 2016). In Germany, Steins (2016) points out that negative relationships between teachers and students are often reinforced by the behavior of the teachers. Therefore, principals and teachers in successful SSDCs in Germany, too, placed an emphasis on developing positive relationships with their students (Racherbäumer/van Ackeren 2014).

Moreover, research carried out by Louis and Murphy (2017) showed that when teachers felt that they were working in a caring environment, they also illustrated more improvement activities; other studies indicate that empowering teachers seems to be an important prerequisite for teachers to take responsibility and be prepared for the hard work of school improvement, whereas a lack of caring and empowering leadership can lead to adversity and isolation (Klein/Bremm 2019).

4. Discussion – What we know, what we need to know more about

Effective school leaders are critical to school improvement, particularly in SSDCs. With the introduction of improved research designs and statistical methods, a growing body of empirical evidence demonstrates that principals

have an important impact on schools, teachers and student learning. In this chapter, we have examined research from the USA and Germany focused on leadership in SSDCs. SSDCs represent unique contexts within both countries. While government entities in the USA gather extensive data on SSDCs, allowing researchers, principals and policy leaders to track student, teacher and school performance, there is a growing trend within German states to develop strategies to support SSDCs, even when most states do not have accurate data on their schools' student composition, and only limited data on their performance.

Regardless of the available data on SSDCs, there is a growing body of research from which implications for effective practice can be drawn (Young/Mawhinney 2012). As discussed above, there are certain beliefs and practices that are unique to the American and German settings, particularly with regard to the level of authority that principals have compared to their teaching staff. However, these two contexts appear to have more in common than one might expect.

In both contexts, SSDCs serve a large portion of low income students and students representing diversity in terms of race, ethnicity, immigrant status and language, factors which must be taken into consideration when determining how best to support the learning and achievement of their particular student population. The research we summarized in this paper indicates that principals of SSDCs must be able to lead change by working with staff and school community members to implement and evaluate a continuous, responsive, sustainable school improvement process focused on improving student learning. Furthermore, this work must be done collaboratively, with significant attention dedicated to developing a safe, caring, inclusive and responsive school culture that embraces the belief that all students can learn at high levels. Finally, in order for principals of SSDCs to ensure equity, they must support the ability of teachers and other staff members to recognize, respect and employ students' strengths, diversity and culture as assets for teaching and learning; to recognize and redress biases, marginalization, and deficit-based thinking; and monitor and address individual and institutional biases to ensure each student and adult is treated fairly, respectfully, in a responsive manner. This support towards teachers should be evidence-based, which inherently means that teachers must be able to read, understand and accept data that can facilitate sensemaking processes.

Although more research is needed in both countries in order to guide effective leadership practice in SSDCs, even the limited research has begun to paint a picture of effective practice as reflective, collaborative and equity-focused. We would recommend that scholars in both countries continue to examine the practices associated with effective leadership of SDCCs. It will be important in both countries to conduct mixed methods research comparing practices to a variety of outcome measures over time. Comparative work that

takes into consideration the unique histories and cultures of the USA and Germany would be particularly useful for identifying practices that work across contexts versus those that are unique.

References

- Böse, Susanne/Neumann, Marko/Gesswein, Therese/Maaz, Kai (2017): Das Berliner Bonus-Programm zur Förderung von Schulen in schwieriger Lage – Eckpunkte des Programms und erste Ergebnisse der BONUS-Studie. In: Manitiuss, Veronika/Dobbelstein, Peter (eds.): Beiträge zur Schulentwicklung. Schulentwicklungsarbeit in herausfordernden Lagen. Münster: Waxmann, pp. 179-203.
- Böse, Susanne/Neumann, Marko/Gesswein, Therese/Lee, Eunji/Brauckmann, Stefan/Maaz, Kai (2018a): Supporting Schools in Challenging Circumstances in Germany: The Berlin Bonus-Program. In: Meyers, Coby V./Darwin, Marlene J. (eds.): International Perspectives on Leading Low-Performing Schools. Charlotte, NC: Information Age Publishing, pp. 189-216.
- Böse, Susanne/Neumann, Marko/Becker, Michael/Maaz, Kai/Baumert, Jürgen (2018b): Kooperationsbereit oder Innovationsgegner? In: Zeitschrift für Erziehungswissenschaft, 21, 6, pp. 1157-1186.
- Böse, Susanne/Neumann, Marko/Becker, Michael/Maaz, Kai/Baumert, Jürgen (2019): Empirische Arbeit: Bereit für Veränderungen? Zum Zusammenhang von Akzeptanz und handlungsbezogener Auseinandersetzung von Schulleiterinnen und Schulleitern in Reformprozessen. In: Psychologie in Erziehung und Unterricht, 66, 0, pp. 51-68.
- Brewer, Dominic/Smith, Joanna (2006): Evaluating the “Crazy Quilt”: Educational Governance in California. Stanford, CA: Institute for Research on Education Policy & Practice.
- Briffault, Richard (2005): The Local School District in American Law. In: Howell, William G. (ed.): Besieged: School Boards and the Future of Education Politics. Washington, DC: Brookings Institution Press, pp. 24-55.
- Brüsemeister, Thomas (2012): Von der bürokratischen Schulverwaltung zum Bildungsmanagement? In: Geiss, Michael/De Vincenti, Andrea (eds.): Verwaltete Schule. Geschichte und Gegenwart. Wiesbaden: VS-Verlag für Sozialwissenschaften, pp. 181-206.
- Bryk, Anthony S./Gomez, Louis M./Grunow, Alicia/LeMahieu, Paul G. (2015): Learning to improve: How America’s schools can get better at getting better. Cambridge, MA: Harvard Education Press.
- Bryk, Anthony S./Schneider, Barbara (2003): Trust in Schools: A Core Resource for School Reform. In: Educational Leadership, 60, 6, pp. 40-45.
- Capper, Colleen A./Young, Michelle D. (2014): Ironies and limitations of educational leadership for social justice: A call to social justice educators. In: Theory into Practice, 53, 2, pp. 158-164.

- Dedering, Kathrin/Müller, Sabine (2011): School Improvement Through Inspections? First Empirical Insights From Germany. In: *Journal of Educational Change*, 12, 3, pp. 301-322.
- Doyle, Dennis/Thomas, David A./Childress, Stacey M. (2009): *Leading for equity: The pursuit of excellence in Montgomery County Public Schools*. Cambridge, MA: Harvard Education Press.
- Drucks, Stephan/Bremm, Nina/van Ackeren, Isabell/Klein, Esther Dominique (forthc.): Recognizing the Strengths of 'Failing Schools' – An Evidence-based Way to Sustainable Change? In: Meyers, Coby V./Darwin, Marlene J. (eds.): *School Turnaround in Secondary Schools. Possibilities, Complexities, & Sustainability*. Charlotte, NC: Information Age Publishing.
- Duke, Daniel/Salmonowicz, Michael (2010): Key decisions of a first-year 'turnaround' principal. In: *Educational Management Administration & Leadership*, 38, 1, pp. 33-58.
- Elmore, Richard F./Anrig, Gregory R. (2012): Foreword. In: Roberts, John E. (ed.): *Instructional Rounds in Action*. Cambridge, MA: Harvard Education Press, pp. ix-xi.
- Gomolla, Mechtild/Radtke, Frank-Olaf (2002): *Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule*. Wiesbaden: VS-Verlag für Sozialwissenschaften.
- Halverson, Richard (2010): School formative feedback systems. In: *Peabody Journal of Education*, 85, 2, pp. 130-146.
- Hemmings, Annette (2012): Four Rs for Urban High School Reform. Re-envisioning, Reculturation, Restructuring, and Remoralization. In: *Improving Schools*, 15, 3, pp. 198-210.
- Herman, Rebecca (2012): Scaling School Turnaround. In: *Journal of Education for Students Placed at Risk (JESPAR)*, 17, 1-2, pp. 25-33.
- Huggins, Kristin Shawn/Scheurich, James Joseph/Morgan, James R. (2011): Professional Learning Communities as a Leadership Strategy to Drive Math Success in an Urban High School Serving Diverse, Low-Income Students. A Case Study. In: *Journal of Education for Students Placed at Risk (JESPAR)*, 16, 2, pp. 67-88.
- Jacobson, Stephen L. (2011): Leadership Effects on Student Achievement and Sustained School Success. In: *International Journal of Educational Management*, 25, 1, pp. 33-44.
- Jacobson, Stephen L./Brooks, Sharon/Giles, Corrie/Johnson, Lauri/Ylimaki, Rose (2007): Successful Leadership in Three High-Poverty Urban Elementary Schools. In: *Leadership and Policy in Schools*, 6, 4, pp. 291-317.
- Jacobson, Stephen L./Johnson, Lauri/Ylimaki, Rose/Giles, Corrie (2005): Successful Leadership in Challenging US Schools: Enabling Principles, Enabling Schools. In: *Journal of Educational Administration*, 43, 6, pp. 607-618.
- Jann, Werner (2005): Neues Steuerungsmodell. In: Blanke, Bernhard/von Bandemer, Stephan/Nullmeier, Frank/Wewer, Göttrik (ed.): *Handbuch zur Verwaltungsreform*. Wiesbaden: VS, pp. 74-84.
- Jimerson, Jo Beth (2014): Thinking about data. Exploring the development of mental models for "data use" among teachers and school leaders. In: *Studies in Educational Evaluation*, 42, pp. 5-14.

- Khalifa, Muhammad (2018): *Culturally Responsive School Leadership*. Cambridge, MA: Harvard Education Press.
- Klar, Hans W./Brewer, Curtis A. (2013): Successful Leadership in High-Needs Schools. An Examination of Core Leadership Practices Enacted in Challenging Contexts. In: *Educational Administration Quarterly*, 49, 5, pp. 768-808.
- Klein, Esther Dominique (2016): Instructional Leadership in den USA. Ein Modell für Schulleiterinnen und Schulleiter in Deutschland? In: *Tertium Comparationis*, 22, 2, pp. 203-229.
- Klein, Ester Dominique (2017): Bedingungen und Formen erfolgreicher Schulentwicklung in Schulen in sozial deprivierter Lage. Eine Expertise im Auftrag der Wübben Stiftung. SHIP Working Paper Reihe, No. 01. Essen. doi: 10.17185/dupublico/44384
- Klein, Esther Dominique (2018a): Erfolgreiches Schulleitungshandeln an Schulen in sozial deprivierter Lage. Eine Zusammenschau zentraler Grundlagen und Befunde aus der nationalen und internationalen Bildungsforschung. Expertise im Auftrag der Wübben Stiftung. SHIP Working Paper Reihe, No. 02. Essen. doi: 10.17185/dupublico/45206.
- Klein, Esther Dominique (2018b): Transformationale Führung und Daten in Schulen in sozial deprivierter Lage. In: *Die Deutsche Schule*, 110, 1, pp. 27-46.
- Klein, Esther Dominique/Bremm, Nina (2019): "It's almost as if I treat the teachers as I want them to treat the students". Caring als Element von Führung an Schulen in sozial deprivierter Lage. In: *Zeitschrift für Bildungsforschung*, 9, 1, pp. 89-108.
- Klein, Esther Dominique/Bremm, Nina (eds.). (2020): *Unterstützung – Kooperation – Kontrolle. Zum Verhältnis von Schulaufsicht und Schulleitung in der Schulentwicklung*. Wiesbaden: Springer VS.
- Klein, Esther Dominique/Tulowitzki, Pierre (2020): Die Fortbildung von Schulleiter*innen in Forschung und Praxis – ein Systematisierungsversuch. *Die Deutsche Schule*, 112, 3, pp. 257-276.
- Kronsfoth, Katharina/Muslic, Barbara/Graf, Tanja/Kuper, Harm (2018): Der Zusammenhang zwischen Führungsdimensionen in der Schulleitung und der Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten. In: *Die Deutsche Schule*, 110, 1, pp. 47-64.
- Ladd, Helen F. (2011): Teachers' Perceptions of Their Working Conditions. In: *Educational Evaluation and Policy Analysis*, 33, 2, pp. 235-261.
- Leithwood, Kenneth/Louis, Karen Seashore/Anderson, Stephen/Wahlstrom, Kyla (2004): *How Leadership Influences Student Learning*. Ontario: Center for Applied Research and Educational Improvement & Ontario Institute for Studies in Education.
- Louis, Karen Seashore/Murphy, Joseph (2017): Trust, caring and organizational learning: the leader's role. In: *Journal of Educational Administration*, 55, 1, pp. 103-126.
- Luhmann, Niklas/Schorr, Karl Eberhard (1982): Das Technologiedefizit der Erziehung und die Pädagogik. In: Luhmann, Niklas/Schorr, Karl Eberhard (eds.): *Zwischen Technologie und Selbstreferenz. Fragen an die Pädagogik*. Frankfurt am Main: Suhrkamp, pp. 11-40.
- Marzano, Robert J./Frontier, Tony/Livingston, David (2011): *Effective Supervision. Supporting the Art and Science of Teaching*. Alexandria, VA: ASCD.

- Mintrop, Rick (2015): Public Management Reform Without Managers. The Case of German Public Schools. In: *International Journal of Educational Management*, 29, 6, pp. 790-795.
- Mintrop, Rick/Klein, Esther Dominique (2017): Schulentwicklung in den USA. Nützliches Lehrstück für die deutsche Praxis? In: Manitiuis, Veronika/Dobbelstein, Peter (eds.): *Schulentwicklungsarbeit in herausfordernden Lagen*. Münster: Waxmann, pp. 63-83.
- Muslic, Barbara (2017): *Kopplungen und Entscheidungen in der Organisation Schule. Organisationsbezogenes Schulleitungshandeln im Kontext von Lernstandserhebungen*. Wiesbaden: Springer VS.
- National Policy Board for Educational Administration (2015): *Professional Standards for Educational Leaders 2015*. Reston, VA.
- Nelson, Sarah W./Guerra, Patricia L. (2014): Educator Beliefs and Cultural Knowledge. Implications for School Improvement Efforts. In: *Educational Administration Quarterly*, 50, 1, pp. 67-95.
- Park, Vicki (2018): Leading Data Conversation Moves. Toward Data-Informed Leadership for Equity and Learning. In: *Educational Administration Quarterly*, 54, 4, pp. 617-647.
- Pashiardis, Petros/Brauckmann, Stefan/Kafa, Antonios (2018): Leading low performing schools in Cyprus: finding pathways through internal and external challenges. In: *Leading and Managing*, 24, 2, pp. 14-27.
- Penuel, William R./Riel, Margaret/Joshi, Aasha/Pearlman, Leslie/Kim, Chong Min/Frank, Kenneth A. (2010): The alignment the informal and formal organizational supports for reform: Implications for improving teaching in schools. In: *Educational Administration Quarterly*, 46, 1, pp. 57-95.
- Pfaff, Nicole/Fölker, Laura/Hertel, Thorsten (2015): Schule als Gegenraum zum Quartier. Sozialräumliche Segregation und die Prekarität grundlegender schulischer Bedingungen. In: Fölker, Laura/Hertel, Thorsten/Pfaff, Nicolle (eds.): *Brennpunkt(-)Schule. Zum Verhältnis von Schule, Bildung und urbaner Segregation*. Opladen: Verlag Barbara Budrich, pp. 67-84.
- Pollack, Terry M. (2012): The Miseducation of a Beginning Teacher. One Educator's Critical Reflections on the Functions and Power of Deficit Narratives. In: *Multicultural Perspectives*, 14, 2, pp. 93-98.
- Potter, David/Reynolds, David/Chapman, Christopher (2002): School Improvement for Schools Facing Challenging Circumstances. A Review of Research and Practice. In: *School Leadership & Management*, 22, 3, pp. 243-256.
- Racherbäumer, Kathrin/Funke, Christina/van Ackeren, Isabell/Clausen, Marten (2013): Datennutzung und Schulleitungshandeln an Schulen in weniger begünstigter Lage. Empirische Befunde zu ausgewählten Aspekten der Qualitätsentwicklung. In: van Ackeren, Isabell/Heinrich, Martin/Thiel, Felicitas (eds.): *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund*. Münster: Waxmann, pp. 225-253.
- Racherbäumer, Kathrin/van Ackeren, Isabell (2014): Lehrer-Schüler-Beziehungen als Qualitätskriterium erwartungswidrig guter Schulen in schwieriger Lage. In: Tillack, Carina/Fetzer, Janina/Raufelder, Diana (eds.): *Beziehungen in Schule und Unterricht. Teil 3: Soziale Beziehungen im Kontext von Motivation und Leistung*. Prolog-Verlag, pp. 50-69.

- RAND Corporation (2019): Disadvantaged students.
<https://www.rand.org/topics/disadvantaged-students.html>
- Robinson, Vivianne M.J./Lloyd, Claire A./Rowe, Kenneth J. (2008): The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. In: *Educational Administration Quarterly*, 44, 5, pp. 635-674.
- Rürup, Matthias (2007): Innovationswege im deutschen Bildungssystem. Die Verbreitung der Idee „Schulautonomie“ im Ländervergleich. Wiesbaden: VS.
- Sebastian, James/Camburn, Eric M./Spillane, James P. (2018): Portraits of Principal Practice. Time Allocation and School Principal Work. In: *Educational Administration Quarterly*, 54, 1, pp. 47-84.
- Senatsverwaltung für Bildung, Jugend und Wissenschaft (2013): Bonus-Programm: Unterstützung für Schulen in schwieriger Lage. Berlin.
- Steins, Gisela (2016): Classroom Management an Schulen in sozialräumlich deprivierter Lage unter besonderer Berücksichtigung des Lehrer-Schüler-Verhältnisses. In: *Die Deutsche Schule*, 108, 4, pp. 340-353.
- Tschannen-Moran, Megan (2009): Fostering Teacher Professionalism in Schools. The Role of Leadership Orientation and Trust. In: *Educational Administration Quarterly*, 55, 1, pp. 217-247.
- Tyack, David B. (1974): *The One Best System. A History of American Urban Education*. Cambridge, MA: Harvard University Press.
- Valencia, Richard R. (2010): *Dismantling Contemporary Deficit Thinking. Educational Thought and Practice*. New York: Routledge.
- Weishaupt, Horst (2016): Schulen in schwieriger Lage und Schulfinanzierung. *Die Deutsche Schule*, 108, 4, pp. 354-369.
- Wiesner, Christian/George, Ann Cathrice/Kemethofer, David/Schratz, Michael (2015): School leadership in German speaking countries with an emphasis on Austria: a re-vision. In: *RICERCAZIONE*, 7, 2, pp. 65-90.
- Ylimaki, Rose/Jacobson, Stephen L. (2011): *US and Cross-national Policies, Practices, and Preparation. Implications for Successful Instructional Leadership, Organizational Learning, and Culturally Responsive Practices*. Dordrecht: Springer.
- Young, Michelle D./Anderson, Erin/Nash, Angel Miles (2017): Preparing school leaders: Standards-based curriculum in the United States. In: *Journal of Leadership and Policy in Schools*, 16, 2, pp. 228-271.
- Young, Michelle D./Mawhinney, Hanne (eds.) (2012): *The Research Base Supporting the ELCC Standards*. Charlottesville, VA: UCEA.

Leadership for Learning in Germany and the US: Commonalities and Differences

Pierre Tulowitzki¹, Marcus Pietsch² and James Spillane³

1. Leadership for learning as an integrated model: An introduction

Over the last two decades, *leadership for learning* (LFL) has emerged as a concept that integrates various educational leadership theories and concepts into a more comprehensive theoretical model, i.e. instructional leadership, transformational leadership and shared leadership (Daniëls/Hondeghem/Dochy 2019; Hallinger 2011; Townsend/MacBeath 2011). Although the model encompasses a variety of assumptions and practices, at its core it can be viewed as a set of principles woven around the notion that every member of a school's staff should have a stake in creating optimal conditions for learning, and that the role of a formal educational leader in this context is to provide school-wide, learning-focused leadership (MacBeath/Dempster 2009). An underlying understanding is that principals become effective (mostly) indirectly and that leadership behavior as well as its connections to learning and its antecedents are shaped by a school's context and culture (Goldring/Porter/Murphy/Elliott/Cravens 2009; Hallinger 2011; Murphy/Neumerski/Goldring/Grissom/Porter 2016).

Thus, LFL is understood as a process where whole school communities actively engage in purposeful and effective interactions that nurture relationships focused on improving (interconnected) learning on all levels of a school (Day 2011): the organizational learning, the professional learning of employees and the individual learning of students.

There is a large overlap between instructional leadership and leadership for learning, as both concepts emphasize the relevance of leading and supervising the instructional and curricular program of a school, defining a school's mission and promoting a positive school learning climate (Boyce/Bowers 2018a). But while the ultimate goal within both leadership

-
- 1 Pierre Tulowitzki is Chair of Educational Management and School Improvement at the School of Education of the FHNW University of Applied Sciences and Arts Northwestern Switzerland. Email: pierre.tulowitzki@fhnw.ch
 - 2 Marcus Pietsch is Visiting Professor for Empirical Research in Primary Education at the Leuphana University of Lüneburg. Email: pietsch@leuphana.de
 - 3 James Spillane is Spencer T. and Ann W. Olin Professor at the Northwestern University School of Education and Social Policy. Email: j-spillane@northwestern.edu

concepts is to improve student learning, instructional leadership mainly tries to reach this goal by optimizing the instructional program, whereas leadership for learning “aims at building the academic capacity of schools as means of improving student outcomes” (Hallinger/Heck 2010: 654). The concept of leadership for learning goes beyond the idea of instructional leadership by incorporating a broader range of leadership activities to support learning and learning outcomes (Bush/Glover 2014: 556). One main characteristic of LFL is that learning-oriented principals focus on “school-wide alignment of all aspects of a school with instructional-centered leadership at its core” (Boyce/Bowers 2016: 2).

Seen from this angle, the improvement of student learning is mainly reached through interactive organizational resources that support school-wide reform work and teacher change (Cosner 2009) and through capacity building (Daniëls et al. 2019). On this account, leadership within the LFL framework is conceptualized as a dynamic process of (micro) interactions within an organizational entity by incorporating aspects of laterality (Harris 2008; Harris/Leithwood/Day/Sammons/Hopkins 2007). Laterality refers to an understanding that leadership can be shared, and thus not only happens along a vertical (usually top-down), but also a lateral path (for example, teacher to teacher). Consequently, the concept of leadership for learning is also closely related to pluralistic leadership models like shared, distributed and collaborative leadership (Denis/Langley/Sergi 2012). In contrast to the concept of instructional leadership, where leadership is usually understood to be exerted by holders of a formal position, leaders within the leadership for learning framework are understood as emergent leaders, irrespective of whether they have been appointed to an official position or not. This – at its core – can be seen as a distributed perspective.

This chapter starts by offering some conceptual notions about leadership for learning, especially regarding the contextual factors that (might) shape it. It then provides a brief overview of factors that shape leadership for learning in Germany and the US. This overview is structured along the lines of input, process and output factors.

2. Assessing leadership commonalities under a common framework

Leadership is a cultural phenomenon linked to the values and customs of a group of people (Gerstner/Day 1994). Thus, a sound framework for the assessment of leadership commonalities and differences among and between cultures must take into account specific aspects of the underlying cultural

systems. For our analysis – to describe differences and similarities between Germany and the US – we refer to the well-established frameworks of educational effectiveness research, namely the Context - Input - Process - Output model (CIPO model, Scheerens/Bosker 1997). The model groups together factors and its (simple) heuristic makes it possible to describe relationships between Inputs, Processes and Outputs in educational settings within certain contexts. It should be noted that the model is not a logic model (Astbury/Leeuw 2010) in the pure sense, as it lacks dynamic as well as reciprocal aspects, and thus does not allow to prompt unambiguous research hypotheses about mechanisms and influencing paths among the incorporated factors or categories (Kuger/Klieme/Jude/Kaplan 2016). Drawing on the four dimensions of the model, we will focus on the following aspects of LFL:

- C Contextual Conditions for Leadership for Learning (e.g. educational policy, support system)
- I Input of and for Leadership for Learning (e.g., training and recruitment of principals)
- P Process of Leadership for Learning (e.g., procedures of leading and learning)
- O Output(s) of Leadership for Learning (e.g., anticipated and realized outcomes of school leadership)

In conceptualizing leadership for learning, one critical challenge involves conceptualizing and understanding relationships between school leadership and teaching and learning. Teaching is not a simple reflex of learning; teaching and learning are distinct practices, and we need to understand how both practices not only connect with one another, but also with leadership practice. Recent work argues for attention to conceptualize leadership, teaching and learning in terms of the relationships among these practices (Spillane 2015). Scholars of human practice, working in several disciplinary traditions, argue for attention to activity systems that take into account how persons interact with one another using aspects of their environment (Engeström 2001; Engstrom 2001; Cook/Brown 1999). Teaching or leading, for example, is often conceptualized as what the teacher or leader does, roughly equivalent to a teacher's or leader's behavior. In contrast, scholars of human activity argue that practice is not about the actions of individuals but about interactions – it is about what people do together using key aspects of their situation rather than what they do on their own (Spillane 2006). Hence, the challenge to understanding relationships among leadership and learning fundamentally concerns understanding relationships among leading practice, teaching practice, and learning practice (Spillane 2015).

3. Contextual Conditions for Leadership for Learning

3.1 Germany

With regard to the relationship between leadership and context, there's hardly any German research that makes use of quantitative designs. However, the existing findings support the nowadays common wisdom that context matters. For example, Schwarz & Brauckmann (2015) drew upon survey data to show that the area close to schools (ACTS) influences among other things school principals' perceptions of student-related challenges at school, workload and what is done during the work time.

Furthermore, Pietsch and Leist (2018) demonstrated that competition between schools (to attract students) has a major impact on the LFL behavior of principals in German secondary schools: the stronger the competition between schools, the more pronounced the leadership activities of principals. The nature of school leadership varies directly with the level of competition, even when controlled for other potential contextual confounding variables such as the socioeconomic status of students' families and school organization factors. What was striking was that all facets of LFL, i.e. instructional, transformational and shared leadership, were positively associated with competition. Thus, the LFL climate of a school as indicated by a principal's leadership behavior directly reflects a school's competitive context, in that principals seem to react to the (perceived) competitive pressure by adapting their leadership style accordingly. In contrast to American findings, the social context of schools does not seem to have an impact on the leadership behavior of principals in Germany.

3.2 USA

Advocates of the Leadership for Learning model argue "that leadership is enacted within an organizational and environmental context" (Hallinger 2011: 127), with context referring to features of the broader organizational and environmental setting within which the school and the principal are located (Hallinger 2016). From a distributed perspective context is understood not as something external to leadership and as something that influences it from the outside, but rather as something that is constitutive of leadership practice, influencing it from the inside out. Put metaphorically, context is not a stage on which individuals practice and that influences what individuals do; it defines the practice as it is the medium for practice and for interactions.

Fittingly, existing research indicates that school contextual and compositional factors may have effects on all three leadership styles incorporated

into the LFL model (Hallinger/Murphy 1986; Liu/Bellibas/Printy 2016; Smith/Bell 2011). It underscores that a school's context can influence the way the school is led and/or its priorities. In other words, the national, regional and local as well as the social and organizational context of a school can be considered to be inextricably linked to school leadership and its consequences.

4. Input of and for Leadership for Learning

4.1 Germany

Principals of public schools in Germany are usually recruited exclusively among the teaching staff. Teachers go through a master-degree level higher education qualification that ends in a state-recognized "Master of Education" (for more details, see Tulowitzki/Krüger/Roller 2018). They then have to undergo a mandatory period as teachers in training for 1-2 years in school before becoming "full" teachers. Teachers interested in becoming a principal apply for vacant positions that are in many circumstances publicly listed. The vetting process usually involves a check of an applicant's career achievements and teaching abilities. While having a teacher-type master's degree is a hard requirement, additional qualifications are often also desired; for example, experience as vice principal or having had special responsibilities in schools, or having completed a voluntary further qualification in educational management. However, the teaching competencies and teaching evaluations are often given the most weight when assessing an application. Candidates applying for a position as principal will usually have to undergo a series of interviews; the ministry of education and cultural affairs, and usually also the school where the candidate applies, get to weigh in on whether or not the position should be awarded to the candidate. In many but not all states, they are required to undergo a short course in the form of preparatory or in-service training. The training usually covers aspects of management, judicial aspects as well as aspects of quality management. Once appointed, principals are usually civil servants or on indefinite contracts, meaning they are appointed for life.

The position of the German school principal has received more attention over the last decades because as schools have gained more autonomy, the responsibilities of school-based leaders expanded accordingly (Tulowitzki 2015). Among other things, this has led to an increased need for professionalization and support.

4.2 USA

Writing about school leadership in the US is difficult because there is not one US school system. Rather, in the US there are multiple school systems – some public, some private, and some hybrid – from local school districts to charter school networks to religious based school systems. Even public school systems vary radically, depending on whether they serve urban, suburban, or rural communities. Moreover, these school systems operate in rather different government/policy environments, depending on the state (Manna 2015). The policy and government environments in which schools operate – for instance, in New York and New Mexico – are not the same. For example, some states approve curricular materials for core school subjects for use in schools, whereas other states leave such matters to local school systems. Overall, state governments have a variety of avenues through which they can leverage influence on school principals, including establishing leadership standards, influencing leadership preparation programs, principal licensure and principal evaluation. However, there remains considerable variability among states in how they deploy these policy levers in practice. And within states, there can be considerable variability on everything from principal recruitment to formal preparation and professional development.

Nevertheless, there appear to be some broad patterns about school leadership that mostly hold across state policy environments and many school systems. Principals are hired by local school system leaders (e.g., the local school district), though there are exceptions to this pattern; for example, in Chicago, where the majority of school principals are hired by the Local School Council (LSC), which is elected by members of the community served by the school. Typically, the school principal hires teachers, often with input from school staff, depending on the school system. In some school systems, system leaders can also play a role in teacher recruitment (for example in some of the Charter School Networks).

5. Process of Leadership for Learning

5.1 Germany

While the German education system has many unique features compared to other European countries (for a detailed presentation, see Döbert 2015), there are strong indications that educational leadership practices share common characteristics across the globe (see for example Leithwood/Harris/Hopkins 2008, 2019; OECD 2014). One particularity, however, is that in Germany,

principals only have little authority over teacher recruitment and appointment as well as over teacher salaries and teacher promotion. Consequently, principals hold less than 20 percent of the responsibility for resources (the OECD average is 38 percent, OECD 2016).

The formally assigned authority of principals over staff varies from federal state to state (*Land*), in many cases the teachers are free to teach as they deem appropriate (as they have what is called “freedom of teaching” and “pedagogical freedom”, see Wermke 2011: 681f). That means that principals in Germany typically are limited in terms of influence on teaching practices and pedagogical approaches used in schools. As principals in Germany work in a low-accountability context and – like teachers – are civil servants in many states, their position is rather secure (Huber/Gördel/Kilic/Tulowitzki 2016). In many schools, the principal and deputy principal additionally work with several teachers on matters of leadership and management such as organizing processes of quality management, initiating and implementing school improvement projects, forming an extended leadership team (in German *Steuergruppe*, which translates to “steering group”). Through their work on selected management or leadership issues as well as on various projects and initiatives, they have a significant influence on matters of school improvement as well as on practices of teaching staff (Feldhoff 2010; Feldhoff/Rolff 2008).

5.2 USA

Traditionally, with respect to teaching and learning the two images of the school principal in the literature were the principal as buffering teachers from external interference especially with respect to their classroom practice, causing a perpetual tension between principal’s desire to focus her/his time on improving instruction and what Larry Cuban refers to as the “managerial imperative” of the job (Cuban 1988). While managing the tension between the managerial and the instructional continues to be an issue for principals, increasingly they cannot afford to buffer teachers from external environmental pressures to improve teaching and learning (Spillane/Lowenhaupt 2019). This added pressure can cause teachers to focus their efforts on (relatively) easy-to-teach students, thus putting students who traditionally have been disenfranchised by the school system at risk.

Since the 1980s there have been dramatic shifts in the policy environment in which US schools and school systems operate, regardless of state with local, state, and federal policy makers in the USA directing their attention and policy initiatives on classroom teaching and student learning, specifying what teachers should teach, in some cases how they should teach, and acceptable levels of student achievement. Mobilizing policy instruments – in particular

rewards and sanctions – for compliance with externally imposed performance standards are sought by federal and state policy makers. As a result of the dramatic change in the institutional environment of US schools over the last 25 years, curriculum standards and test-based accountability have become staples. Moreover, requirements to report student achievement data by different subpopulations of students (e.g., race, class) has foreground tremendous inequities in students' opportunities to learn. As the pressure on school leaders and teachers to improve the quality of teaching and learning from beyond the schoolhouse has increased, principals can no longer buffer teachers from external initiatives intended to draw attention to teaching and learning.

Policy makers are not the only ones pressing for school leaders to pay attention to teaching and learning. Extra-system agents and agencies such as philanthropic institutions, university preparation programs and national associations have also played a prominent role, often with government support and incentives, in transforming the American education sector. One such effort is the Interstate School Leaders Licensure Consortium (ISLLC) standards, recently revised and renamed as the Practice Standards for Educational Leaders (PSEL), that lay out expectations for school and district leaders regarding practice (Young/Crow/Murphy/Ogawa 2009; Young/Mawhinney/Reed 2016). Designed primarily as a foundation for thinking about leadership practice, the PSEL standards have also been influential in leadership development work. Based on a review of the empirical literature and the educational landscape together with input from researchers and practitioners, the standards are intended to guide the practice of educational leaders by identifying the nature of the work and defining what counts as quality work. Teaching and learning and its improvement figures prominently in these standards for leadership practice. Furthermore, recent work reports that all 50 states in the United States have either adopted or adapted the ISLLC standards (Anderson/Reynolds 2015).

These shifts in the institutional environment of America's schools represent a considerable departure from business as usual for teaching and learning in schools, and for leadership in particular. For example, supporting instruction, leading instructional improvement and monitoring the quality of instruction are increasingly central to the work of school leadership. While the tension between the managerial and the instructional persists, improving teaching and learning are integral to the work of the school principal and educational leadership more broadly.

6. Output(s) of Leadership for Learning

6.1 *Germany*

German research explicitly based on LFL is virtually non-existent. To the best of our knowledge, only a handful of studies exist (Ammann 2018; Pietsch/Leist 2018; Pietsch/Tulowitzki/Koch 2018). Studies considering effectiveness criteria, that is student learning and achievement gains of students, as outcomes measures are – with only one exception (Pietsch/Lücken/ Thonke/Klitsche/Musekam 2016) – not available. Regarding the scarce empirical knowledge base from Germany, Pietsch, Tulowitzki and Koch (2018) explored multilevel associations of LFL, teachers' job satisfaction and organizational commitment, drawing on survey data from the school inspection of the German federal state of Hamburg. Their findings indicated that shared leadership is a strong predictor of individual and shared job satisfaction as well as organizational commitment of teachers' job satisfaction and organizational commitment, and that LFL is contextually bound. The social background of a school's student population had a statistically significant impact on teachers' organizational commitment and job satisfaction at the school level. Teachers who worked in schools with a higher amount of socially privileged students were more strongly committed to their schools and more satisfied with their jobs than their colleagues who work at schools in challenging social circumstances. Additionally, results indicated that the association of an instructional leadership culture and the shared organizational commitment and shared job satisfaction of teachers varied with the social and structural context of a school in its entirety. Thus, with regard to the structural and social contexts of a school, the study also showed that instructional management and its relation to the shared job satisfaction and shared organizational commitment of teachers seem to be contextually contingent.

Using teacher survey data from the federal state of Hamburg, Pietsch et al. (2016; 2017) also investigated the direct and indirect ties between various leadership styles, namely, instructional, transformational, transactional, and laissez-faire leadership, and the instructional practices of teachers by applying a structural equation model. Results revealed that mediating variables – e.g. organizational commitment, and motivation of teachers, capacity (beliefs) of teachers and working conditions of teachers – are influenced by a leadership core as well as by all leadership facets, and that the leadership behavior varied systematically with a schools' achievement context.

In addition to these studies, which explicitly focus on LFL in its totality, there exists research from Germany into educational leadership that covers individual facets of LFL, though again the evidence base is sparse. There is

very little research looking into instructional leadership in Germany (Brauckmann/Geissler/Feldhoff/Pashiardis 2016; Klein 2016). Similarly, only a small number of studies dealing with practices akin to shared leadership and transformational leadership have been produced. For example, Schaarschmidt and colleagues found that a participatory and supportive leadership style led to more intact interpersonal relationships among staff, and acted as a buffer for stressors of the day-to-day work (Schaarschmidt/Kieschke 2013: 93). Similarly, a study conducted in North-Rhine Westphalia, one of the most populated federal states in Germany, found evidence that transformational leadership, participation (in other words, sharing of tasks and responsibilities) as well as the work climate in schools correlate highly with the affective commitment of teachers (Harazd/Gieske/Gerick 2012). Findings from a mixed-methods study (Gieske 2013), also conducted in North-Rhine Westphalia, echo this: Data indicated that teaching staff had a stronger organizational commitment in schools that were led by what Gieske dubbed “rational school principals”. These were principals who tried to lead by presenting issues in a transparent manner, winning staff over through arguments and tried to involve staff in the decision-making process (Gieske 2013: 131ff). None of those studies focused on linking leadership to student achievement.

6.2 USA

In the US, there is a relatively long history of efforts to document relations between aspects of what we refer to as leadership for learning and school outcomes, dating back to at least the beginning of school effectiveness research. Research on school effectiveness, starting with work by Lezotte and Brookover in the 1970s, documented how schools can organize to create conditions necessary to improve teaching and student learning. Among other things, scholars working in this tradition (see Purkey/Smith 1983; Lezotte 2001; Brookover/Lezotte 1977) have identified conditions that characterize effective schools as measured in terms of student outcomes including:

- Strong school leadership focused on quality instruction
- High expectations for students
- Planned curriculum coordination and organization
- Linking professional development to the expressed needs of the staff
- Clear and focused mission
- An orderly and safe school environment
- Frequent monitoring of student progress as basis for improvement
- Positive home-school relations.

Though work in this tradition has been critiqued methodologically, it had a strong influence on the field and subsequent scholarship.

In the 1980s, research in the ‘instructional leadership’ tradition identified both the roles and functions of instructional leaders, including defining and communicating a clear mission for instruction, managing a program for instruction by coordinating curriculum and supervising teaching and students’ progress, recognizing achievement, and nurturing a positive learning climate for both children and adults in schools (Hallinger/Murphy 1985; Hallinger 2009; Heck et al. 1990, 1991; Marks/Printy 2003). A major meta-analysis of research on school leadership involving 27 research studies (two thirds of which were conducted in the US) focused on relationships between school leadership and student outcomes. The meta-analysis shows that the closer school leaders’ work is to teaching and learning, the more likely they are to have a positive influence on student outcomes (Robinson/Lloyd/Rowe 2008).

Over the past quarter century a large number of studies dealing with facets of leadership for learning have been undertaken (Boyce/Bowers 2018b; Daniëls et al. 2019; Hallinger 2011). Successful principals in this context are seen as value-driven, cooperation-oriented, aiming at building the school's capacity for improvement, sharing and empowering leadership where appropriate, and then developing suitable strategies only after having understood the context (Hallinger 2011: 137-138). Particularly, a large body of quantitative empirical LFL research is based upon data from the School And Staffing Survey (SASS, Boyce/Bowers 2018b), which (together with its successor the National Teacher and Principal Survey) is the largest, most comprehensive survey of schools and school staff, which provides descriptive data on the context of elementary and secondary education on a wide range of topics. Within the SASS LFL is conceptualized assuming that

teacher autonomy and influence and principal leadership serve as the foundation of instructional leadership with a reciprocal relationship between them, adult development is affected by teacher autonomy and influence, and all of these three factors contribute to school climate, which in turn acts as a significant bridge between instructional leadership and the three emergent factors. [...] teacher satisfaction, teacher commitment, and teacher retention. (Boyce/Bowers 2018b: 171)

Taking advantage of longitudinal administrative data, several recent studies show reasonably large ‘principal effects’ on student outcomes, typically test scores (Branch/Hanushek/Rivkin 2012; Grissom/Kalogrides/Lobe 2015). Furthermore, several recent studies show a relationship between school leadership and both teacher retention and teacher satisfaction (Boyd et al. 2011; Grissom 2011; Ladd 2011; Sebastian/Allensworth 2012). Empirical findings indicate that effective American schools have principals who focus on curricula and instruction by shaping a schools’ climate and culture, defining and communicating missions and visions, recognizing and awarding success and accomplishments, maintaining good internal and external relations, and investing in the schools’ personnel (Daniëls et al. 2019).

Table 1. Summary based on table on relationships between instructional leadership themes and human resource factors (i.e. teacher satisfaction, commitment, retention), expanded to account for studies from Germany

Country	Number of studies	Level of evidence	Rationale
USA	42	Limited to moderate	Sufficient number of primary studies, but lack of multilevel modeling
Germany	3	Limited	Lack of primary studies

Source: Boyce/Bowers 2018b (USA); own research (Germany)

7. Discussion and Conclusion

Based on our overview, we come to the conclusion that school leadership per se and LFL in particular are far less discussed as well as empirically investigated in Germany than in the US. Furthermore, we observe that the scholarly discussion on school leadership in Germany – unlike in the US – does not seem to focus much on effectiveness, i.e. student learning and achievement gains of students. There are preliminary indications pointing to the social context of a school not being as relevant for shaping principal leadership in Germany compared to the US.

Furthermore, the dearth on studies on educational leadership, and by extension on leadership for learning in Germany, may be indicative of a key difference between the US and Germany when it comes to the professional culture in schools: Teachers in Germany are far more autonomous than their US colleagues. Possibly due to their more independent status and their extensive preparatory training, they are relatively resistant to influences of school principals on the classroom level. By that logic, principals in Germany serve more as a buffer for teachers against disruptions, and as mediators and administrative managers. American principals, by contrast, seem to have a more pronounced role in terms of influencing instructional practices, human resource management and leadership in general. It seems plausible that American principals can’t afford to buffer their teachers from external environmental pressures anymore due to the high-stakes accountability context they are operating in. While standards-based accountability plays a major role in the US but not in Germany, this can be seen as another explanation for differences in terms of educational leadership: the German low-stakes accountability system offers German teachers and principals more

room to maneuver in terms of leadership and teaching practices than their American counterparts.

Nevertheless, the empirical results suggest that LFL in both contexts share more communalities than differences. Thus, on the one hand principals on both sides of the Atlantic seem to have a strong influence on the working conditions of teachers, their professional capacities, personnel development and mediated by that on teaching practices. On the other hand, this is reached by the same means: instructional, transformational and distributed/shared leadership practices. Furthermore, there is evidence that the local context of a school shapes the behavior of principals in Germany as well as in the US – independently from the national context in which principals and schools are situated. However, while the social context plays a major role in the US regarding how and how successfully principals lead, the social context in Germany appears to have less of an influence on leadership practices and their success. Other context factors, especially those of the administrative kind, have a more pronounced influence.

Ultimately, this comparative contribution shows that international comparative research allows us to reflect on particular national situations and provide an opportunity for understanding implicit and culturally specific theories, assumptions and empirical findings concerning how school principals influence the teaching and learning within schools as well as relevant determinants, interactions and results. Furthermore, the contribution points to the fact that LFL is an under-researched topic on both sides of the Atlantic, being nearly non-existent in Germany. Nonetheless, it underscores the relevance of LFL and its viability, irrespective of any national context. It furthermore paints a picture of emergent research to be conducted in order to better understand links between practices of principals and teachers on the one hand, and students and learning on the other.

References

- Ammann, Markus (2018): Leadership for Learning as Experience: Introducing the Use of Vignettes for Research on Leadership Experiences in Schools. In: *International Journal of Qualitative Methods*, 17, 1, pp. 1-13.
<https://doi.org/10.1177/1609406918816409>
- Anderson, Erin/Reynolds, Amy L. (2015): A Policymaker's Guide – Research-Based Policy for Principal Preparation Program Approval and Licensure [Report from the University Council for Educational Administration]. Charlottesville, VA: UCEA.

- Astbury, Brad/Leeuw, Frans L. (2010): Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation. In: *American Journal of Evaluation*, 31, 3, pp. 363-381. <https://doi.org/10.1177/1098214010371972>
- Boyce, Jared/Bowers, Alex J. (2016): Different levels of leadership for learning: Investigating differences between teachers individually and collectively using multilevel factor analysis of the 2011-2012 Schools and Staffing Survey. In: *International Journal of Leadership in Education*, 21, 2, pp. 197-225. <https://doi.org/10.1080/13603124.2016.1139187>
- Boyce, Jared/Bowers, Alex J. (2018a): Different levels of leadership for learning: Investigating differences between teachers individually and collectively using multilevel factor analysis of the 2011-2012 Schools and Staffing Survey. *International Journal of Leadership in Education*, 21, 2, pp. 197-225. <https://doi.org/10.1080/13603124.2016.1139187>
- Boyce, Jared/Bowers, Alex J. (2018b): Toward an evolving conceptualization of instructional leadership as leadership for learning: Meta-narrative review of 109 quantitative studies across 25 years. In: *Journal of Educational Administration*, 56, 2. <https://doi.org/10.1108/JEA-06-2016-0064>
- Boyd, Donald/Grossman, Pam/Ing, Marsha/Lankford, Hamilton/Loeb, Susanna/Wyckoff, James (2011): The Influence of School Administrators on Teacher Retention Decisions. In: *American Educational Research Journal*, 48, 2, pp. 303-333. <https://doi.org/10.3102/0002831210380788>
- Branch, Gregory/Hanushek, Eric A./Rivkin, Steven G. (2012): Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals (No. w17803). <https://doi.org/10.3386/w17803>
- Brauckmann, Stefan/Geissler, Gert/Feldhoff, Tobias/Pashiardis, Petros (2016): Instructional Leadership in Germany: An Evolutionary Perspective. In: *International Studies in Educational Administration*, 44, 2, pp. 5-20.
- Brookover, Wilbur B./Lezotte, Lawrence W. (1979): Changes in School Characteristics Coincident With Changes in Student Achievement. Occasional Paper No. 17. In: *Occasional Papers*, 17, pp. 1-2.
- Burkhauser, Susan (2017): How Much Do School Principals Matter When It Comes to Teacher Working Conditions? In: *Educational Evaluation and Policy Analysis*, 39, 1, pp. 126-145. <https://doi.org/10.3102/0162373716668028>
- Bush, Tony/Glover, Derek (2014): School leadership models: What do we know? In: *School Leadership & Management*, pp. 1-19. <https://doi.org/10.1080/13632434.2014.928680>
- Cook, Scott D. N./Brown, John Seely (1999): Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizational Knowing. In: *Organization Science*, 10, 4, pp. 381-400. <https://doi.org/10.1287/orsc.10.4.381>
- Cosner, Shelby (2009): Building Organizational Capacity Through Trust. *Educational Administration Quarterly*. <https://doi.org/10.1177/0013161X08330502>
- Cuban, Larry (1988): *The Managerial Imperative and the Practice of Leadership in Schools*. New York: SUNY Press.
- Daniëls, Ellen/Hondeghem, Annie/Dochy, Filip (2019): A review on leadership and leadership development in educational settings. In: *Educational Research Review*, 27, pp. 110-125. <https://doi.org/10.1016/j.edurev.2019.02.003>

- Day, Christopher (2011): The Layering of Leadership. In: Robertson, Jan/Timperley, Helen (eds.): *Leadership and Learning*, pp. 13-26.
<https://doi.org/10.4135/9781446288931>
- Denis, Jean-Louis/Langley, Ann/Sergi, Viviane (2012): Leadership in the Plural. In: *The Academy of Management Annals*, 6, 1, pp. 211-283.
<https://doi.org/10.1080/19416520.2012.667612>
- Döbert, Hans (2015): Germany. In: Hörner, Wolfgang/Döbert, Hans/Reuter, Lutz/von Kopp, Bodo (eds.): *The Education Systems of Europe* 2nd edition. Cham: Springer International Publishing, pp. 305-333.
- Engeström, Yrjö (2001): Expansive Learning at Work: Toward an activity theoretical reconceptualization. In: *Journal of Education and Work*, 14, 1, pp. 133-156.
<https://doi.org/10.1080/13639080020028747>
- Feldhoff, Tobias (2010): Schulentwicklung durch Steuergruppen und Organisationales Lernen – Eine empirische Studie zum Einfluss von schulischen Steuergruppen und der Kapazität des Organisationalen Lernens auf Erfolgsfaktoren des Modellvorhabens "Selbstständige Schule" in Nordrhein-Westfalen (Unveröffentlichte Dissertation). Dortmund: Technische Universität Dortmund.
- Feldhoff, Tobias/Rolff, Hans-Günter (2008): Einfluss von Schulleitungs- und Steuergruppenhandeln. In: Holtappels, Heinz Günter/Klemm, Klaus/Rolff, Hans-Günter (eds.): *Schulentwicklung durch Gestaltungsautonomie. Ergebnisse der Begleitforschung zum Modellvorhaben "Selbstständige Schule"* in Nordrhein-Westfalen. Münster: Waxmann, pp. 293-303.
- Gerstner, Charlotte R./Day, David V. (1994): Cross-cultural comparison of leadership prototypes. In: *The Leadership Quarterly*, 5, 2, pp. 121-134.
[https://doi.org/10.1016/1048-9843\(94\)90024-8](https://doi.org/10.1016/1048-9843(94)90024-8)
- Gieske, Mario (2013): Mikropolitik und schulische Führung: Einflussstrategien von Schulleitern bei der Gestaltung organisationalen Wandels. Bad Heilbrunn: Klinkhardt.
- Goldring, Ellen/Porter, Andrew/Murphy, Joseph/Elliott, Stephen N./Cravens, Xiu (2009): Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices. In: *Leadership and Policy in Schools*, 8,1, pp. 1-36. <https://doi.org/10.1080/15700760802014951>
- Grissom, Jason A./Kalogrides, Demetra/Loeb, Susanna (2014): Using Student Test Scores to Measure Principal Performance. *Educational Evaluation and Policy Analysis*. <https://doi.org/10.3102/0162373714523831>
- Grissom, Jason A. (2011): Can Good Principals Keep Teachers in Disadvantaged Schools? Linking Principal Effectiveness to Teacher Satisfaction and Turnover in Hard-to-Staff Environments. In: *Teachers College Record*, 113, 11, pp. 2552-2585.
- Hallinger, Philip (2011): Leadership for learning: Lessons from 40 years of empirical research. In: *Journal of Educational Administration*, 49, 2, pp. 125-142.
<https://doi.org/10.1108/09578231111116699>
- Hallinger, Philip (2016): Bringing context out of the shadows of leadership. In: *Educational Management Administration & Leadership*, 46, 1, pp. 5-24.
<https://doi.org/10.1177/1741143216670652>
- Hallinger, Philip/Heck, Ronald H. (2010): Leadership for Learning: Does Collaborative Leadership Make a Difference in School Improvement? In:

- Educational Management Administration & Leadership, 38, 6, pp. 654-678.
<https://doi.org/10.1177/1741143210379060>
- Hallinger, Philip/Murphy, Joseph F. (1986): The Social Context of Effective Schools. In: American Journal of Education, 94, 3, pp. 328-355.
- Harazd, Bea/Gieske, Mario/Gerick, Julia (2012): Was fördert affektives Commitment von Lehrkräften? Eine Analyse individueller und schulischer (Bedingungs-) Faktoren. In: Zeitschrift für Bildungsforschung, 2, 2, pp. 151-168.
<https://doi.org/10.1007/s35834-012-0039-z>
- Harris, Alma (2008): Distributed leadership: According to the evidence. In: Journal of Educational Administration, 46, 2, pp. 172-188.
<https://doi.org/10.1108/09578230810863253>
- Harris, Alma/Leithwood, Kenneth/Day, Christopher/Sammons, Pam/Hopkins, David (2007): Distributed leadership and organizational change: Reviewing the evidence. In: Journal of Educational Change, 8, 4, pp. 337-347.
<https://doi.org/10.1007/s10833-007-9048-4>
- Huber, Stephan Gerhard/Gördel, Bettina-Maria/Kilic, Selin/Tulowitzki, Pierre (2016): Accountability in the German school system. In: Easley II, Jacob/Tulowitzki, Pierre (eds.): Educational Accountability – International perspectives on challenges and possibilities for school leadership. London: Routledge, pp. 165-183.
- Klein, Esther Dominique (2016): Instructional Leadership in den USA. Ein Modell für Schulleiterinnen und Schulleiter in Deutschland? In: Tertium Comparationis, 22, 2, pp. 203-229.
- Kuger, Susanne/Klieme, Eckhard/Jude, Nina/Kaplan, David (eds.) (2016): Assessing Contexts of Learning – An International Perspective.
https://doi.org/10.1007/978-3-319-45357-6_1
- Ladd, Helen F. (2011): Teachers' Perceptions of Their Working Conditions: How Predictive of Planned and Actual Teacher Movement? In: Educational Evaluation and Policy Analysis, 33, 2, pp. 235-261.
<https://doi.org/10.3102/0162373711398128>
- Leithwood, Kenneth/Harris, Alma/Hopkins, David (2008): Seven strong claims about successful school leadership. In: School Leadership & Management, 28, 1, pp. 27-42. <https://doi.org/10.1080/13632430701800060>
- Leithwood, Kenneth/Harris, Alma/Hopkins, David (2019): Seven strong claims about successful school leadership revisited. In: School Leadership & Management, 0(0), pp. 1-18. <https://doi.org/10.1080/13632434.2019.1596077>
- Lezotte, Lawrence W. (2001): Revolutionary and evolutionary: The effective schools movement. Okemos, MI: Effective Schools Products Ltd.
- Liu, Yan/Bellibas, Mehmet Sukru/Printy, Susan (2016): How school context and educator characteristics predict distributed leadership: A hierarchical structural equation model with 2013 TALIS data. In: Educational Management Administration & Leadership, 46, 3, pp. 401-423.
<https://doi.org/10.1177/1741143216665839>
- MacBeath, John/Dempster, Neil (eds.) (2009): Connecting Leadership and Learning: Principles for Practice. London: Routledge.
- Manna, Paul (2015): Developing Excellent School Principals to Advance Teaching and Learning: Considerations for State Policy. Wallace Foundation.

- Murphy, Joseph/Neumerski, Christine M./Goldring, Ellen/Grissom, Jason/Porter, Andy (2016): Bottling fog? The quest for instructional management. In: *Cambridge Journal of Education*, 46, 4, pp. 455-471.
<https://doi.org/10.1080/0305764X.2015.1064096>
- OECD (2014): TALIS 2013 Results – An International Perspective on Teaching and Learning. <https://doi.org/10.1787/23129638>
- OECD (2016): PISA 2015 Results (Volume II) – Policies and Practices for Successful Schools. <https://doi.org/10.1787/19963777>
- Pietsch, Marcus/Leist, Sebastian (2018): The effects of competition in local schooling markets on leadership for learning. In: *Zeitschrift Für Bildungsforschung*, 9, pp. 109-134. <https://doi.org/10.1007/s35834-018-0224-9>
- Pietsch, Marcus/Lücken, Markus/Thonke, Franziska/Klitsche, Stefan/Musekamp, Frank (2016): Der Zusammenhang von Schulleitungshandeln, Unterrichtsgestaltung und Lernerfolg: Eine argumentbasierte Validierung zur Interpretier- und Nutzbarkeit von Schulinspektionsergebnissen im Bereich Führung von Schulen. In: *Zeitschrift für Erziehungswissenschaft*, 19, 3, pp. 527-555. <https://doi.org/10.1007/s11618-016-0692-4>
- Pietsch, Marcus/Tulowitzki, Pierre (2017): Disentangling school leadership and its ties to instructional practices – an empirical comparison of various leadership styles. In: *School Effectiveness and School Improvement*, 28, 4, pp. 629-649. <https://doi.org/10.1080/09243453.2017.1363787>
- Pietsch, Marcus/Tulowitzki, Pierre/Koch, Tobias (2018): On the Differential and Shared Effects of Leadership for Learning on Teachers' Organizational Commitment and Job Satisfaction: A multilevel perspective. In: *Education Administration Quarterly*, 55, 5, pp. 705-741.
<https://doi.org/10.1177/0013161X18806346>
- Purkey, Stewart C./Smith, Marshall S. (1983): Effective schools: A review. In: *The Elementary School Journal*, 83, 4, pp. 427-452.
- Schaarschmidt, Uwe/Kieschke, Ulf (2013): Beanspruchungsmuster im Lehrerberuf – Ergebnisse und Schlussfolgerungen aus der Potsdamer Lehrerstudie. In: Rothland, Martin (ed.): *Belastung und Beanspruchung im Lehrerberuf: Modelle, Befunde, Interventionen*. 2nd edition. Wiesbaden: Springer VS, pp. 81-97.
- Scheerens, Jaap/Bosker, Roel (1997): *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schwarz, Alexandra/Brauckmann, Stefan (2015): Between facts and perceptions: The area close to school as a context factor in school leadership.
<https://www.econstor.eu/handle/10419/111690>
- Sebastian, James/Allensworth, Elaine (2012): The Influence of Principal Leadership on Classroom Instruction and Student Learning: A Study of Mediated Pathways to Learning. In: *Educational Administration Quarterly*, 48, 4, pp. 626-663.
<https://doi.org/10.1177/0013161X11436273>
- Smith, Phil/Bell, Les/Davies, Patricia M./Popescu, Ana-Cristina/Gunter, Helen M. (2011): Transactional and transformational leadership in schools in challenging circumstances: A policy paradox. In: *Management in Education*, 25, 2, pp. 58-61.
<https://doi.org/10.1177/0892020611399608>
- Spillane, James P. (2006): *Distributed leadership*. San Francisco: Jossey-Bass.

- Spillane, James P. (2015): Leadership and Learning: Conceptualizing Relations between School Administrative Practice and Instructional Practice. *Societies*, 5,2, pp. 277-294. <https://doi.org/10.3390/soc5020277>
- Spillane, James P./Lowenhaupt, Rebecca (2019): Navigating the Principalship: Key Insights for New and Aspiring School Leaders. Alexandria, VA: ASCD.
- Townsend, Tony/MacBeath, John (eds.) (2011): *International Handbook of Leadership for Learning*. Dordrecht: Springer Netherlands.
- Tulowitzki, Pierre (2015): The Development of Educational Leadership and Teaching Professions in Germany. In: *ECPS – Educational, Cultural and Psychological Studies*, 11, pp. 45-55. <https://doi.org/10.7358/ecps-2015-011-tulo>
- Tulowitzki, Pierre/Krüger, Michael/Roller, Marvin (2018): Education Policies, the Teaching Profession and Teacher Training in Germany – The Ever-Evolving 16-Piece Mosaic. In: Normand, Romuald/Liu, Min/Carvalho, Luis Miguel/Oliveira, Dalila Andrade/LeVasseur, Louis (eds.): *Education Policies and the Restructuring of the Educational Profession: Global and Comparative Perspectives*, pp. 71-87. https://doi.org/10.1007/978-981-10-8279-5_6
- Wermke, Wieland (2011): Continuing professional development in context: Teachers' continuing professional development culture in Germany and Sweden. In: *Professional Development in Education*, 37, 5, pp. 665-683. <https://doi.org/10.1080/19415257.2010.533573>
- Young, Michelle D./Crow, Garry M./Murphy, Joseph/Ogawa, Rodney T. (eds.) (2009): *Handbook of research on the education of school leaders*. New York, NY: Routledge.
- Young, Michelle D./Mawhinney, Hanna/Reed, Cynthia J. (2016): Leveraging Standards to Promote Program Quality. In: *Journal of Research on Leadership Education*, 11, 1, pp. 12-42. <https://doi.org/10.1177/1942775116641086>

Distributed Leadership in Schools: German and American Perspectives

Barbara Muslic¹, Jonathan Supovitz² and Harm Kuper³

1. Introduction

Over the past two decades distributed leadership has increasingly entered into leadership conversations across the world (Camburn et al. 2003; Diamond/Spillane 2016; Harris 2008; Spillane 2006). While there is no singular universally accepted definition for the concept (Woods et al. 2004), it is generally understood to expand investigations of school leadership beyond the activity of the school principal. In this chapter we outline the basis for the development of scientific discussions on distributed leadership in a comparison of the German and American contexts. Thereby we highlight this leadership model as a starting point to analyze new organizational (management) structures in schools and to present the leadership and management of schools in conceptual terms for empirical studies.

Two assumptions guide our deliberations. First, while schooling has much in common across the world, we assume specific areas of priority and focus in its discussion in national contexts. According to a proposition of Ballantine and Spade – “schooling is ubiquitous in the world” (2008: xii) – educational interaction generates a universal form of organization. Without exception, it is described as a professional organization with high autonomy on the operational level, flat hierarchies, and a strong importance of professional guidelines for practice. Nevertheless, the basic constellation described here allows considerable scope for elaboration in the details and the accentuation of structural aspects, which are undertaken against the backdrop of national traditions. This was taken into consideration in our comparison of the American and German discussion on distributed leadership. In the American discussion on school leadership, an understanding of school *management* was consolidated much earlier and more clearly, and this is reflected in the debate about distributed leadership. In Germany, by contrast, the individualized responsibility of professional teachers traditionally has a

-
- 1 Barbara Muslic is Postdoctoral Researcher at the Department of Education and Psychology at the Freie Universität Berlin. Email: barbara.muslic@fu-berlin.de
 - 2 Jonathan Supovitz is Professor at the Education Policy Division of the University of Pennsylvania’s Graduate School of Education. Email: jons@upenn.edu
 - 3 Harm Kuper is Professor at the Department of Further Education and Educational Management of the Freie Universität Berlin. Email: harm.kuper@fu-berlin.de

central place in considerations on the structure and management of schools which is also revealed by the hesitant reception of the concept of distributed leadership.

Second, we assume that distributed leadership can not only set normative requirements for the leadership of schools, but also point to basic theoretical or conceptual principles for the analysis of management and leadership in schools. In an applied understanding of education science it is important to separate the two perspectives, but also not lose sight of the connections between them. The analytical perspective represented here is intended to gain insight into the existing practices of distributed leadership. With a research program based on the concept of distributed leadership, these practices can be described and analyzed according to social science theories of interaction, networking or professionalization. Thus, the groundwork is laid for the discussion of practical possibilities in the leadership of schools.

In the following, we first outline the evolution of research topics on the concept of distributed leadership in the United States, and subsequently present the research topic in Germany which was inspired by this concept. In the conclusion, we examine research implications, questions and challenges for the field.

2. The evolution of research on distributed leadership in the American context

American research framing of distributed leadership in the first two decades of the 21st century is broadly acknowledged to emanate from the theoretical work of Peter Gronn and James Spillane. Gronn (2000, 2002) theorized leadership as a joint and interactive performance, and was heavily influenced by Engeström's (1999) activity theory. Gronn conceptualized leadership as the interdependent and coordinated activity of school actors mediated by the tools of their environment. Spillane's theory of distributed leadership (2006) further moved leadership away from attention on the individual and conceptualized leadership as that which emerged from the interactions amongst leaders and followers, regardless of their title or hierarchical position, engaged in specific task-based contexts. In doing so, both Gronn and Spillane challenged our notion of leadership as an individual activity conceptually separable from the context within which it was enacted.

The theory of leadership as distributed practice has opened up several avenues for educational research which American scholars have begun to transverse. First, and more straight-forward distributed leadership, is used to study reforms that expand leadership responsibilities beyond the traditional

role of the school principal. Second, and more conceptually challenging, distributed leadership theory broadens the notion of what constitutes leadership by expanding attention to the professional and social interactions amongst school actors as they engage in their professional work, as well as the social contexts in which leadership activity is embedded.

These two conceptualizations refer closely to important distinctions drawn in the research literature. Mayrowetz examined the different conceptions of distributed leadership used by researchers and distinguished between what he called “distributed leadership for efficiency and effectiveness” (Mayrowetz 2008: 429) and research that uses distributed leadership as a conceptual lens on leadership. Studies using the former tend to be examinations of normative models of how the distribution of leadership tasks influences participants and impacts school outcomes. In these kinds of studies, leadership is still the bailiwick of the individual, but it is spread across a broader set of school actors.

Studies that use the latter conception tend to dig into the complex interactions amongst people, and produce over time different degrees of joint activity. The conceptual perspective of distributed leadership allows for a more nuanced depiction of leadership activity in schools, involving multiple actors and the myriad ways in which they interact, as well as attending to the contextual forces which shape (and in some ways define) their activity. These conceptualizations try to make sense of the complexity by which leadership practice occurs in schools. Further, they de-privilege the roles or positions of school actors and emphasize the activity that emerges from the interactions amongst both formal and informal leaders within educational settings.

Examples of these two strains abound. The first set of research that uses a distributed leadership framework investigates the spread of leadership responsibilities in school reform efforts and how they influence schools. Several studies in the literature illustrate this perspective. Camburn, Rowan and Taylor (2003), for example, examined the ways that three comprehensive school reform models used distributed leadership to rearrange school leadership responsibilities and socialize leaders into their roles. Their conceptualization of distributed leadership came from what Rowan (1990) called “‘network’ patterns of control, where leadership activities are distributed widely across *multiple* roles and role incumbents” (Rowan 1990: 348). Following this emphasis on leadership roles, they used survey data to compare the spread and instructional focus of leadership activity of reform and compared school leaders, and found that the reform models had more leadership roles and enabled more attention to instructional leadership as a consequence of the distribution of leadership.

As another example of this kind of research on distributed leadership, Goldstein (2004) used mixed methods to examine a different configuration of distributed leadership by studying schools that shifted formal leadership

responsibility for teacher evaluation from principals to teachers. She argued that distributed leadership meant expanding leadership responsibility across more school actors. Goldstein framed her study as an extension of policy approaches that have attempted to “alter education’s longstanding hierarchical authority structure, distributing leadership responsibility beyond administrators to include teachers” (Goldstein 2004: 175). She found that the tradition of hierarchy in education, the difficulty of conducting evaluations, district leadership, and program ambiguity were challenges for distributing leadership.

The second vein of scholarship of distributed leadership in the US focuses on how individuals interact around school tasks. This set of research frames leadership as a complex set of interactions amongst educators, and how they shape the ideas and actions that emerge. For example, Scribner, Sawyer, Watson & Myers used the distributed leadership perspective to understand how teacher teams “are embedded in an interactive network of interdependent school activities that collectively constitute leadership” (Scribner et al. 2007: 68). As an element of their conceptualization of leadership, they view decisions as emerging from “dialogue amongst individuals, engaged in mutually dependent activities” (Scribner et al. 2007: 70). Through a discourse analysis of team discussions, they found that the purpose of teams, the autonomy that members felt as decision-makers, and the patterns of discussion amongst team members influenced both group functioning and the exercise of leadership.

In another study, Park & Datnow (2009) used a distributed leadership lens to investigate how teams co-constructed the meaning and structure of data use. Like Scribner et al. (2007), these researchers used a perspective on leadership that emphasized its interactive nature amongst a broad set of school actors in service of social ends. Consequently, they viewed the unit of analysis as “the social interaction within the organization as a whole” (p. 479), rather than the individual. Using interviews and observations, they found that leaders co-constructed data-driven decision-making as a process of continuous learning and diffused decision-making authority to different levels of the system.

The evolving research on distributed leadership in the United States raises several important questions for international scholars. First, an unstated tension underlies this literature. Is distributed leadership a descriptive theoretical perspective from which to gain insights into the workings of schools? Or is distributed leadership a normative statement of how schools should strive to operate? With today’s emphasis on putting knowledge into practice, what are the implications of transporting distributed leadership from a theory into a theory of action?

Second, the theory of distributed leadership opens the door to viewing leadership as an organizational, as well as individual, characteristic. Stressing

leadership as the interactions amongst people embedded within social contexts that bound their choices, raises important questions about the organizational attributes that shape leadership activity in schools.

3. State of research on distributed leadership in the German context

In contrast to the Anglo-American setting, the leadership concept embodied in *distributed leadership* is less well-known and not as widely spread in German-speaking countries. In Germany, the relevant international literature has only attracted interest since the 2000s, and has thus far only been hesitantly received.

There have been a few exceptions in the form of national publications on *distributed leadership* by Bonsen (2009; 2010), who addresses the topic generally, as well as Muslic (2015; et al. 2015; 2016), who considers it in the special context of new governance and the use of evaluation or performance data. The specified publications can primarily be classified as conceptually focused rather than empirical literature.

In German-speaking countries, distributed leadership has been translated or understood literally as “shared leadership.” In terms of understanding, there is an assumed sharing of leadership in the school across the different formal departments or groups, organizational members or units. These mainly include steering groups and committees as well as school management teams or extended school management (Feldhoff/Rolff 2008). The school-specific involvement of these rather cooperative steering groups or teams with school management tasks indicates a new understanding of leadership and a reorganization of the division of responsibility in schools. This concept is thereby linked with *professional learning communities* (Bonsen/Rolff 2006). These describe teams of teachers who are involved in structured development processes through cooperative enquiry, and who are thus intended to contribute to lesson quality assurance in their area of responsibility.

Because of the sparse background in the German-speaking context, the leadership concept presented in *distributed leadership* can be understood as a new analytical perspective, in order to focus on school management teams and school organization. This innovative leadership concept can therefore be seen as the starting point to describe new organizational (management) structures in schools, and to conceptualize the leadership and management of schools for empirical studies.

In German-speaking countries, the discussion about distributed leadership is closely linked to the understanding that schools are considered

as places (organizations) of professional work. Traditionally, this understanding is particularly associated with the individual responsibility of each teacher for their pedagogical practice. The understanding of schools as organizations is still a very new or not fully established point of view in research on school improvement or school effectiveness. However, the term is more intensively used in the context of new governance, in order to examine the effects of new governance mechanisms on the individual school as organization (van Ackeren et al. 2013). Generally a school has little hierarchical and no consistently formalized organizational structure, which has impacts on school management as well as the implementation of quality assurance measures for teaching (Feldhoff 2011). The influence of school management on teaching is described rather cautiously as indirect, whereas the responsibility of teachers and the significance of cooperation between colleagues are very much emphasized. Thus, the development, of teaching is traditionally more strongly anchored in bottom-up communication processes rather than top-down ones.

School organization in Germany has long been considered a bureaucratic matter (Terhart 1986), and the management of schools has been seen as administrative tasks (Bonsen et al. 2002; Rosenbusch 2002); for a long while this also accounted for the relative separation of school management as well as curricular and instructional lesson content. However, in the course of the current reform processes in the context of new governance, there have been far-reaching changes for school organization and its constellations of actors and responsibilities: through decentralized control of schools on the basis of standardized comparative assessments, there has been a resultant strengthening of the autonomy of individual schools and their actors. This results in a transfer of management competence and decision-making authority from the institutional school administration or school system level down to the level of the school organization (Bonsen 2010; Fuchs 2008; Rürup 2007), and also to the functional area of the school management (Fend 2011; Pfeiffer 2002; Rosenbusch 2005; Schleicher 2009). School leaders are moving into a position where they initiate, moderate and give structural support to the development of teaching. This means that the management of schools has evolved into an increasingly complex leadership role, which implies new and changed activities and responsibilities (e.g. increased managerial functions and tasks) (Böttcher 2002; Brauckmann/Hermann 2012; Schleicher 2009). From the American perspective, Mintrop incisively describes this development in the German school system as “management reform without managers” (2015: 791). In this context, a growing number of collaboratively organized forms of management responsibilities are establishing themselves, where school management is becoming intermeshed with the bottom-up processes of teaching staff. This includes divisional responsibilities for

subjects or subject groups, year groups, pedagogical coordinating bodies, but also the less formalized votes within subject committees.

Over a long period in the discussion of the management of schools in Germany there has been a shift in perspective from a traditional bureaucratic model to a management-oriented model of school: schools no longer correspond to the former image of an administratively led organization with professionals acting to a great extent independently, but rather fit far more the image of a management-oriented organization with professionals who develop a joint program for the individual school and collectively supported quality standards for teaching. In this respect, new or innovative forms of functional differentiation play a role, in which departments, subject-specific committees, cooperation and coordination are experiencing increased importance (Thiel 2008).

Against this background, distributed leadership acquires increased relevance. Triggered by test-based school reform, internal school coordination requirements, which until now were barely developed structurally, are becoming clearer, with greater need to be anchored in internal school organization and responsibility frameworks (Muslic et al. 2015). Management functions are attracting greater attention and require a connection with the horizontal arrangement of the organizational structures of a professional organization. This influences the school management's understanding and practice of leadership. The innovative, management-oriented leadership concept represented in distributed leadership can be linked to this, as it is primarily characterized by a horizontal leadership level in the school organization or a decentralized idea of leadership. Management responsibility should accordingly be transferred via the organization to further internal school actors and departments, thereby also strengthening the organizational responsibility in a formal sense.

Originating from the conceptual idea of distributed leadership, suppositions or hypotheses can be established and become the subject of empirical studies. This means we can envisage an indirect impact of school management on teaching, to the extent that following the communication channels through departments, whereby questions can reach the teaching staff. It should be examined by what means binding decisions are made on teachers' development and on the quality assurance of pedagogical work. This connection becomes clear, for example, in relation to the use of returned evaluation and test data in the context of test-based school reform: early findings suggest that a school management which acts according to distributed leadership – in the case of weak performance or evaluation results – promotes a higher responsibility of the professionals in the whole school organization related to quality assurance measures (Gronn 2002; Bonsen 2010; Huber 2008; Muslic 2016). In this case, the school management can specifically influence how evaluation or test results are handled in the context

of teaching development, by addressing, for example, the subject groups or committees as the responsible persons for the operational processing of these results and for coordinating the examination of these results. These observations support the assumption that distributed leadership in German schools is accelerating the development of management structures which mediate between a single organizational head and the teachers responsible at an individual level for their teaching (Muslic 2017).

4. Discussion

The chapter has outlined the basis of the development of scientific discussions of distributed leadership by a comparison of its use and connections within the German and American contexts.

Both contexts are characterized by different lines of development and traditions in the respective school systems. The reception as well as practical anchoring of this leadership concept thus correlates to the differing premises and structural factors inherent in both contexts.

Further research perspectives can be inferred from this discussion. First, the distributed leadership perspective raises both challenges and opportunities for researchers in each context. To identify, capture, and make sense of complex leadership interactions over time, we need better tools and methods. Extensions of social network analysis offer some promising opportunities in this regard, but this method is still in its nascence. The field also needs to have a better understanding of the relational qualities embedded within professional interactions and how these lead to different kinds of interactions, as well as the contextual mediators of these interactions. Additionally, if we view interactions as the unit of analysis, there are important conceptual and analytical implications for both qualitative and quantitative researchers, for interactions are multi-perspectival and ephemeral. Despite these challenges, distributed leadership is changing the way that both scholars and practitioners understand leadership practice in schools.

The distributed leadership perspective also raises the important question of where tasks can be specified in the school organization. In this regard, the themes of functional differentiation or internal school task sharing, distribution of professional responsibility, participative decision-making processes as well as the management-oriented coordination of school and teaching themes, all come to the fore. This innovative and complex leadership model can be seen as a starting point to describe new organizational (management) structures in schools, and to present the leadership and management of schools in conceptual terms for empirical studies. The analytical perspective and the theoretical-conceptual understanding of distributed

leadership could in the future contribute to this leadership concept experiencing increased consideration and a wider reception – precisely because it is viewed as an effective form of leadership with regard in particular to school change processes and social change (Harris 2004; Leithwood et al. 2004; Supovitz 2018).

Moreover, there is potential for a unifying perspective: the analytical perspective of distributed leadership allows school leaders and the school organization, or also the interaction or teaching – either as separate areas or in a connected manner – to be more closely considered. This theoretical concept is therefore characterized by a high level of flexibility.

At the same time, as a theoretical concept, distributed leadership corresponds to a universal idea of schools. This theoretical concept is thus particularly suitable for a comparison in different contexts and countries, since it not only offers a general or overarching basis for comparison, but also allows flexibility and sensitivity with regard to different contexts and specific characteristics. That means the distributed leadership model is compatible to different contexts (like low vs. high stakes contexts) in different countries (like USA, European countries, Singapore) and provides in a first instance a cross-cultural transferability (Hairon/Goh 2015). Further empirical research is needed to explore whether cultural and contextual factors have an impact on shaping distributed leadership practices in schools.

References

- Ballantine, Jeanne/Spade, Joan (2008): *Schools and Society: A Sociological Approach to Education*. 3rd edition. Thousand Oaks: Sage.
- Bonsen, Martin (2010): *Distributed Leadership in der Schule*. In: *Erziehung und Unterricht*, 7/8, pp. 619-627.
- Bonsen, Martin (2009): *Führung, Delegation und 'distributed leadership'*. *Unterrichtswirksame Schulleitung in Zeiten der datengestützten Schulentwicklung*. In: Bohl, Thorsten/Kiper, Hanna (eds.): *Lernen aus Evaluationsergebnissen. Verbesserungen planen und implementieren*. Bad Heilbrunn: Klinkhardt, pp. 267-282.
- Bonsen, Martin/Rolff, Hans-Günter (2006): *Professionelle Lerngemeinschaften von Lehrerinnen und Lehrern*. Paralleltitel: *Professional learning communities of teachers*. In: *Zeitschrift für Pädagogik*, 52, 2, pp. 167-184.
- Bonsen, Martin/von der Gathen, Jan/Iglhaut, Claus/Pfeiffer, Hermann (2002): *Die Wirksamkeit von Schulleitung. Empirische Annäherungen an ein Gesamtmodell schulischen Leitungshandelns*. Weinheim und München: Beltz Juventa.
- Böttcher, Wolfgang (2002): *Kann eine ökonomische Schule auch eine pädagogische sein? Schulentwicklung zwischen Neuer Steuerung, Organisation, Leistungsevaluation und Bildung*. Weinheim und München: Beltz Juventa.

- Brauckmann, Stefan/Herrmann, Christoph (2012): Schulleitungshandeln im Rahmen neuer Steuerung – Belastung oder Chance?! In: Zeitschrift für Bildungsverwaltung, 28, 1, pp. 87-98.
- Camburn, Eric/Rowan, Brian/Taylor, James E. (2003): Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. In: Educational evaluation and policy analysis, 25, 4, pp. 347-373.
- Diamond, John B./Spillane, James P. (2016): School leadership and management from a distributed perspective: A 2016 retrospective and prospective. In: Management in Education, 30, 4, pp. 147-154.
- Engeström, Yrjö (1999): Activity theory and individual and social transformation. In: Engeström, Yrjö/Miettinen, Reijo/Punamäki, Raija-Leena (eds.): Perspectives on activity theory. Cambridge: Cambridge University Press, pp. 19-38.
- Feldhoff, Tobias (2011): Schule organisieren. Der Beitrag von Steuergruppen und Organisationalem Lernen zur Schulentwicklung. Wiesbaden: VS.
- Feldhoff, Tobias/Rolff, Hans-Günter (2008): Einfluss von Schulleitungs- und Steuergruppenhandeln. In: Holtappels, Heinz Günter/Klemm, Klaus/Rolff, Hans-Günter (eds.): Schulentwicklung durch Gestaltungsautonomie. Ergebnisse der Begleitforschung zum Modellvorhaben „Selbstständige Schule“ in Nordrhein-Westfalen, Münster: Waxmann, pp. 293-302.
- Fend, Helmut (2011): Die Wirksamkeit der Neuen Steuerung – theoretische und methodische Probleme ihrer Evaluation. In: Zeitschrift für Bildungsforschung, 1, 1, pp. 5-24.
- Fuchs, Hans-Werner (2008): Educational Governance und neue Steuerung: Grundsätze – Beispiele – Erwartungen. In: Langer, Roman (ed.): „Warum tun die das?“ Governanceanalysen zum Steuerungshandeln in der Schulentwicklung. Wiesbaden: VS Verlag, pp. 19-29.
- Goldstein, Jennifer (2004): Making sense of distributed leadership: The case of peer assistance and review. In: Educational evaluation and policy analysis, 26, 2, pp. 173-197.
- Gronn, Peter (2000): Distributed properties: A new architecture for leadership. In: Educational management & administration, 28, 3, pp. 317-338.
- Gronn, Peter (2002): Distributed leadership. In: Leithwood, Kenneth A./Hallinger, Philip (eds.): Second international handbook of educational leadership and administration. Dordrecht: Kluwer, pp. 653-696.
- Hairon, Saleh/Goh, Jonathan W. P. (2015): Pursuing the elusive construct of distributed leadership: Is the search over? In: Educational Management Administration and Leadership, 43, 5, pp. 693-718.
- Harris, Alma (2004): Distributed Leadership and School Improvement. Leading or Misleading? In: Educational Management Administration & Leadership, 32, 1, pp. 11-24.
- Harris, Alma (2008): Distributed Leadership: According to the Evidence. In: Journal of Educational Administration, 46, 2, pp. 172-188.
- Huber, Stephan Gerhard (2008): Steuerungshandeln schulischer Führungskräfte aus Sicht der Schulleitungsforschung. In: Langer, Roman (ed.): „Warum tun die das?“. Wiesbaden: VS Verlag, pp. 95-126.
- Leithwood, Kenneth/Seashore Louis, Karen/Anderson, Stephen/Wahlstrom, Kyla (2004): How leadership influences student learning.

- <http://www.wallacefoundation.org/knowledge-center/school-leadership/key-research/Documents/How-Leadership-Influences-Student-Learning.pdf>
- Mayrowetz, David (2008): Making sense of distributed leadership: Exploring the multiple usages of the concept in the field. In: *Educational Administration Quarterly*, 44, 3, pp. 424-435.
- Mintrop, Rick (2015): Public Management Reform Without Managers. The Case of German Public Schools. In: *International Journal of Educational Management*, 29, 6, pp. 790-795.
- Muslic, Barbara (2015): Distributed Leadership. Wissenschaftliche Befunde zur Wirksamkeit. *Schulmanagement*. In: *Die Fachzeitschrift für Schul- und Unterrichtsentwicklung*, 3, pp. 8-11.
- Muslic, Barbara (2016): Schulleitungshandeln. Distributed Leadership und Lernstandserhebungen. *Schulmanagement*. In: *Die Fachzeitschrift für Schul- und Unterrichtsentwicklung*, 6, pp. 22-24.
- Muslic, Barbara (2017): Kopplungen und Entscheidungen in der Organisation Schule. Organisationsbezogenes Schulleitungshandeln im Kontext von Lernstandserhebungen. Wiesbaden: Springer VS.
- Muslic, Barbara/Brauckmann, Stefan/Basold, Katrin (2015): Distributed Leadership: Why does it matter? In: *Evangelische Akademie Loccum* (ed.): Was wirklich wirkt!? Effektive Lernprozesse und Strukturen in Lehrerfortbildung und Schulentwicklung. Loccumer Protokolle, pp. 107-124.
- Park, Vicki/Datnow, Amanda (2009): Co-constructing distributed leadership: District and school connections in data-driven decision-making. In: *School leadership and Management*, 2, 5, pp. 477-494.
- Pfeiffer, Hermann (2002): Forschung zur Schulleitung – Schwerpunkte und Perspektiven. In: Huber, Stephan Gerhard/Wissinger, Jochen (eds.): *Schulleitung – Forschung und Qualifizierung*. Opladen: Leske + Budrich, pp. 21-32.
- Rolff, Hans-Günther (2010): Leitung als Dreiklang. In: Rolff, Hans-Günther (ed.): *Führung, Steuerung, Management*. Seelze: Klett.
- Rosenbusch, Heinz S. (2002): Schulleitung als Beruf. Hintergründe einer historischen Zäsur. In: *Zeitschrift Schulmanagement*, 33, 1, pp. 20-22.
- Rosenbusch, Heinz S. (2005): *Organisationspädagogik der Schule. Grundlagen pädagogischen Führungshandelns*. Neuwied: Luchterhand.
- Rowan, Brian (1990): Commitment and control: Alternative strategies for the organizational design of schools. In: *Review of research in education*, 16, 1, pp. 353-389.
- Rürup, Matthias (2007): *Innovationswege im deutschen Bildungssystem. Die Verbreitung der Idee „Schulautonomie“ im Ländervergleich*. Wiesbaden: VS Verlag.
- Schleicher, Andreas (2009): Moderne Schulleitung im Wandel. Schlussfolgerungen aus OECD-Analysen. In: *Die Deutsche Schule*, 101, 4, pp. 311-322.
- Scribner, Jay Paredes/Sawyer, Keith/Watson, Sheldon T./Myers, Vicki L. (2007): Teacher teams and distributed leadership: A study of group discourse and collaboration. In: *Educational Administration Quarterly*, 43, 1, pp. 67-100.
- Spillane, James P. (2006): *Distributed leadership*. John Wiley & Sons.
- Supovitz, Jonathan A. (2018): Teacher leaders' work with peers in a Quasi-formal teacher leadership model. In: *School Leadership & Management*, 38, 1, pp. 53-79.

- Terhart, Ewald (1986): Organisation und Erziehung. Neue Zugangsweisen zu einem alten Dilemma. In: Zeitschrift für Pädagogik 32, 2, pp. 205-223.
- Thiel, Felicitas (2008b): Organisationssoziologische Vorarbeiten zu einer Theorie der Schulentwicklung. In: Journal für Schulentwicklung, 12 2, pp. 31-39.
- Van Ackeren, Isabell/Heinrich, Martin/Thiel, Felicitas (2013): Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund. In: Die Deutsche Schule, 12. Beiheft. Münster: Waxmann.
- Woods, Philip A., Bennett, Nigel/Harvey, Janet A./Wise, Christine (2004): Variabilities and dualities in distributed leadership: findings from a systematic literature review. In: Educational Management Administration and Leadership, 32, 4, pp. 439-457.

II. Migration, Refugees, and Public Education

Section Editors:

Lisa Damaschke-Deitrick, University of Tübingen

Alexander W. Wiseman, Texas Tech University

Migration, Refugees, and Education: Challenges and Opportunities

Lisa Damaschke-Deitrick¹ and Alexander W. Wiseman²

1. Introduction

Education is universally presented to migrant and refugee youth and their families as a panacea to help them transition smoothly into their new communities (Wiseman/Damaschke-Deitrick/Galegher/Park 2019). As such, education is expected to deliver opportunities beyond academic schooling and is viewed as a mechanism to socially integrate youth into their new communities as well as transform them into productive citizens (Beirens et al. 2007; Kia-Keating/Ellis 2007). However, in many cases, education systems and educators are not prepared for the unique needs and challenges of refugee and forced migrant students.

The contexts of transition for refugee and forced-migrant youth is also key to education as a panacea. As the chapters in this section suggest, the unique experiences of refugee youth along the path from their home communities, through different displacement experiences, and eventually into a relatively permanent new community is quite varied. The experience of an affluent Syrian family and its children from war-torn Syria to Western Europe is quite different from that of an unskilled and illiterate Somalian refugee youth who finds herself permanently residing in a refugee camp in Turkey. And, both are uniquely different from the experiences of an ethnic-minority Congolese refugee youth fleeing extreme violence and human rights violations who has been vetted and resettled in the United States as an officially-designated refugee.

The challenge for education in receiving communities is to balance both humanitarian needs reflected in the diverse set of experiences and history outlined above and the demand from mainstream communities for education that creates productive citizens in terms of social and economic mobility as well as contributions to both individual and community well-being. This is a difficult enough task when students are already diverse in local communities,

1 Lisa Damaschke-Deitrick is Senior Lecturer and Researcher at the Institute of Political Science, University of Tübingen. Email: lisa.damaschke@uni-tuebingen.de

2 Alexander W. Wisemann is Professor of Educational Psychology & Leadership at the College of Education, Texas Tech University Lubbock. Email: Alexander.Wiseman@ttu.edu

but it becomes even more challenging for refugee and forced migrant youth as well as local educators in their new communities.

Chapters in this section on *Migration, Refugees, and Public Education* address both challenges and opportunities in education for refugee students, migrant families, and their teachers and educators using evidence from new research. To contextualize these chapters, we provide a definitional and conceptual framework for understanding the characteristics and contexts of refugee and migrant students. We then discuss the role that “education as a panacea” plays in both refugee and migrant students’ transition as well as the provision of education that follows. The unique intersection of trauma, identity, and language issues (TIDAL), which defines the refugee experience, is then explored. Finally, we introduce each chapter in this section, which both individually and collectively contribute to a broader understanding of refugee and migrant education.

2. Refugee and migrant students

There are over 79.5 million refugees, asylees, and other forced migrants worldwide (UNHCR 2020). The experience of refugees and others fleeing refugee-like situations is embedded with experiences of violence and trauma starting in their home communities, then again as they flee and migrate, and finally during resettlement in their receiving communities (UNICEF 2016). Among those refugee and forced migrant students who have access to education, there are many challenges they must still overcome including past traumas, unstable home environments, and socio-cultural instability. These factors combine to create frequent and persistent risks to their psychological and social well-being (Hadfield/Ostrowski/Ungar 2017).

The United Nations High Commissioner for Refugees (UNHCR) defines a “refugee” as “someone who has been forced to flee his or her country because of persecution, war, or violence” with “a well-founded fear of persecution for reasons of race, religion, nationality, political opinion or membership in a particular social group” (para. 1). As a result of “war, ethnic, tribal, and religious violence,” (UNHCR n.d.: para. 1), refugees cannot return home. These circumstances mean that refugees have special legal status and protections in most receiving countries, which are not available to other migrants (Buckner et al. 2018). Refugees’ rights and privileges in receiving countries are politically-constructed by receiving countries’ foreign policies regarding the provision and timing of assistance as well. This creates an important distinction between who is a refugee and who is a migrant.

Distinctions between refugees, including forced migrants, and those who migrate or immigrate for economic reasons include two key factors. First,

refugees and forced migrants face a change in their living conditions that jeopardize their lives and are unrelated to their economic situation (Joly 2002). Second, economic migrants may leave their homes out of optimism for what is possible even though they could remain in their current locations; whereas, refugees and forced migrants flee for their lives and cannot remain in their original locations (Joly 2002).

Displacement occurs for a variety of contextual reasons, and the distinction between documented and undocumented refugees or asylum-seekers is often a question of politics (Bartlett/Ghaffar-Kucher 2013). Asylum-seekers do not always have the legal protection that a recognized refugee's status brings. An asylum-seeker has been defined as "someone whose request for sanctuary has yet to be processed" (UNHCR 2017). Approximately 3.5 million persons were waiting for a decision on their asylum claims worldwide in 2018, and 1.7 million new asylum requests were submitted that year (UNHCR 2019). According to UNHCR, the United States received the highest rate of new asylum requests with 254,300 claims, followed by Peru with 192,500 claims and Germany with 161,900 claims in 2018. Overall, most asylum applications came from Syria, with over half a million claims, followed by people from Venezuela with 341,800 asylum requests (UNHCR 2019).

From October 2016 to September 30, 2017, the United States granted asylum status to 26,568 people (Blizzard/Batalova 2019), most of them coming from countries in Central America (El Salvador, Guatemala, Honduras and Mexico) and Venezuela. It has, until recently, been common for individuals from Central America seeking asylum from gang violence and domestic abuse to be granted asylum in the United States. However, public perception of migrants varies based on their legal status, meaning that unauthorized immigrants in the US are sometimes viewed as a threat (Oliviera/Lima Becker 2019). Beyond granting protection to asylum seekers who claim asylum from within the country, the United States also accepts refugees for resettlement (Blizzard/Batalova 2019). As a result of changing political attitudes and changes in policies, the number of resettled refugees has significantly decreased in the United States (Fratzke 2017). The US only resettled approximately 23,000 refugees in 2018 compared to 97,000 in 2016 (Radford/Connor 2019).

In Germany the number of asylum applications peaked in 2016 after almost one million refugees entered the country in 2015. There were 745,545 initial and subsequent applications for asylum in 2016. Since then, the number of applications has decreased. In 2017, Germany counted a total of 222,683 initial and subsequent applications for asylum (bpb 2019a). About one third of asylum seekers are granted a refugee status in Germany and are allowed to stay in the country (bpb 2019b).

The official UNHCR definition of refugees does not explicitly mention migrants and forced migration. The International Organization for Migration (IOM) (n.d.) defines forced migration as “migratory movement in which an element of coercion exists, including threats to life and livelihood, whether arising from natural or man-made causes.” Internally displaced persons (IDPs) are also migrants who are not officially classified as refugees. IDPs remain in their own countries and are legally protected by their governments, but they are still highly vulnerable people, who are often denied access to humanitarian aid and education. There are currently more than 41 million IDPs worldwide due to “armed conflict, generalized violence or human rights violations” (UNHCR n.d.; UNHCR 2019).

Refugee identity is less static than the legal definition of refugee. The experiences of refuge seekers suggest that their identity is fluid and dependent upon context. One of the more well-known explanations of this experience comes from Hannah Arendt (1994), who described her experience as a refugee in the 1940s as akin to arriving in a new location without resources and needing help. Arendt explains the lack of agency that refugees experience during their forced displacement by emphasizing the ways in which refugees are victims and that their actions are not the cause of their situations.

Likewise, the label of ‘refugee’ is often applied to those who are forced migrants to make benefits or resources available to them in their receiving country, but these labels are often stigmatizing (Burnett 2013; Zetter 2007). On the one hand, the stigma of being a refugee is frequently oppressive and those experiencing that stigma may seek alternative labels and roles to alleviate the stigma as much as possible. For example, Galegher’s chapter in this section on *Migration, Refugees, and Public Education* documents how refugees in Egypt hid their refugee status, and instead shared their new identities as university students (see Damaschke-Deitrick/Galegher/Park 2019). On the other hand, being labeled a refugee or asylum-seeker allows some forced migrants to be less vulnerable and more stable in their role and community (Oliveira/Becker 2019). Unfortunately, documentation of legal refugee status does not ensure that there will be consistency in the ways that refugees experience their situation or their identity.

The balance between the shared experiences of refugees and other forced migrants and their unique contexts and experiences is important to note. While forced migrants often share the experiences of war, persecution, and violence as they are displaced from their homes. They also are consistently unwilling victims of the injustices associated with these experiences. Most refugees experience significant trauma as they are forcibly displaced, too. Yet, there are different experiences in the ways that refugee youth navigate their documentation status in receiving countries. They also each build a new identity and reconcile their existing identity differently depending on where they relocate and how the receiving community facilitates that relocation.

This section on *Migration, Refugees, and Public Education* uses a more inclusive definition of refugee, asylum-seeking and migrant youth, which echoes the need for flexibility and contextualization that refugee voices have raised. The use of these terms also acknowledges that mass refugee crises in the 21st century are significantly different from refugee and other forced migration in the 20th century and earlier (Zetter 2007). Changes in refugee populations in the 21st century are expected due to an increase of the intensity of climate change and natural disasters, a rise in terrorism, an increase in IDPs, and an escalation of severe socioeconomic deprivation (McBrien 2016). Each of the chapter contributions to this section embraces both the political definition as well as the more figurative definition of refugees and asylum-seeking youth, which may change “based upon the individual, society and place: ranging from those in camp situations to someone awaiting an asylum decision to a refugee successfully integrated into his/her new host society” (Burnett 2013: 2).

If refugee and forced migrant youth participate in some form of schooling in their new locations, it is far from home and often separated from parents and family. The institution of schooling is remarkably stable and stabilizing for refugees and forced migrants because many experienced it in their home communities before being displaced. It is also a mechanism for the delivery of resources, care, counseling, and opportunities as they build a new life and recreate their identity in their new homes once relocated. School is a constant in the lives of refugee and migrant youth, even when they experience instability in most other aspects of their lives (Wiseman/Damaschke-Deitrick/Galegher/Park 2019).

3. Education as a panacea

Education is and has historically been viewed – whether appropriately or not – as a cure for problems beyond academic knowledge and skills (Amos/Wiseman/Rohstock 2014; Wiseman/Damaschke-Deitrick/Bruce/Davidson/Taylor 2016). The use of education as a panacea has been especially prevalent since the expansion of mass education beginning in the early 20th century. Since then, politicians, parents, teachers, and community leaders have systematically used it – often in the form of formal schooling – as a tool to supposedly cure social, economic, political, and many other problems whose origins lie beyond schools. In other words, education is often viewed as a panacea for problems out of the scope of schools or academic teaching and learning. It also carries with it a significant disadvantage. Not only is education unable to consistently resolve problems outside of the scope of the school building, but also policymakers and others have used the taken-for-

granted expectation that schooling is a way to resolve social, economic, and political problems to blame schools and teachers for these problems (Wiseman et al. 2019).

Since refugee and migrant youth are significantly affected by social, economic, and political problems, or may have been forcibly displaced because of these problems, education is often seen as a panacea for the trauma, identity issues, communication difficulties, and other problems that they may bring with them to their receiving communities. Education is also often viewed as a stabilizing force, which can have positive benefits for youth who experience instability and displacement, as refugees and other forced migrants do (Damaschke-Deitrick/Bruce 2019).

The war in Syria has led to the forced migration of more than 12 million people, of which at least six million were school-aged children (Sirin/Rogers-Sirin 2015). The six million of those school-aged children are likely victims of trauma, violence, and persecution either in their home country, during their relocation, or since they relocated in their receiving communities. They might have been given the opportunity to attend some sort of schooling, if they stayed at any point during their relocation at a refugee camp. And, once they reach their receiving country, they are likely to be expected to attend school or are given the option to attend school alongside the school-aged children who are native to that community. In each of these instances, the role of education is expected by educators, their community, and often the parents themselves to provide more than an academic education. The expectation is often that education for refugees provides a foundation for social and economic mobility; for civic education and how to be a good citizen, and for socialization and acculturation into the host or receiving community's society and culture. In short, education is sought to be a panacea for these youth at every opportunity, regardless of the possibility of it really being able to provide that level of service (Wiseman et al. 2019).

Education does provide some solutions to the problems that refugees and forced migrants face, but they are often not unique to the needs or contexts of those youth. For example, education is a mechanism for integrating refugee and forced migrant youth into their receiving communities, which can also improve their social opportunities (Beirens et al. 2007; Kia-Keating/Ellis 2007). Participation in education is indeed often a way to encourage social mobility among refugee and forced migrant populations over time, and it is especially helpful with poor, marginalized, and often under-educated youth in immediate conflict and post-conflict situations. Historically, refugees have had low levels of schooling, little vocational skills, and few financial resources (Strekalova/Hoot 2008); however, little is known about the social mobility effects of education on refugee and forced migrant youth who are less marginalized and more highly-educated when they migrate. This has

frequently been the situation across Europe when working with refugee communities from Syria (Sasnal 2015).

The conflict in Syria has led to many families and their children resettling outside of conflict zones. In these cases, refugee youth may not have experienced the same levels of extreme violence and trauma as some others, but they still find themselves with little knowledge of their new, unfamiliar locations (Strekalova/Hoot 2008). Socialization is defined as the “process of acquiring the norms to which all the members of a society conform” (Arnstine 1995: 5). Teachers and other educational professionals are key contributors to refugee, forced migrant youth socialization in their receiving communities (Mickan et al. 2007). As microcosms of the broader society, schools and in turn classrooms afford refugee and forced migrant youth the opportunity to experiment with socio-cultural norms and values in a closed environment first. They can then use their new-found understanding of socio-cultural norms and values in the wider society and cultural community outside of the school or classroom (Mickan et al. 2007).

Sometimes educational systems and schools in receiving countries plan the experiences and socialization of refugee and forced migrant youth through specific educational policies and training programs for educators. Although these policies and trainings are useful, the educators responsible for implementing the policies and enacting their training are themselves the products of their own socio-cultural experiences and contexts (Schmidt/Datnow 2005). As a result, these educators individually interpret and enact policies and trainings related to refugee and forced migrant student needs (Spillane et al. 2002). For example, teachers may engage in either planned or impromptu “social pedagogy” to develop intercultural identity awareness, teach socio-cultural norms and values, or emphasize communication in the local language (Schneider 2018).

4. Intersection of trauma, identity, and language (TIDAL)

Most education-related studies, as McBrien (2005) points out, consider both refugee and migrant education simultaneously. Studies on the education of immigrant children and adolescents have mainly focused on educational outcomes (see Portes/Rumbaut 1996; 2001), the relation of language learning and academic achievement (Azzolini/Schnell/Palmer 2012; Cobb-Clark/Sinning/Stillman 2012; Entorf/Minoiu 2004; OECD 2012), and lastly on multiculturalism and diversity (Banks 2004). However, the conditions under which refugees are forced to leave their country differ significantly from other immigrants’ experiences, which can pose specific challenges and opportunities to education systems and schools. Evidence suggests that the unique

intersection of trauma, identity, and language issues (TIDAL) defines the refugee and forced migrant experience and needs to be considered by educators and researchers alike working with refugee and forced migrant youth.

Trauma. Refugee and forced migrant children and adolescents often share experiences with conflict, war, persecution and violence as well as displacement from their home. These experiences create high risks to their psychological and social well-being. Research has shown various experiences of trauma and loss that refugee children go through that impact how refugee students learn, behave, and interact with others (Mendenhall/Bartlett 2018; Fegert/Diehl/Leyendecker/Hahlweg/Prayon-Blum 2018; Dryden-Peterson 2015). As Fegert et al. (2018) point out, refugee children and youth are at risk of experiencing trauma in their home country, while fleeing, and when resettling and trying to adapt to the new receiving community. They are at great risk for developing mental and socio-emotional illnesses as well as long-lasting developmental disorders (Fegert et al. 2018). Children with unstable homes or with traumatized parents are at an even higher risk. This shows the need for educators and teaching staff at schools and universities to understand how to recognize symptoms of trauma and how to respond to them in order to support those students. It is important to note that educational support for refugee and forced migrant youth has been shown to be more impactful as a long-term approach, rather than a quick fix (Francis/Yan 2016).

Research suggests that teachers need better professional preparation to support students with trauma. Teachers and educators are rarely trained for trauma-informed teaching (Phifer/Hull 2016; Thomas 2016; Wiseman/Galegher 2019). Additionally, there is often a lack of professional training for teachers to work with students from diverse cultural backgrounds, which enables them to recognize and value existing cultural competencies as well as existing language competencies (Gitlen/Buendía/Crosland/Doumbia 2003: 118).

Identity. Newcomers often experience feelings of disconnection, social and cultural isolation and a “culture shock” in their host countries (Abu El-Haj 2007; Wiseman/Galegher 2019). Being in a new place, they need to reconcile their existing identity into the host or receiving community’s society and culture. In addition to that, immigrant and refugee students and their parents are often not familiar with the education system, and refugee students, in particular, do not always have access to the same educational opportunities and extracurricular activities as their native peers (Schnepf 2007). Many schools and universities struggle to bridge the refugee students’ previous education with that received in the host countries’ classrooms and to offer support on an individual basis. Refugee students are more frequently marginalized and attend less academically demanding schools or school tracks. For example in the case of Germany, most adolescent refugees are placed into

less academic school tracks upon their arrival, which makes access into university more difficult afterwards (UNESCO 2018; Damaschke-Deitrick/Bruce 2019). This practice does not only lead to lower educational qualifications and degrees but also to lower social recognition.

Research shows that the experience of trauma, fear and safety concerns affect both the ability to learn and identity development (Collet/Bang 2016). This is a unique challenge for refugee students, and evidence suggests that educators must be better prepared for it (Wiseman/Galegher 2019). Other obstacles are challenging immigration laws or negative public attitudes towards refugees or immigrants, as discussed Filsecker and Abs that can impact teachers' attitudes. Bias in schools or among teachers negatively affects immigrant and refugee youth in the classroom, including underestimating their competencies (Wiseman et al. 2019). Also, in order to build a supportive and inclusive school environment, there is a need for teachers and educators to challenge negative or dismissive rhetoric spread by some media outlets or politicians about immigrants and refugees (Mendenhall/Bartlett 2018).

In addition to that, social and cultural marginalization and disconnection from more typical life and education experiences in a host country can lead to personal challenges and identity crises among refugee youth. Unsurprisingly, the immediate needs and crises of refugee and forced migrant youth that teachers must acknowledge and address often overshadow the necessity of developing cultural awareness and social competencies among these youth. Evidence suggests it is crucial, however, to develop conceptualizations of how refugee students can be integrated in schools and universities in a balanced and inclusive way without being negatively stigmatized or being solely treated as a victim of trauma instead of as resilient individuals (Dryden-Peterson 2011, 2016; Dryden-Peterson et al. 2018; Taylor/Sidhu 2012). In this way, schools and universities can serve as a constant, stabilizing force for refugee students and as a "return to normalcy", even when they face instability in other spheres of their lives.

Language. Language is one of the main sources for one's social identity and belonging to a social group and context. The acquisition of the host country language is seen as precondition for newcomers to be able to interact socially and as a result integrate into a new community. Language skills are also described as key for immigrants and refugees to achieve success in school or university (see chapter Fleckenstein/Maehler/Pötzschke/Ramos/Pritchard). However, the experience of trauma and the feeling of loss of their "old" identity can impact the openness and ability of a person to learn a new language.

The language of the host country also makes a difference, as some comparative studies suggest. Educational achievement is higher for those students that immigrated to an English-speaking country (Schnepf 2007). Also, younger students are more likely to become fluent in their second language

than their parents or students that left their home country at an older age (Azzolini/Schnell/Palmer 2012; Cobb-Clark/Sinning/Stillman 2012). Overall, a lack of skills in the host country's language is linked to lower educational achievements for immigrant and refugee students. At schools and higher education institutions, however, teaching modifications are not always available, and even switching to a different language to assist students is often not practiced (Damaschke-Deitrick/Bruce 2019). Studies suggest that most teachers are not prepared to work with new language learners (Lucas/Villegas 2010).

Research shows that teaching approaches involving translanguaging can be beneficial, which involve the integration of native languages in the classroom. Translanguaging values the students' existing language competencies and it helps to bridge across languages (Bajaj/Bartlett 2017). However, it is important to note that educators and schools should not only focus on second language learning but also on the interrelation between trauma, identity, and language.

5. Contributions in the section on migration, refugees, and public education

The movement of people through both voluntary and forced migration poses unique challenges for public education systems in receiving or host countries. In many contexts, educators and educational systems may not be prepared for the unique concerns and real problems that migration and refugee needs pose. Yet, there are examples of programs and contexts where refugee and migrant students are served and may even complement the ongoing education of mainstream students in receiving countries' schools. The contributions in the section on *Migration, Refugees, and Public Education* address the challenges that youth and educators face posed by refugees and other migrant students in public education systems in different country contexts. Both the challenges and opportunities for refugee children and youth, migrant families, and their teachers and educators are addressed in these chapters.

The chapter by Fleckenstein, Maehler, Pötzschke, Ramos, and Pritchard examines language as a predictor and an outcome of acculturation. Acquiring the language skills of the host country is a central predictor of educational outcomes and vocational success. Considering the relevance of language skills in the acculturation process, there has been surprisingly little research on the topic in the context of refugee children and youth. A literature search of English-language publications found 22 peer-reviewed empirical studies that investigate language skills of young refugees, only some of which provided relevant information on age, sex/gender, length of stay, educational

background, or country of origin of their sample. The chapter provides an overview of these studies and points out research gaps pertaining to refugee children and youth language acquisition.

Attitudes towards refugees is the focus of the chapter by Filsecker and Abs. The authors develop and scale items for the measurement of attitudes towards refugees. First, they describe the current practices of item development and its challenges. Second, the authors argue for a new perspective on attitude measurement. Finally, they provide an illustration of a concrete scale under the guidelines of a specific scaling model and discuss the potential of this approach.

Finally, Galegher examines female refugees' experiences in Egyptian higher education. The author describes the opportunities and challenges for female refugee students from Syria and Yemen enrolled in universities in Egypt. Using qualitative data analysis of interviews with female university refugees, findings suggest that cultural and linguistic similarities along with universities' pre-existing infrastructure significantly ease transitions and provide greater access to non-English speaking refugees, often the most marginalized. Although significant differences exist between experiences in public versus private universities, all women expressed the opportunity to attend university as life-changing and empowering. As a result, higher education institutions in the Middle East must be acknowledged and utilized as an investment in long-term durable solutions for refugees.

Through the lens provided by these three chapters and the contextualization of ways of identifying, defining, and giving voice to refugee and similar youth, the education of refugee and migrant youth may be more clearly and comprehensively understood. Awareness and understanding are key first steps in most change processes, which suggest that changes in national policies, international actions, and local accommodations and supports that are provided for refugee and migrant youth may begin with this section on *Migration, Refugees, and Public Education*. Further, understanding of the impact that the application of trauma-informed teaching, civic and social identity formation, and translanguaging may contribute to the development of policies and their implementation for the support and accommodation of refugee and migrant youth. In other words, this section is a foundation for both understanding and action, and as such is not only relevant to researchers and scholars, but is useful for policymakers, development officials, and educators at all levels who are part of the refugee and migrant experience.

References

- Abu El-Haj, Thea Renda (2007): "I was born here but my home isn't here": Educating for Democratic Citizenship in an Era of Transnational Migration and Global Conflict. In: *Harvard Educational Review*, 77, 3, pp. 285-316.
- Amos, Karin/Wiseman, Alexander/Rohstock, Anne (2014): EKSS-Change: Education and scientization as the dynamic forces of late modern 'knowledge societies'. A new order of things?: The scientization of late modern societies. Unpublished manuscript, Institute of Educational Sciences, University of Tübingen & College of Education, Lehigh University.
- Arendt, Hannah (1943): *We Refugees*. In: *Menorah Journal*, 31, 1, pp. 69-77.
- Arnstine, Donald (1995): *Democracy and the arts of schooling*. Albany, NY: SUNY Press.
- Azzolini, Davide/Schnell, Philipp/Palmer, John R.B. (2012): Educational Achievement Gaps between Immigrant and Native Students in Two "New" Immigration Countries: Italy and Spain in Comparison. In: *The Annals of the American Academy of Political and Social Science*, 643, pp. 46-77.
- Bajaj, Monisha/Bartlett, Lesley (2017): Critical transnational curriculum for immigrant and refugee students. In: *Curriculum Inquiry*, 47, 1, pp. 25-35. doi: 10.1080/03626784.2016.1254499
- Banks, James A. (2004): Multicultural Education: Historical Development, Dimensions, and Practice. In: Banks, James A./McGee Banks, Charry A. (eds.): *Handbook of Research on Multicultural Education*. 2nd edition. San Francisco, CA: Jossey-Bass, pp. 3-29.
- Bartlett, Lesley/Ghaffar-Kucher, Ameena (eds.) (2013): *Refugees, immigrants and education in the global south: Lives in motion*. New York, NY: Routledge.
- Beirens, Hanne/Hughes, Nathan/Hek, Rachel/Spicer, Neil (2007): Preventing social exclusion of refugee and asylum seeking children: building new networks. In: *Social Policy and Society*, 6, 2, pp. 219-229.
- Blizzard, Brittany/Batalova, Jeanne (2019): *Refugees and Asylees in the United States*. Washington, DC: Migration Policy Institute.
- Buckner, Elizabeth/Spencer, Dominique/Cha, Jihae (2018): Between Policy and Practice: The Education of Syrian Refugees in Lebanon. In: *Journal of Refugee Studies*, 31, 4, pp. 444-465.
- Bundeszentrale für politische Bildung [bpb] (2019a): *Asylanträge in Deutschland (Asylum applications in Germany)*.
<https://www.bpb.de/gesellschaft/migration/flucht/265708/asylantraege-und-asylsuchende>
- Bundeszentrale für politische Bildung [bpb] (2019b): *Asylentscheidungen und Klagen (Asylum decisions and complaints)*.
<https://www.bpb.de/gesellschaft/migration/flucht/265711/entscheidungen-und-klagen>
- Burnett, Kari (2013): *Feeling like an outsider: A case study of refugee identity in the Czech Republic*. New Issues in Refugee Research. Research paper no. 251. Geneva, Switzerland: UNHCR.

- Cobb-Clark, Deborah A./Sinning, Mathias/Stillman, Steven (2012): Migrant Youth's Educational Achievement: The Role of Institutions. In: *The Annals of the American Academy of Political and Social Science*, 643, pp. 18-45.
- Collet, Bruce A./Bang, Hyeyoung (2016): The Securitisation of Refugee Flows and the Schooling of Refugees: Examining the Cases of North Koreans in South Korea and Iraqis in Jordan. In: *Compare: A Journal of Comparative and International Education*, 46, 2, pp. 272-292.
- Damaschke-Deitrick, Lisa/Bruce, Elizabeth (2019): Education as a Panacea for Refugee Integration? Evidence from Germany and the United States. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka L./Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education: Dreams and Realities in Educational Systems Worldwide*. New York: Routledge.
- Damaschke-Deitrick, Lisa/Galegher, Ericka L./Park, Maureen F. (2019): Female refugees' transitions into higher education: Comparative perspectives from Germany, Egypt, and Kyrgyzstan. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka L./Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education: Dreams and Realities in Educational Systems Worldwide*. New York: Routledge.
- Dryden-Peterson, Sarah (2011): *Refugee Education: A Global Review*. New York: United Nations High Commissioner for Refugees.
- Dryden-Peterson, Sarah (2015): *The Educational Experiences of Refugee Children in Countries of First Asylum*. Washington, DC: Migration Policy Institute.
- Dryden-Peterson, Sarah (2017): Refugee education: Education for an unknowable future. In: *Curriculum Inquiry*, 47, 1, pp. 14-24.
- Dryden-Peterson, Sarah/Adelman, Elizabeth/Alvarado, Sagra/Anderson, Katelin/Bellino, Michelle J. et al. (2018): Inclusion of refugees in national education systems. Paper commissioned for the 2019 Global Education Monitoring Report Migration, displacement and education: Building bridges, not walls. Paris, France: UNESCO.
- Entorf, Horst/Minoiu, Nicoleta (2004): *PISA Results: What a Difference Immigration Law Makes*. IZA Discussion Paper No. 1021. Bonn, Germany: IZA.
- Fegert, Joerg M./Diehl, Claudia/Leyendecker, Birgit/Hahlweg, Kurt/Prayon-Blum, Valeria (2018): Psychosocial problems in traumatized refugee families: Overview of risks and some recommendations for support services. In: *Child and Adolescent Psychiatry and Mental Health*, 12, 5.
- Francis, Jenny/Yan, Miu-Chung (2016): Bridging the gaps: Access to formal support services among young African immigrants and refugees in Metro Vancouver. In: *Canadian Ethnic Studies*, 48, 1, pp. 77-100.
- Fratzke, Susan (2017): *Engaging communities in refugee protection: The potential of private sponsorship in Europe*. Brussels: Migration Policy Institute Europe.
- Hadfield, Kristin/Ostrowski, Aly/Ungar, Michael (2017): What can we expect of the mental health and well-being of Syrian refugee children and adolescents in Canada? In: *Canadian Psychology/Psychologie canadienne*, 58, 2, pp. 194-201.
- Joly, Danièle (2002): *Odyssean and rubicon refugees: Toward a typology of refugees in the land of exile*. In: *International Migration*, 40, 6, pp. 3-23.
- Kia-Keating, Maryam/Ellis, B. Heidi (2007): Belonging and connection to school in resettlement: Young refugees, school belonging, and psychosocial adjustment. In: *Clinical Child Psychology and Psychiatry*, 12, 1, pp. 29-43.

- Lucas, Tamara/Villegas, Anna Maria (2010): The Missing Piece in Teacher Education: The Preparation of Linguistically Responsive Teachers. In: Yearbook of the National Society for the Study of Education, 109, 2, pp. 297-318.
- McBrien, Jody Lynn (2005): Educational Needs and Barriers for Refugee Students in the United States: A Review of the Literature. In: Review of Educational Research, 75, 3, pp. 329-364.
- McBrien, Jody Lynn (2016): Refugees and asylum seekers. In: Peterson, Andrew/Hattam, Robert/Zembylas, Michalinos/Arthur, James (eds.): The Palgrave international handbook of education for citizenship and social justice. London: Palgrave MacMillan.
- Mendenhall, Mary/Bartlett, Lesley (2018): Academic and extracurricular support for refugee students in the US: Lessons learned. In: Theory into Practice, 57, 2, pp. 109-118.
- Mickan, Peter/Lucas, Katrina/Davies, Brien/Lim, Mi-Ok (2007): Socialisation and contestation in an ESL class of adolescent African refugees. In: Prospect, 22, 2, pp. 4-24.
- OECD (2012): Untapped Skills: Realising the Potential of Immigrant Students. Paris: OECD Publishing.
- OECD (2015): Immigrant Students at School: Easing the Journey towards Integration. Paris: OECD Publishing.
- Oliveira, Gabrielle/Becker, Mariana Lima (2019): Immigrant Latina youth and their education experience in the United States. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka/Park, Maureen F. (eds.): Comparative Perspectives on Refugee Youth Education: Dreams and Realities in Educational Systems Worldwide. New York: Routledge.
- Portes, Alejandro/Rumbaut, Rubén G. (1996): Immigrant America: A portrait. Berkeley, CA: University of California Press.
- Portes, Alejandro/Rumbaut, Rubén G. (2001): Legacies: The story of the immigrant second generation. Berkeley, CA: University of California Press.
- Radford, Jynnah/Connor, Phillip (2019): Canada now leads the world in refugee resettlement, surpassing the U.S. Pew Research Center.
<https://www.pewresearch.org/fact-tank/2019/06/19/canada-now-leads-the-world-in-refugee-resettlement-surpassing-the-u-s/>
- Sasnal, Patrycja (2015): Who are they? Two profiles of Syrian refugees. Warsaw: Polski Instytut Spraw Międzynarodowych.
- Schmidt, Michèle/Datnow, Amanda (2005): Teachers' sense-making about comprehensive school reform: The influence of emotions. In: Teaching and Teacher Education, 21, 8, pp. 949-965. doi: 10.1016/j.tate.2005.06.006
- Schneider, Britta (2018): Methodological nationalism in linguistics. In: Language Sciences. doi: 10.1016/j.langsci.2018.05.006
- Schnepf, Sylke (2007): Immigrants' Educational Disadvantage: An Examination across Ten Countries and Three Surveys. In: Journal of Population Economics, 20, 3, pp. 527-545.
- Sirin, Selcuk R./Rogers-Sirin, Lauren (2015): The educational and mental health needs of Syrian refugee children. Washington, DC: Migration Policy Institute.
- Spillane, James/Diamond, John B./Burch, Patricia (2002): Managing in the middle: School leaders and the enactment of accountability policy. In: Educational Policy, 16, pp. 731-762.

- Strekalova, Ekaterina/Hoot, James L. (2008): What is special about special needs of refugee children?: Guidelines for teachers. In: *Multicultural Education*, 16, 1, pp. 21-24.
- Taylor, Sandra/Sidhu, Ravinder (2012): Supporting Refugee Students in Schools: What Constitutes Inclusive Education? In: *International Journal of Inclusive Education*, 16, 1, pp. 39-56.
- Thomas, Rebecca Leela (2016): The Right to Quality Education for Refugee Children Through Social Inclusion. In: *Journal of Human Rights and Social Work*, 1, 4, pp. 93-201.
- UNESCO (2018): *Global Education Monitoring Report 2019: Migration, displacement, and education: Building bridges, not walls*. Paris, France. <https://en.unesco.org/gem-report/report/2019/migration>
- UNHCR (2017): *2017 Global Trends: Forced Displacement in 2017*. <https://www.unhcr.org/5b27be547.pdf>
- UNHCR (n.d.): What is a refugee? <http://www.unrefugees.org/what-is-a-refugee/>
- UNHCR (2019): *Figures at a glance*. <https://www.unhcr.org/ph/figures-at-a-glance>
- UNHCR (2020): *Figures at a Glance*. <https://www.unhcr.org/figures-at-a-glance.html>
- UNICEF (2016): *Broken dreams: Central American children's dangerous journey to the United States*. UNICEF Refugee and Migrant Crisis: Child Alert. [https://www.unicef.org/infobycountry/files/UNICEF_Child_Alert_Central_America_2016_report_final\(1\).pdf](https://www.unicef.org/infobycountry/files/UNICEF_Child_Alert_Central_America_2016_report_final(1).pdf)
- Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Bruce, Elizabeth/Davidson, Petrina/Stevens Taylor, Calley (2016): Transnational Scientized Education Discourse: A Cross-national Comparison. In: Schmid, Josef/Amos, Karin/Schrader, Josef/Thiel, Ansgar (eds.): *Internationalisierte Welten der Bildung – Bildung und Bildungspolitik im Globalen Vergleich*. Baden-Baden: Nomos, pp. 121-146.
- Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka/Park, Maureen F. (2019): The Contested Expectations of Education as a Panacea for Refugee Transitions. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka/Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education: Dreams and Realities in Educational Systems Worldwide*. New York: Routledge.
- Wiseman, Alexander W./Galegher, Ericka (2019): Teacher Preparation, Classroom Pedagogy, and the Refugee Crisis in National Education Systems. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka/Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education: Dreams and Realities in Educational Systems Worldwide*. New York: Routledge.
- Zetter, Roger (2007): More labels, fewer refugees: Remaking the refugee label in an era of globalization. In: *Journal of Refugee Studies*, 20, 2, pp. 172-192.

Language as a Predictor and an Outcome of Acculturation: A Review of Research on Refugee Children and Youth

Johanna Fleckenstein¹, Débora B. Maehler², Steffen Pötzschke³, Howard Ramos⁴ and Paul Pritchard⁵

1. Introduction

Language skills are of vital importance for the acculturation of immigrants because proficiency in the language of the host country plays a key role in social, educational, and occupational contexts. Despite the indisputable importance of language, its consideration in the acculturation of young refugees⁶ has been a blind spot in educational research (Behrensen/Westphal 2009; Liebau/Schacht 2016; Maehler/Pötzschke/Ramos/Pritchard/Fleckenstein 2020"). This is a major lacuna because sound research is needed for government agencies and service providers to offer evidence-based actions to support the educational and social integration of refugee children and youth in receiving countries.

This chapter presents a literature review of research on acculturation in the educational domain with a focus on language learning of refugee children and youth. The chapter aims to give a methodological overview of the existing research on the host country language skills of refugees, and to identify gaps in research that need to be addressed.

-
- 1 Johanna Fleckenstein is Researcher at the Leibniz Institute for Science and Mathematics Education (IPN) in Kiel. Email: fleckenstein@leibniz-ipn.de
 - 2 Débora B. Maehler is Senior Researcher and Head of the Research Data Centre PIAAC at GESIS – Leibniz Institute for the Social Sciences in Mannheim. Email: debora.maehler@gesis.org
 - 3 Steffen Pötzschke is Researcher at GESIS – Leibniz Institute for the Social Sciences in Mannheim. Email: steffen.poetzschke@gesis.org
 - 4 Howard Ramos is Professor at the Department of Sociology at the Western University in Ontario. Email: howard.ramos@uwo.ca
 - 5 Paul Pritchard is a PhD student in Sociology at the University of Toronto. Email: paul.pritchard@mail.utoronto.ca
 - 6 The Geneva Refugee Convention of 1951 defines a “refugee” as a person that seeks international protection (asylum) against political or other persecution and is unable to return to their country of origin (UNHCR 2017). We apply this broad understanding to our own arguments in this contribution.

2. The relevance of language skills for immigrants' acculturation into new societies

Proficiency in the language of the host country is a central issue in the education of immigrant and refugee students and for refugees too. Mastering the language of the host country plays a key role in the occupational integration of adults, and is associated with positive employment outcomes such as finding a job and earnings (Chiswick/Miller 2002; Shields/Price 2002), and for successful social integration (Martinovic/van Tubergen/Maas 2009). Immigrants' language proficiency also has important consequences for the integration of their children as parents' language skills influence the educational and occupational careers of their offspring (Heath/Rothon/Kilpi 2008). Thus, taking a closer look at the determinants of immigrant language learning is a highly relevant endeavor and is key to understanding refugee integration.

Most studies on the determinants of immigrants' language acquisition focus on individuals who migrated for labor or family related reasons, while the language skills of refugees have been left largely unexamined (Fennelly/Palasz 2003; Van Tubergen 2010). Due to the specific characteristics associated with forced migration, researchers cannot assume the same patterns occur for refugees than for immigrants, as they experience profoundly different pre-migration and post-migration issues that affect the process of settlement. For instance, displaced immigrants may not have the opportunity to learn the language of the host country in advance, may have experienced traumatic events, and may face limitations because of their legal status in the new country. Each of these factors pose particular challenges that may affect the process of language acquisition.

Across disciplines, researchers find three general mechanisms that underlie immigrants' acquisition of the host language (Chiswick/Miller 2007; Esser 2006). These mechanisms are associated with language exposure, economic incentives, and the efficiency with which immigrants learn new languages. These are operationalized through observable individual and contextual determinants of language proficiency, for example, age, sex/gender, and length of stay in the host country (Carliner 2000; Chiswick/Miller 2001, 2007; Hwang/Xi 2008; Stevens 1999; van Tubergen/Kalmijn 2009). A growing body of literature also investigated the determinants and correlates of refugees' host language skills and whether or how they differ from other immigrants. Van Tubergen (2010), for example, found that the main factors relevant for the language acquisition of family and labor immigrants are also predictive of refugees' language skills (i.e., age at arrival, educational background, sex/gender, length of stay in the host country, and settlement intentions). Other studies have also investigated the language skills of young refugees. For example, Liebau and Schacht (2016) found that the language

proficiency of refugees in Germany was comparable to that of non-refugee immigrants. While host language skills at the time of arrival may lag behind, immigrants with refugee backgrounds close the gap over time. A number of characteristics have been found to be positively associated with language proficiency. These include being younger at the age of arrival and possessing a stronger educational background. Post-migration factors that positively affect language acquisition include a longer length of stay in the host country, higher rates of participation in the host country's education system, and higher frequency in the usage of the host country's dominant language.

3. Reviewing studies on the language skills of refugees

Research on the integration of young refugees identifies several methodological shortcomings largely due to a lack of consistency in the operationalization of key concepts and inconsistencies in methodological approaches (Allen/Vaage/Hauß 2006; Pritchard/Maehler/Pötzschke/Ramos 2019). Based on the findings reported by Van Tubergen (2010), we investigated (1) whether studies on young refugees' language skills report information on acculturation factors at both the individual and macro-level, e.g., age, sex/gender, length of stay, educational background, country of origin, host country. We also discussed (2) the central findings of these studies.

To this end, we reviewed and analyzed 22 out of 178 peer-reviewed articles that are available on the Education Resources Information Center (ERIC) database. Studies included in our review look at individuals aged 19 and younger and were published between 1987 and 2016. The sample for our literature search was constructed through a multilevel set of inclusion criteria, consisting of key search terms grouped by three levels and used in combination with each other. The first level of search terms served to define the target group ("refugees"); the search terms at the second level delimited the desired age range (e.g., "child", "adolescent"); the terms at the third level comprised several keywords relevant to language and learning. Only documents containing at least one keyword from each of the three levels were retained. The search yielded a working sample of 421 English-language articles that constituted the broad basis for further selection and coding. The selection and coding procedure followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model (Moher/Liberat/Tetzlaff/Altman/The PRISMA Group 2009). In two rounds of filtering, duplicates, non-peer-reviewed-articles, articles not published in the target languages concerning divergent target groups (not refugees or not within the specified age range), literature reviews, and non-empirical contributions were removed. This left 178 English-language publications carried out in educational

contexts. From these, we kept all publications that included “language” OR “literacy” as a variable in the study. The search yielded 22 English-language studies that met the selection criteria.

4. Characteristics of research on the language skills of young refugees

All of the articles we look at were published between 1999 and 2015 – despite the search going back to 1986. In reviewing the methodological approaches of these studies, we found that eleven studies used qualitative research methods and six used quantitative methods, exclusively; while five used mixed research methods. The research designs employed in these studies were most frequently cross-sectional (ten), followed by ethnographic (five) and case studies (two). Only four studies used a longitudinal research design. Most studies (ten) were based on small samples between $n=3$ and $n=20$, six were based on medium-sized samples between $n=56$ and $n=110$, and only three studies were based on large samples between $n=182$ and $n=1051$. In three studies, the sample size was not specified at all.

We next report the degree to which the 22 studies analyzed provided information on background characteristics found to be significant to the process of language learning: age, sex/gender, length of stay, educational background, country of origin, and host country. Each is a characteristic identified as key to acculturation by Van Tubergen (2010). Our assessment finds that only 15 studies specified the age-range of participants, three studies specified school grade, and four did not provide any indication regarding the age of individuals in their sample at all. The sex/gender of the participants was reported in 13 studies, 11 of which investigated both male and female refugees. Only two studies focused on female or male children only, one on each. The duration of the refugees’ residence in the host country was reported in eight of the studies, while the other 14 did not provide any information on the length of stay. A minority of studies mentioned details on the educational background of the sample: Nine studies provided details on prior education in the host country and/or in the country of origin. As Table 1 reports, the regional/national origin or ethnic group of the participants was specified in 14 studies. Most of the studies investigated young refugees of African and Southeast Asian origin. Field work was most frequently conducted in the United States ($n=8$), followed by Australia ($n=7$), Canada ($n=5$), the United Kingdom ($n=2$), Scotland ($n=1$), Greece ($n=1$), and Colombia ($n=1$). The geographical distribution of studies and the regional/national or ethnic groups

most studied reflects the countries/languages of studies found via ERIC (which includes publications in English only).

Table 1. Host country and nationality/ethnic group

Host country	Nationality/Ethnic group
Australia	African descent
Australia	Afghan; Sudanese
Australia	Dinka Sudanese
Australia	Sudanese
Australia	<i>diverse or not specified</i>
Australia; Hong Kong	<i>diverse or not specified</i>
Canada	Somali
Canada	<i>diverse or not specified</i>
Canada	Hispanic
Canada	Karen; Iranian
Canada	Chinese; South Asian
Colombia	<i>diverse or not specified</i>
Greece	<i>diverse or not specified</i>
UK	Pakistani; Indian; Somali; Congolese
UK	<i>diverse or not specified</i>
US	Soviet Jewish
US	Soviet Jewish
US	Sudanese
US	Cambodian
US	Karen; Burmese Muslim; Burmese Karen; Poe Karen
US	Vietnamese
US	Vietnamese

5. Content analysis of studies on the language skills of young refugees

Only two studies in our sample statistically analyzed predictors that were specified in our first research question: The first is a longitudinal study by Birman and Trickett (2001) that focused on the acculturation of first-generation Soviet Jewish refugee adolescents and their parents who resettled in the United States. The study examined the contributions of parent education, sex/gender, age of migration, and length of residence in the country for both children and adults in predicting language acculturation. The study results show that the age of arrival in the host country significantly predicted first and second language skills for adolescents (the earlier the better) but not for adults. Of the other variables analyzed, only the degree of parent education was predictive for second language proficiency (the higher the better).

The second study by Mitakidou, Tourtouras, and Tressou (2008) aimed to compare the performance in language and mathematics of 1,051 repatriate and refugee children from the former USSR, who started school in Greece in first grade, with children of the same group, who joined school at a later grade. Contrary to the authors' hypothesis their findings showed that immigrant children who started school in Greece performed better than their peers who arrived to Greece at a later age and/or entered at a higher grade in Greek schooling. However, these findings should be interpreted carefully as prior school experience is confounded with age of arrival and length of stay.

All of the other studies investigated language as an outcome of the acculturation process using qualitative methods. We find ethnographic descriptions of other factors that contribute to language acquisition, such as refugee-specific educational programs (e.g., a gardening program or after-school homework tutoring centers) that provide the opportunity to learn the host-country's language (Cutter-Mackenzie 2009; Naidoo 2008). Furthermore, some qualitative studies used action research methods to investigate particular instructional approaches; often they had very small samples (including single case studies). The instructional approaches aimed to foster the development of host-country language skills through the use of visual texts (Arizpe/Bagelman/Devlin/Farrell/McAdam 2014), digitally supported process drama (Dunn/Bundy/Woodrow 2012), and differentiated instruction (Niño Santisteban 2014). None of these studies presented data on the variables that were specified in our first research question (age, sex/gender, length of stay, educational background, country of origin, host country).

Moreover, studies investigated language not as an outcome but as a predictor or validation criterion. Trickett and Birman (2005), for example, focused on (self-rated) English and Russian language competence as a predictor of school outcomes (Grade Point Average, disciplinary infractions,

school belonging). Nguyen, Messé, and Stollak (1999) used English and Vietnamese language skills as an external criterion for their validation of an acculturation scale. Poppitt and Frey (2007) identified the concern over English language proficiency as the main source of acculturative stress in a qualitative study. Again, these studies do not investigate any of the variables in our first research question.

6. Conclusion

Our results show that there is a general dearth of research in the field of young refugees' language skills. First, the literature review showed that only some of the studies we identified included information on relevant predictors of refugees' second language acquisition, whereas some did not specify age, gender, length of stay, educational background, or country of origin of their sample. Second, a content analysis showed that only very few publications present quantitative analyses on these factors and their relation to refugees' language learning. More studies that consider these variables are needed in order to establish a solid base for educational policy and practice. In particular, longitudinal studies with large sample sizes are missing.

In line with the call to action by several international agencies (UNICEF et al. 2018) we recommend taking into account individual-level variables (e.g., age, gender) as well as macro-level variables (e.g., receiving country) and using longitudinal research designs to study refugee acculturation processes over time. Based on prior research (e. g., Van Tubergen 2010), among the relevant variables to be considered in future research are age, country of origin, host country, gender, length of stay, and educational background. Last but not least, the findings of this review can provide guidance for further studies dealing, for example, with the large number of children and adolescents that came to the European Union during the so-called European refugee crisis of 2015/16. More high-quality research in the domain of language and literacy can lead to evidence-based educational policy and practice.

References

- Allen, James/Basilier Vaage, Aina/Hauff, Edvard (2006): Refugees and asylum seekers in societies. In: Sam, David L./Berry, John W. (eds.): *The Cambridge handbook of acculturation psychology*. Cambridge, UK: Cambridge University Press, pp. 198-217.
- Arizpe, Evelyn/Bagelman, Caroline/Devlin, Alison M./Farrell, Maureen/McAdam, Julie E. (2014): Visualizing intercultural literacy: engaging critically with diversity and migration in the classroom through an image-based approach. In: *Language and Intercultural Communication*, 14, pp. 304-321. doi: 10.1080/14708477.2014.903056
- Behrensen, Birgit/Westphal, Manuela (2009): Junge Flüchtlinge – ein blinder Fleck in der Migrations- und Bildungsforschung. Bildung junger Flüchtlinge als Randthema in der migrationspolitischen Diskussion. In: Krappmann, Lothar/Lob-Hüdepohl, Andreas/Bohmeyer, Axel/Kurzke-Maasmeier, Stefan (eds.): *Bildung für junge Flüchtlinge – ein Menschenrecht. Erfahrungen, Grundlagen und Perspektiven*. Bielefeld: Bertelsmann, pp. 45-58.
- Birman, Dina/Trickett, Edison (2001): Cultural transitions in first-generation immigrants: Acculturation of Soviet Jewish refugee adolescents and parents. In: *Journal of Cross-Cultural Psychology*, 32, pp. 456-477. doi: 10.1177/0022022101032004006
- Bloch, Alice/Chimienti, Milena/Counilh, Anne-Laure/Hirsch, Shirin/Tattolo, Giovanna/Ossipow, Laurence/Wihtol de Wenden, Catherine (2015): The children of refugees in Europe: Aspirations, social and economic lives, identity and transnational linkages. Final report – working paper. Geneva, Switzerland: SNIS.
- Carliner, Geoffrey (2000): The language ability of U.S. immigrants: Assimilation and cohort effects. In: *International Migration Review*, 34, 1, pp. 158-82.
- Chiswick, Barry R./Miller, Paul W. (2001): A model of destination-language proficiency acquisition: Application to male immigrants in Canada. In: *Demography*, 38, pp. 391-409.
- Chiswick, Barry R./Miller, Paul W. (2007): *The economics of language: International analysis*. London: Routledge.
- Cutter-Mackenzie, Amy (2009): Multicultural school gardens: Creating engaging garden spaces in learning about language, culture, and environment. In: *Canadian Journal of Environmental Education*, 14, pp. 122-135.
- Dunn, Julie/Bundy, Penny/Woodrow, Nina (2012): Combining drama pedagogy with digital technologies to support the language learning needs of newly arrived refugee children: a classroom case study. In: *RiDE: The Journal of Applied Theatre and Performance*, 17, pp. 477-499. doi: 10.1080/13569783.2012.727622
- Esser, Hartmut (2006): *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt, Germany: Campus Verlag.
- Fennelly, Katherine/Palasz, Nicole (2003): English language proficiency of immigrants and refugees in the Twin Cities metropolitan area. In: *International Migration*, 41, pp. 93-125.
- Grant, Maria J./Booth, Andrew (2009): A typology of reviews: An analysis of 14 review types and associated methodologies. In: *Health Information and Libraries Journal*, 26, pp. 91-108.

- Heath, Anthony F./Rothon, Catherine/Kilpi, Elina (2008): The second generation in Western Europe: Education, unemployment, and occupational attainment. In: *Annual Review of Sociology*, 34, pp. 211-35.
- Hwang, Sean-Shong/Xi, Juan (2008): Structural and Individual Covariates of English Language Proficiency. In: *Social Forces*, 86, pp. 1079-1104.
- Liebau, Elisabeth/Schacht, Diana (2016): Language acquisition: Refugees nearly achieve proficiency level of other migrants. DIW Economic Bulletin. Berlin, Germany: DIW.
- Maehler, Débora B./Pöttschke, Steffen/Ramos, Howard/Pritchard, Paul/Fleckenstein, Johanna (2020): Studies on the acculturation of young refugees in the educational domain: A scoping review of research and methods. *Adolescent Research Review*. <https://doi.org/10.1007/s40894-019-00129-7>
- Martinovic, Borja/van Tubergen, Frank/Maas, Ineke (2009): Dynamics of interethnic contact: A panel study of immigrants in the Netherlands. In: *European Sociological Review*, 25, pp. 303-18.
- Mitakidou, Soula/Tourtouras, Christos/Tressou, Evangelia (2008): Performance of students from the former USSR in Greek schools. In: *Journal of Early Childhood Research*, 6, pp. 163-187. doi: 10.1177/1476718X08088676
- Moher, David/Liberat, Alessandro/Tetzlaff, Jennifer/Altman, Douglas G./The PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6, 7: e1000097. doi: 10.1371/journal.pmed.1000097
- Naidoo, Loshini (2008): Supporting African refugees in Greater Western Sydney: a critical ethnography of after-school homework tutoring centres. In: *Educational Research for Policy and Practice*, 7, pp. 139-150. doi: 10.1007/s10671-008-9046-1
- Niño Santisteban, Liliana (2014): The effects of differentiated instruction on the literacy process of learners with interrupted schooling. In: *Gist Education and Learning Research Journal*, 9, pp. 31-49.
- Nguyen, Huong H./Messé, Lawrence A./Stollak, Gary E. (1999): Toward a more complex understanding of acculturation and adjustment: Cultural involvements and psychosocial functioning in Vietnamese youth. In: *Journal of Cross-Cultural Psychology*, 30, 1, pp. 5-31.
- Pritchard, Paul/Maehler, Débora B./Ramos, Howard/Pöttschke, Steffen (2019): Integrating refugee child and youth: A scoping review of English and German research literature. In: *Journal of Refugee Studies*. doi: 10.1093/jrs/fez024
- Poppitt, Gillian/Frey, Ron (2007): Sudanese adolescent refugees: Acculturation and acculturative Stress. In: *Australian Journal of Guidance & Counselling*, 17, pp. 160-181.
- Shields, Michael A./Price, Stephen Wheatley (2002): The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. In: *Journal of Population Economics*, 15, pp. 137-60.
- Stevens, Gillian (1999): Age at immigration and second language proficiency among foreign born adults. In: *Language in Society*, 28, 4, pp. 555-578.
- Tricket, Edison J./Birman, Dina (2005): Acculturation, School context, and school outcomes: Adaptation of refugee adolescents from the former Soviet Union. In: *Psychology in the Schools*, 42, pp. 27-38. doi: 10.1002/pits.20024

- Van Tubergen, Frank (2010): Determinants of second language proficiency among refugees in the Netherlands. In: *Social Forces*, 89, 2, pp. 515-534.
- Van Tubergen, Frank/Kalmijn, Matthijs (2005): Destination-language proficiency in cross-national perspective: A study of immigrant groups in nine Western countries. In: *American Journal of Sociology*, 110, 5, pp. 1412-57.
- UNHCR (2017): Global trends. Forced displacement in 2016.
<http://www.unhcr.org/statistics/unhcrstats/5943e8a34/global-trends-forced-displacement-2016.html>
- UNICEF/IOM/UNHCR/Eurostat/OECD (2018): A call to action: Protecting children on the move starts with better data. https://data.unicef.org/wp-content/uploads/2018/02/Migration_advocacy_brief.pdf

Attitudes Towards Refugees. A Case Study on the Unfolding Approach to Scale Construction

Michael Filsecker¹ and Hermann Josef Abs²

1. Introduction

In 2015, nearly 5 million people migrated to Europe, making the issue of migration and refugee status a source of high concern among Europeans (European Commission 2016). Countries such as Germany, Austria, Hungary and Sweden, which have experienced the largest influx of refugees, have also shown a decline in public support of generous immigration policies towards refugees (Heath/Richards 2019). Since 2015, Germany has received 1,524,205 first-time asylum applications from non-EU countries – mainly Syria (34%), Afghanistan (11%) and Iraq (12%). Most asylum seekers were men (69%), and 21% were boys under the age of 18 years. Girls under the age of 18 accounted for 16%.³ The German federal government launched in 2016⁴ a strategic plan to counteract violent acts against specific groups and the “specific attitudes underlying” these acts. Citizenship education plays an important role in this endeavor. Subject to different projects in schools, citizenship education targets a facilitation of democratic attitudes and counteraction of extremist or negative ideas. Clearly, educators and researchers alike face at least a twofold challenge, that of educating for citizenship and that of understanding attitudes towards migration in the context of a financial crisis, an environmental crisis and the recent “refugee crisis” experienced in Europe (Heath/Richards 2019; Jetten/Esses 2018; Schulz et al. 2017). In this context, this chapter represents an effort to contribute to the understanding of attitudes towards refugees, by arguing for the need to develop better measurements that enable to identify specific groups in need of intervention. Without more specific attitude measurements, the effectiveness of such interventions is difficult to assess. We first highlight why attitudes are important, and then show

-
- 1 Michael Filsecker is Researcher at the Department of Psychology at the University of Erfurt. Email: michael.filsecker_wagner@uni-erfurt.de
 - 2 Hermann Josef Abs is Professor of Educational Sciences at the University Duisburg-Essen. Email: h.j.abs@uni-due.de
 - 3 Own calculation based on the data from Eurostat. <http://appsso.eurostat.ec.europa.eu/nui/show.do>
 - 4 Bundesministerium für Familie, Senioren, Frauen und Jugend (2016): Strategie der Bundesregierung zur Extremismusprävention und Demokratieförderung: <https://www.bmfsfj.de/blob/109002/5278d578ff8c59a19d4bef9fe4c034d8/strategie-der-bundesregierung-zur-extremismuspraevention-und-demokratieforderung-data.pdf>.

what the limitations of current scale development practices are for measuring attitudes using an example from the latest cycle 2016 of the International Civic and Citizenship Study (ICCS 2016)⁵. Next, we show how developing a scale with intermediate items could help deal with some of the current limitations, and why these scales yield a different distribution of attitudes towards refugees on a sample of secondary students. We finally draw some implications for attitude measurement and suggest future lines of research.

2. Why attitudes matter: migration – attitudes – integration

Facts and their evaluation, in our case migration and attitudes towards migration, are two sides of the same coin. The origins of the scholarly interest in attitudes can be traced back to events that occurred in the US in the second half of the twentieth century. At that time, the US experienced several immigration waves that led to social conflict in terms of overt legal discrimination against migrants and social violence in the form of riots, killings and property destruction (Wark/Galliher 2007). In this context, the term “race attitudes” was popularized by the sociologist Emory Borgadus, who created the first attitude scale, called the social distance scale, to capture quantitatively the degree of “intimacy and understanding” that usually governs the interaction between individuals or social groups. The assumption was that “hostile” attitudes were the prerequisite of prejudice, discrimination and violence (Allport 1954). Under the same logic, Germany today, after the so-called “refugee crisis” in 2015, developed a set of governmental initiatives to counteract violent acts and discrimination and to understand the “extremist attitudes” underlying such violence. In the political arena, the idea of “public opinion” [i.e. attitudes] and relevance in democratic societies was also a concern (e.g., Allport/Hartman 1951) and is today a main strategic goal for integration policies because “...without managing public perception [e.g. attitudes], it is difficult or even impossible to manage migration, especially on a European level” (Beutin et al. 2007: 390). This role of attitudes has been echoed more recently by the International Organization for Migration (OIM), which in 2013 called “for a fundamental shift in the public perception of migration” with an emphasis on the “important role migrants can and do play as partners in host and home country development” (OIM 2013: 4)⁶. This reflects a top-down political strategy trying to frame the discourse on migra-

5 ICCS 2016 assessed students (13.5 years old) from 25 countries enrolled in the eighth school grade. For more information see the website of the International Association for the Evaluation of Educational Achievement (IEA): <https://www.iea.nl/iccs>

6 Extracted from <https://www.iom.int/files/live/sites/iom/files/What-We-Do/docs/IOM-Position-Paper-HLD-en.pdf>

tion in terms of valuing diversity and conceiving the receiving societies as a “welcoming culture”, which in turn could lead to more positive attitudes towards migrants and refugees. From the perspective of democracy, positive attitudes are vital for keeping it healthy by acting as glue that can sustain social cohesion (Chan et al. 2006). Finally, attitudes seem to be ubiquitous (Allport 1954): Nationals may have negative attitudes towards Muslims (Wirtz/van der Pligt/Doosje 2016) or immigrants in general (Jetten/Esses 2018), Muslims against nationals (Vedder/Wenink/van Geel 2016; Maliepaard/Verkuyten 2018), or well-integrated immigrants may have negative attitudes towards the host society, “integration paradox” (De Vroome/Martinovic/Verkuyten 2014) and so on. All these attitudinal tendencies, including “both the migrants’ identification with the receiving society and the receiving society’s inclusive attitudes and acknowledgement of cultural heterogeneity” (Beutin et al. 2007), may present difficulties for cultural integration. Nevertheless, these and other attitudes can be changed at least in the short-term (Lai et al. 2016). And education systems are a central actor in this endeavor (e.g., Schachner/Van de Vijver/Noack 2018).

In the following, we briefly discuss the challenges of measuring such a central construct as attitudes towards refugees.

3. Challenges and limitations in the measurement of attitudes

As relevant outcomes in education and as key factors in public opinion for integration purposes, attitudes and other non-cognitive constructs (e.g., interests, motivation) need better measurement (e.g., Danner et al. 2016; Filsecker 2019). Indeed, we need better measurements in order to understand the formation of attitudes in social life and the effectiveness of specific interventions aiming at changing attitudes in specific populations.

Researchers have mostly measured attitudes on the basis of self-reports with Likert-type items. The procedure to get the final set of items is well known: First, assisted by experts, relevant literature and early qualitative approaches (e.g., focus groups), researchers determine the possible areas or aspects of the construct of interest. Several items are then in a second step written according to Likert (1932); that is, items should be short, not double-barrel or ambiguous, relatively extreme, and formulated as positive and negative (the latter are afterwards reverse-scored). In a third step, the selection of the items during a pilot version is based on factorial analysis and classical test theory (i.e., items with high item total correlations, no factor cross-loadings are kept showing high internal consistency) and scaling

methods such as Partial Credit Models (Master 1982). This procedure has been used in large-scale assessments such as the International Civic and Citizenship Study (Schulz et al. 2017), the Programme for International Student Assessment (PISA) (OECD 2017), and also in applied research developing attitude measures, such as the acculturation attitude scale (Berry et al. 1989) or the attitudes towards integration of refugees (Beverluis et al. 2016).

Given the ubiquity of this approach for scale development, it is reasonable to ask what might be wrong with it and secondly, what can be done better. However, before turning to the scale development process, a few key ideas of psychometrics need to be explicated. First, measurement is the effort of locating individuals and items in the same imaginary trait continuum. Second, this continuum is not directly observable and must be measured indirectly through different items. Third, respondents and items interact with one another in producing a response process (cf., discriminant process). Fourth, this response process is characterized by an item-response function (IRF), which defines the probability of an individual – given his/her ability level – of answering an item correctly or agreeing to an item. Finally, there are two main assumed response processes: a dominance process with S-shape IRFs and an unfolding process with bell-curved shape IRFs.

Returning to ubiquitous scale development, it can be said that most of such developments in attitude research assumed a dominance process. We argue that this dominance process is inappropriate for measuring attitudes, and that the unfolding process is more suitable for such endeavors. Indeed, dominance processes prescribe a monotonic relationship between individuals' trait level and their scale scores. That is, if we locate both person and items on a continuum representing an attitude, then a person will endorse an item when her or his position on the continuum is more positive than that of the item. Applied to attitude items, it means that individuals will agree with a positive item (e.g., *Moving/<Immigrant> children should have the same opportunities for education*, ICCS 2016) if their position on the latent attitude continuum is higher than the position of the item on the continuum. In the context of cognitive ability measures (e.g., intelligence, problem solving, achievement), the assumed monotonic relationship seems appropriate: When people face an ability or knowledge item, the difficulty of the item is a burden that individuals need to overcome using as much of their available ability as possible – this was called “maximal” performance by Cronbach (1949). These observations have led to two basic ideas: 1) The higher the ability of the person, the more likely she or he is to answer an ability item correctly or to agree with an attitude item; 2) the item difficulty parameter indicates on which trait level a person has a 50% probability of answering the ability item correctly or agreeing to an attitude item.

On the other hand, attitude items apparently evoke other types of activities or response processes in the respondents. Respondents may compare themselves with the item and decide to what extent the position of the item coincides with their own location on the attitude continuum. Respondents' trait and probability of endorsing an item would then follow a bell-curved *non-monotonic* relation. That is, an unfolding process is operating. This phenomenon was said to be ubiquitous for preference data (Coombs/Avrunin 1950; cf. typical performance behavior: Cronbach 1949), such as attitudes and personality constructs as opposed to the achievement/maximal performance data of dominance models. These ideas have important implications for scale development. For the unfolding process, the trait level and the probability of agreeing to an item form a *non-monotonic* bell-curved ("single-peak") relation in which the highest (maximum) possible probability of answering correctly or agreeing to an item occurs when the location of the person (i.e., their "ideal point") and the location of the item coincide (i.e., the difference between the two locations is zero). This probability decreases to the right and left of this "ideal point" as the item and person location increasingly diverge. In scales developed under the dominance approach (i.e., a list of similar positive-worded items, see Figure 1, items 1-5), the more positive the attitude of individuals, the higher the number of items they will agree with compared to individuals with less positive attitudes. By contrast, within the unfolding approach a person with a very positive attitude will not necessarily agree with more items. For example, in a scale assessing attitude towards refugees with items representing low, intermediate and high trait values, a person will more likely agree with extremely positive items (e.g., *Immigrants should have the same rights that everyone else in the country has*, ICCS 2016) but less likely to agree with moderate (e.g., *I can't totally agree with "same rights for every immigrant"*) and negative items (e.g., *Refugees should not have the right to get cash from the state*). Last but not least, the unfolding approach recognizes the fact that someone may *disagree* with an item for two reasons: a person disagrees with an item because they perceive their location on the attitude continuum to be higher than that of the item ("disagree from above"); or because they perceive their location to be lower than that of the item ("disagree from below"). As Andrich puts it: *"Thus there are two latent responses which produce the single manifest Disagree response in the unfolding direct-response design"* (Andrich 1996: 350, emphasis in original). On the contrary, in the dominance process, when a person disagrees with a positively worded item, the direction of the attitudes is immediately assumed to be a negative one (this is also true for negatively formulated items given that they are later reverse-scored). Considering the theoretical differences just presented, we will discuss in the following some limitations of assuming a dominance process in non-cognitive constructs such as attitudes.

Impaired precision. High item-total correlations and clean factor loadings are possible if several items are similar and extremely positively/negatively formulated. Therefore, items reflecting a more mixed or ambivalent attitude, which we call “intermediate items”, are from the beginning discarded as “poor items”, because they are unlikely to show the expected statistical properties of high item-total correlations and no factorial cross-loadings (Davison 1977). Discarding such intermediate items leads to a reduced measurement precision in specific ranges of the attitude trait. It has been shown that intermediate items are more accurate at the lower/higher ends of the trait continuum than the typical positive Likert-type items (e.g., Roberts/Laughlin/Wedell 1999). This is an important property if via large-scale surveys we want to detect the respondents showing moderate to extremely negative attitudes towards refugees. Given the current practices of item development, one solution would be to include more extremely negative items, but this type of item is seldom endorsed by respondents. In short, an impoverished initial item pool leads to less measurement precision in traits levels that are relevant for possible interventions with specific populations.

Social desirability. On the other hand, extremely *positive* items are usually endorsed by almost every respondent (see item 5, Figure 1). This is probably not due to the actual value of respondents on the attitude trait, but to systematic error due to social desirability, a ubiquitous problem in citizenship education research (Ten Dam/Geijsel/Ledoux/Meijer 2013). We argue that current scale development practices, not only in citizenship education research, elicit – by design – such socially desirable responses by creating extremely positive Likert-type items, which respondents find almost impossible to disagree with. This is a fundamental flaw in attitude research that undermines theoretical and empirical efforts to understand the phenomena of attitude formation and change. A lot of effort later needs to be invested in order to “clean” this measurement error that is inserted by design in the current practices of scale development – technically this is done by “common method variance” (e.g., Miller/Ruggs 2014). We turn to this issue and its possible solution in the last paragraph on future lines of research.

Inefficiency. Inclusion of only relatively extreme positive items leads to a sort of inefficiency, given that these items are located in practically the same place on the attitude continuum. For example, in ICCS 2009 three items addressing the issue of attitudes towards immigrant rights were located within -2.64 and -2.06, that is .58 logit of distance. For long and time-consuming large-scale surveys (such as ICCS 2016) with considerable non-response rates, this issue of efficiency is crucial for a successful implementation of such surveys (Stanton/Sinar/Balzer/Smith 2002). A more efficient way might have been to have one item covering the entire range and the other two items covering other areas of the attitude trait. The goal here should be to employ a smaller number of items and cover a wider range of the latent continuum.

This can be achieved by the inclusion of intermediate items and their modeling under the unfolding approach.

Impaired validity. Scoring approaches that do not consider the response process can lead to different results by ranking the persons differently (Stark/Chernyshenko/Drasgow/William 2006). For example, the items developed as a national option here were scored using the dominance and unfolding approach. The correlation between the total scores and the dominance scoring was .99; however, the correlation of total scores with the unfolding scoring was .16. These discrepancies are due to the presence of intermediate items which were included in the scale (such items can unintendedly appear in a traditionally developed scale). If, for example, our attitude scale is used to detect persons with negative attitudes towards refugees, different individuals would have been selected when scored under the unfolding model as compared to any dominance model such as total scores.

Impaired construct validity. When measuring attitudes towards refugees and migrants it is important to relate such attitudes to some criteria such as national identity, prejudice or discriminatory versus prosocial tendencies. It is also important to establish possible predictors of such attitudes, such as the perception of threat and other beliefs and values (e.g., conservatism, religiosity, identity). In order to do so, researchers employ regression analysis and structural equation modeling. These valuable techniques are more useful for handling or uncovering linear rather than curvilinear relations (e.g., Carter et al. 2014). On the other hand, unfolding approaches are more flexible for discovering relationships (linear or nonlinear) and can help advance the field of attitude and attitude change by testing linear and curvilinear relations within the hypothesized nomological network among relevant constructs (Cronbach/Meehl 1950) and not by assuming beforehand linear relations among constructs.

In summary, the application of dominance methods to preference data may have negative implications for the measurement precision of the instrument, can result in misled conclusions from traditional statistical analysis, and misguide theoretical development in the area of attitude research. We first aimed at an improvement by exploring the development of scales explicitly incorporating intermediate items. In the following, we briefly describe the development process and the main results of our efforts.

4. Development of intermediate attitude items as national option in ICCS 2016⁷

The ICCS 2016 study considered two types of items: items that are international, meaning they are administered in all the participant countries, and optional items that only a specific country uses within its national sample. We refer to the former as “ICCS 2016 items” and to the latter as “National Option items”. Germany’s National Option entailed a set of “intermediate” items that were produced for assessing students’ attitude toward refugees.

In line with Andrich (1996), we argue that attitudes are complex and entail compromise, negotiation, and reconciliation of interests that usually compete with each other. Therefore, appropriate scales must contain items expressing such core features of attitudes. We refer to such items as “intermediate” or “neutral” because they are supposed to reflect these tensions, and technically they are located at around the center of the attitude continuum. In the context of the main study of ICCS 2016, we included such items concerning students’ attitudes toward refugees.

4.1 *Developing the item pool*

For the purpose of preparing a short scale with intermediate and non-intermediate items ordered a priori, we implemented the following strategy: First, we searched for published scales developed under the unfolding mechanism described above, and identified the intermediate items that serve as our point of departure for writing our own items reflecting individuals’ attitudes toward refugees. For example, we identified items reflecting attitudes toward church (e.g., Sometimes I feel the church and religion are necessary and sometimes I doubt it; Thurstone/Chave 1929: 33), items reflecting attitudes toward capital punishment (e.g., I do not believe in capital punishment, but I am not sure it is not necessary; Andrich 1995: 277) and items reflecting attitudes toward abortion (e.g., I cannot whole-heartedly support either side of the abortion debate; Roberts et al. 2000: 20).

Second, we analyzed all these items in terms of structure and semantic properties, and adapted them for the purpose of measuring attitude toward refugees. Third, we drafted items that should reflect the sort of ambivalence and competing interest suggested by Andrich (1996) and reflected in the items chosen as examples. Fourth, we – the authors and a third colleague – independently ordered the items we developed in terms of their hypothetical

7 ICCS 2016 entails two types of items. Items that are administered in all countries (here “ICCS 2016 items”) and optional items that a specific country administers to its national sample (here “National option items”).

location on the presumed attitude continuum. Fifth, we compared the ordering of the newly developed items that each of us generated, discussed some discrepancies, and finally compiled a list of agreed upon items. Table 1 summarizes the items and their descriptive statistics. For comparison purposes, the ICCS 2016 items were also included in the figure.

Table 1. Mean, standard deviation and percentage of agree-disagree responses to the national and international items assessing attitude towards refugees and migrants

	Items	<i>M</i>	<i>SD</i>	<i>D-SD</i>	<i>A-SA</i>
ICCS 2016	1. Moving/<Immigrants> should have the opportunity to continue speaking their own language	2.18	0.83	17.70	82.30
	2. Moving/<Immigrant> children should have the same opportunities for education	2.62	0.63	5.62	94.38
	3. Moving/<Immigrants> who live in a country for several years should have the opportunity to vote	2.22	0.80	17.10	82.90
	4. Moving/<Immigrants> should have the opportunity to continue their own customs and lifestyle	2.15	0.81	19.47	80.53
	5. Moving/<Immigrants> should have the same rights that everyone else in the country has	2.56	0.69	7.33	92.67
	Total score ICCS2016 Items	2.34	0.58		
National Option	A. I can't totally agree with “same rights for every immigrant”	1.46	0.94	48.93	51.07
	B. In some situations, immigrants should have the same rights as non-immigrants, but there are situations in which this shouldn't be the case	1.84	0.83	28.34	71.66
	C. I am not sure which rights an immigrant in Germany should have	1.40	0.88	53.50	46.50

Items		<i>M</i>	<i>SD</i>	<i>D-SD</i>	<i>A-SA</i>
National Option	D. Germany shouldn't take more refugees from disaster areas in the next years	1.24	1.02	61.94	38.06
	E. Refugees shouldn't have the right to get cash from the state	1.03	0.88	74.35	25.65
	F. Refugees shouldn't come to Germany because they can't get a job in their country	1.72	0.99	41.08	58.92
	G. Refugees in Germany shouldn't get better apartments than a welfare beneficiary	1.59	0.95	47.55	52.45
	H. I am clueless as to how to get along with refugees in Germany	1.68	0.96	40.78	59.22
	Total score National Option Items	1.49	0.59		

Note. *M* = Mean; *SD*= Standard deviation; *SD-D* = Strongly Disagree-Disagree; *A-SA* = Agree-Strongly agree.

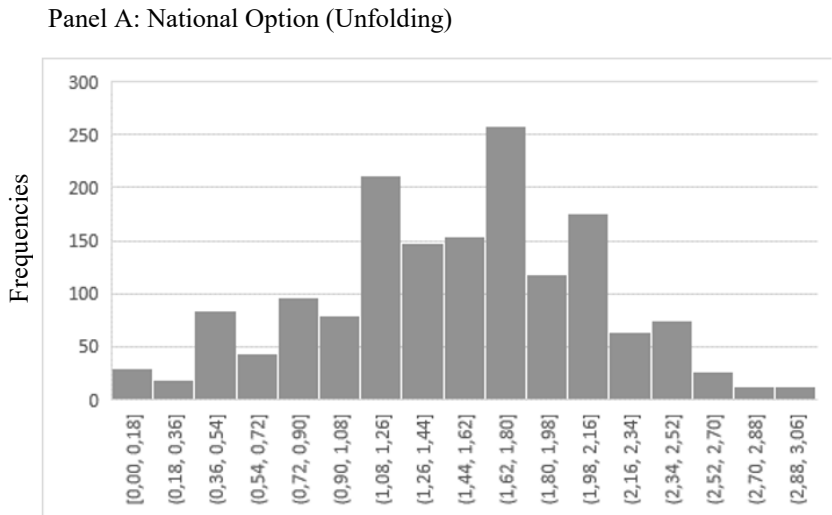
Source: Own representation based on data taken from ICCS 2016

4.2 Preliminary Results

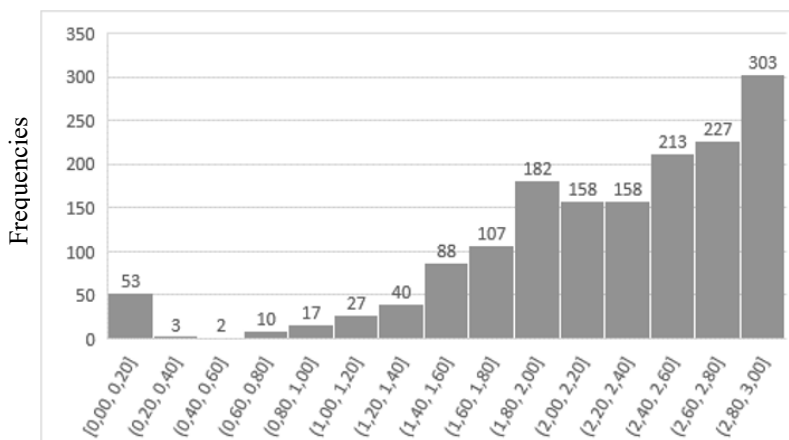
The following analyses are based on the German target population of the ICCS 2016 study. The total sample of students was 1,582 (825 girls and 757 boys) and all of them answered both types of items: ICCCS 2016 items and National Option Items. The descriptive data tell the reader that our items seem to be not that easy to agree with. Especially the intermediate items A, B, C and H were difficult for the respondents to agree with. In fact, the average agreement for these items was around 57%. In particular, most respondents (average 56%) agree with the three statements (Item A, B and C) reflecting a degree of uncertainty regarding the civic principle of “same rights for all”. In particular, the highest percentage of agreement relates to the statement (Item B) that qualifies the principle of “same rights” for all and makes it applicable only to some situations (71%). Similarly, 51% of the respondents agree with the idea of “same rights for every immigrant” (51%). In contrast, on average 87% agree with the ICCS 2016 items regarding the attitudes towards immigrant rights. For reasons already discussed, this discrepancy is expected. Interestingly, 59% of students agree that getting along with refugees is not a

simple matter (Item H). Regarding the more negatively formulated items in our national options, we can see that respondents do not support the statement that refugees should not get money from the host state (Item E, 74%). Students agree with financially supporting refugees, but not in a detrimental way as shown by the 52% agreeing that refugees should not get better apartments than a welfare beneficiary (Item G). Finally, 59% of the students agree that refugees should not come to Germany for economic reasons (Item F). If we calculate an average value across the items for both ICCS 2016 items and the National Option (intermediate) Items, we can see that the frequency distribution of the ICCS 2016 scale is highly skewed, while the newly developed scale shows more of a normal shape (cf. Figure 2), which might be more realistic for the “true” attitude distribution (see Gulliksen 1945).

Figure 2. Frequency distribution of the total scores: Panel A, items developed as national options; Panel B, items used in ICCS 2016



Panel B: ICCS 2016 (Dominance)



Note: Scores range 0 - 3 (0 = strongly disagree; 1 = disagree; 2 = agree; 3 = strongly disagree).

5. Conclusions and future research

After our descriptive analyzes of the data, we can conclude that both types of processes, dominance and unfolding, provide a somewhat different picture when it comes to respondents' attitudes towards migrants and refugees. These differences can be expected when one considers the fundamental assumptions that govern the scale development of both approaches. In particular, on the issue of immigrant rights (a subset of the list of inalienable human rights⁸), we see that students strongly support these macro-normative statements as reflected on the ICCS 2016 items, but at the same time students feel some degree of ambivalence with such principle on the abstract level, and agree with the need for contextualization to the complexities of the social *setting in which they live*. This is an important distinction, because at the macro level students may agree with rights for all (education, work, fair payment, etc.), but at a micro level things may not be so definite and clear-cut. At this level, some ambivalence towards immigrant rights starts to emerge. The national option items do not deal with macro-norms, but with more fine-grained

⁸ See the charter of human rights here: https://www.un.org/en/udhrbook/pdf/udhr_booklet_en_web.pdf [United Nations (2015): Universal Declaration of Human Rights]

aspects of the complex psychology of attitudes. Conceptually, this kind of intermediate items should reflect with higher fidelity the “proximal processes”⁹ (Bronfenbrenner/Morris 2006) associated with the so-called “two-way” cultural integration (Beutin et al. 2007). In short, ICCS 2016 items focus on the “enduring challenges” of democratic nation states, while the items of the national option for Germany focus more on other aspects related to the “emergent” challenges to citizenship education brought about by the “refugee crisis” (Abs/Hahn-Laudenberg 2017). We believe that these two foci should be given the same amount of attention in the future. Although our first attempt at developing intermediate items is not perfect, it shed light into the differences of both approaches. Certainly, other strategies for developing intermediate items should be pursued in the future (e.g., Michell 1994; Cao/Drasgow/Cho 2015). However, regardless of the strategy used, there are important issues to consider in future research on the measurement of attitudes.

First, it is necessary to address the issue of social desirability in attitude research. As already mentioned, some authors accept social desirability as intrinsic to citizenship education (Ten Dam et al. 2013), but we disagree with this view. One possible way of reducing socially desirable responding is straightforward. We believe that exposing respondents to intermediate items reflecting people’s everyday encounters with migrants could lead to the impression that the survey really tries to address the entire complexity of the issue, and that it is expected for people not to have a clearly developed attitude in any defined direction (positive or negative). One can present these items to respondents under different experimental conditions trying to induce positive, neutral, negative reactions to, for example, refugees by showing different stories, and then analyzing the item properties that could provide evidence of socially desirable responding. Another line of research to capture such attitudes with minimum socially desirable responding can be the so-called indirect or objective measurement. In particular, the Conditional Reasoning Technique (LeBreton/Grimaldi/Schoen 2018) has been shown to be robust against “faking” attempts (Wiita/Meyer/Kelly/Collins 2017), and has been used for measuring sensitive constructs such as aggression tendencies. This technique assumes that persons believe reason dictates their decisions to behave and not the other way around. By capitalizing on this idea, the authors develop problems with alternatives that appear to be a problem of logical reasoning. Without knowing it, by choosing one alternative response over the other, respondents reveal tendencies such as aggression among others. Therefore, the new approach seems to be promising for developing a measurement instrument for attitudes towards immigrants,

9 Proximal processes are “particular forms of interaction between organism and environment (...), that operate over time and are posited as the primary mechanisms producing human development” (Bronfenbrenner/Morris 2006: 795).

and towards sensitive issues more generally where social desirability is likely to appear.

Second, concerning the content validity of the attitude construct, it is important to embed into the item pools concrete issues and ideas that can be found in daily life, so that we can enrich the final scales and avoid having only positive and general items which would be endorsed by almost anybody taking the survey (like “everybody should have the same rights”). This should entail not only items that appear in other scales and instruments already developed, but also controversies and discussion topics that can be found in newspapers, on television and on the internet (see discussion forums or discussion threads at the end of online-articles). For example, discourse analysis concerning online discussion forums on the issue of migration as conducted by Fuller (2018) reflected the concerns of samples of people with issues such as tolerance, immigrants’ adherence to the receiving culture and the discrimination/integration coming from both sides. As suggested by Kentmen-Cin and Erisen (2007), it is important to develop items that distinguish important categories such as perceived symbolic (cultural and religious) and security/economic threats from skilled/unskilled, legal/illegal, religious/nonreligious migrants and refugees. Regarding the symbolic threat, it would be necessary to consider what aspect of the refugees’ values is perceived as threatening to the receiving society (e.g., gender roles, respect for authority, child-rearing practices). In short, in developing attitudes scales, we need to move away from macro-normative statements (e.g., “everyone should have the same rights”) which are easy to agree with, to a more fine-grained focus based on people’s everyday experiences with refugees. From this perspective, negative, intermediate and positive items should be developed.

Finally, people need to learn how attitude statements are to be answered. Instructions as to what respondents are expected to do or not to do and the impact of this on the results need to be made explicit to the respondent. We cannot expect respondents to fully grasp their tasks by giving instructions such as “Try to be honest. It is anonymous. Indicate your level of agreement/disagreement with the following statements”. Here it would be helpful to see the detailed instruction respondents receive when answering ability and knowledge tests. Crafting and trying out different sets of instructions before assessing the actual attitude items by means of cognitive interviews should be part of any scale development effort. For example, by asking respondents to carefully read each of the statements first, before attempting to answer them, and then asking respondents to recollect their experiences, thoughts, and feelings about the topic, opens up the opportunity to get an overview of the issue at hand and activate the relevant information from memory.

Regardless of what research on this area will look like in the future, measuring non-cognitive variables takes the form of three “what” questions to be addressed when developing a scale (Filsecker 2019): (1) What is the nature of the construct we need to measure (e.g., knowledge, ability, aptitude, or attitudes); (2) What kind of processes are being elicited by the items and the instructions we develop (maximal performance or personal preferences?); and (3) What are the appropriate psychological models for estimating peoples’ attitudes (i.e., unfolding or dominance models?).

References

- Abs, Hermann Josef/Hahn-Laudenberg, Katrin (eds.) (2017): *Das politische Mindset von 14-Jährigen. Ergebnisse der International Civic and Citizenship Education Study 2016*. Münster: Waxmann Verlag.
- Allport, Gordon W. (1954): *The nature of prejudice*. Cambridge, MA: Addison-Wesley.
- Allport, Floyd Henry/Hartman, D. A. (1925): The measurement and motivation of atypical opinion in a certain group. In: *American Political Science Review*, 19, 4, pp. 735-760.
- Andrich, David (1996): A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. In: *British Journal of Mathematical and statistical Psychology*, 49, pp. 347-365.
- Berry, John W./Kim, Uichol/Power, Simon/Young, Marta/Bujaki, Merridee (1989): Acculturation attitudes in plural societies. In: *Applied Psychology: An international Review*, 38, 2, pp. 185-206.
- Beutin, Ricklef/Canoy, Marcel/Horvath, Anna/Hubert, Agnes/Sochacki, Myriam (2007): Reassessing the link between public perception and migration policy. In: *European Journal of Migration and Law*, 9, pp. 389-418.
- Beverluis, David/Schoeller-Diaz, David/Anderson, Martin/Anderson, Natalie/Slaughter, Amy/Patel, Ronak B. (2016): Developing and validating the refugee integration scale in Nairobi, Kenya. In: *Journal of Refugee Studies*, 30, 1, pp. 106-132.
- Bronfenbrenner, Urie/Morris, Pamela A. (2006): The bioecological model of human development. In: Lerner, Richard M. (ed.): *Handbook of child development: Vol. 1. Theoretical models of human development*. Hoboken, NJ: Wiley, pp. 793-828.
- Cao, Mengyang/Drasgow, Fritz/Cho, Seonghee (2015): Developing Ideal Intermediate Personality Items for the Ideal Point Model. In: *Organizational Research Methods*, 18, 2, pp. 252-275.
- Carter, Nathan T./Dalal, Dev/Boyce, Anthony S./O'Connell, Matthew S./Kung, Mei-Chuan/Delgado, Kristin M. (2014): Uncovering curvilinear relationships between conscientiousness and job performance: how theoretically appropriate measurement makes an empirical difference. In: *Journal of Applied Psychology*, 99, 4, pp. 564-586.

- Coombs, Clyde H./Avrunin, George S. (1977): Single-peaked functions and the theory of preference. In: *Psychological Review*, 84, 2, pp. 216-230.
- Chan, Joseph/Ho-Pong, To/Chan, Elaine (2006): Reconsidering Social Cohesion: Developing a Definition and Analytical Framework for Empirical Research. In: *Social Indicators Research*, 75, pp. 273-302.
- Cronbach, Lee J. (1949): Essentials of psychological testing. New York, NY: Harper.
- Cronbach, Lee J./Meehl, Paul E. (1955): Construct validity in psychological tests. In: *Psychological Bulletin*, 52, 4, pp. 281-302.
- Danner, Daniel/Blasius, Jörg/Breyer, Bianka/Eifler, Stefanie/Menold, Natalja/Paulhus, Delroy L./Rammstedt, Beatrice/Roberts, Richard D./Schmitt, Manfred/Ziegler, Matthias (2016): Current challenges, new developments, and future directions in scale construction. In: *European Journal of Psychological Assessment*, 32, 3, pp. 175-180.
- Davison, Mark L. (1977): On a metric, unidimensional unfolding model for attitudinal and developmental data. In: *Psychometrika*, 42, pp. 523-548.
- De Vroome, Thomas/Martinovic, Borja/Verkuyten, Maykel (2014): The Integration Paradox: Level of Education and Immigrants' Attitudes towards Natives and the Host Society. In: *Cultural Diversity and Ethnic Minority Psychology*, 20, 2, pp. 166-175.
- Filsecker, Michael (ed.) (2019): Latent-Trait Models for the Measurement of Attitudes and personality variables [Special issue]. In: *Applied Psychological Measurement*, 43, 3, pp. 179-180. doi: 10.1177/0146621619835502.
- Fuller, Janet M. (2018): Immigration, integration and Leitkultur in German newspapers: competing discourses about national belonging. In: *Studii de lingvistica*, 8, pp. 175-189.
- Gulliksen, Harold (1945): The relation of item difficulty and inter-item correlation to test variance and reliability. In: *Psychometrika*, 10, 2, pp. 79-91.
- Heath, Anthony/Richards, Lindsay (2019): How do Europeans differ in their attitudes to immigration? Findings from the European Social Survey 2002/03 – 2016/17. OECD Social, Employment and Migration Working Papers, No. 222. Paris: OECD Publishing.
- Jetten, Jolanda/Esses, Victoria M. (2018): The Reception of Immigrants and Refugees in Western Countries: The Challenges of Our Time. In: *Journal of Social Issues*, 74, 4, pp. 662-673.
- Kentmen-Cin, Cigdem/Erisen, Cengiz (2017): Anti-immigration attitudes and the opposition to European integration: A critical assessment. In: *European Union Politics*, 18, 1, pp. 3-25.
- Lai, Calvin K/Marini, Maddalena/Lehr, Steven A./Cerruti, Carlo/Shin, Jiyun Elizabeth/Teachman, Bethany et al. (2016): Reducing implicit racial preferences: II. Intervention effectiveness across time. In: *Journal of Experimental Psychology: General*, 145, 8, pp. 1001-1016.
- Likert, Rensis (1932): A technique for the measurement of attitudes. In: *Archives of Psychology*, 140, pp. 5-53.
- LeBreton, James M./Grimaldi, Elizabeth M./Schoen, Jeremy L. (2019): Conditional Reasoning: A Review and Suggestions for Future Test Development and Validation. In: *Organizational Research Methods*, pp. 1-31.
- Masters, Geoff N. (1982): A Rasch model for partial credit scoring. In: *Psychometrika*, 47, 2, pp. 149-174.

- Maliepaard, Mieke/Verkuyten, Maykel (2018): National disidentification and minority identity: A study among Muslims in Western Europe. In: *Self and Identity*, 17, 1, pp. 75-91.
- Michell, Joel (1994): Measuring Dimensions of Belief by Unidimensional Unfolding. In: *Journal of Mathematical Psychology*, 38, 2, pp. 244-273.
- Miller, Brian/Ruggs, Enrica (2014): Measurement invariance tests of the Impression Management sub-scale of the Balanced Inventory of Desirable Responding. In: *Personality and Individual Differences*, 63, pp. 36-40.
- OECD (2017): PISA 2015 technical report. Paris: OECD Publishing.
- Roberts, James S./Laughlin, James E./Wedell, Douglas H. (1999): Validity Issues in the Likert and Thurstone Approaches to Attitude Measurement. In: *Educational and Psychological Measurement*, 59, 2, pp. 211-233.
- Schachner, Maja K./Van de Vijver, Fons J. R./Noack, Peter (2018): Acculturation and school adjustment of early-adolescent immigrant boys and girls in Germany – Conditions in school, family and ethnic group. In: *Journal of Early Adolescence*, 38, 3, pp. 352-384.
- Schulz, Wolfram/Ainley, John/Fraillon, Julian/Losito, Bruno/Agrusti, Gabriella/Friedman, Tim (2017): *Becoming citizens in a changing world: IEA International Civic and Citizenship Education Study 2016 – International Report*. Amsterdam: IEA.
- Stanton, Jeffrey/Sinar, Evan/Balzer, William/Smith, Patricia (2002): Issues and strategies for reducing the length of self-report scales. In: *Personnel Psychology*, 55, 1, pp. 167-194.
- Stark, Stephen/Chernyshenko, Oleksandr O./Dragow, Fritz/Williams, Bruce A. (2006): Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? In: *Journal of Applied Psychology*, 91, pp. 25-39.
- Ten Dam, Geert T.M./Geijsel, Femke/Guuske Ledoux, Emmanuel/Meijer, Elise J. (2013): Citizenship of students and social desirability: Living apart together? In: *International Journal of Educational Research*, 62, pp. 229-238.
- Thurstone, Louis Leon/Chave, Ernest John (1929): *The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the Church*. Chicago: University of Chicago Press.
- Vedder, Paul/Wenink, Erlijn/van Geel, Mitch (2016): Explaining negative outgroup attitudes between native Dutch and Muslim youth in the Netherlands using the integrated threat theory. In: *International Journal of Intercultural Relations*, 53, pp. 54-64. doi: j.ijintrel.2016.05.001
- Wark, Colin/Galliher, John F. (2007): Emory Bogardus and the origins of the social distance scale. In: *American Sociologist*, 38, pp. 383-395.
- Wiita, Nathan E./Meyer, Rustin D./Kelly, Elnora D./Collins, Brian J. (2017): Not Aggressive or Just Faking It? Examining Faking and Faking Detection on the Conditional Reasoning Test of Aggression. In: *Organizational Research Methods*, pp. 1-28.
- Wirtz, Coen/van der Pligt, Joop/Doosje, Bertjan (2016): Negative attitudes toward Muslims in The Netherlands: The role of symbolic threat, stereotypes, and moral emotions. *Peace and Conflict*. In: *Journal of Peace Psychology*, 22, pp. 75-83.

Refugee Experiences in Higher Education: Female Perspectives from Egypt

Ericka Galegher¹

1. Introduction

Given the displacement of significantly large numbers of university-qualified students, particularly from Syria (Streitwieser/Miller-Idriss/De Wit 2017), there remain significantly limited opportunities to access higher education (HE) in displacement. According to United Nations High Commissioner for Refugees (UNHCR), only 1% of refugees advance to study in a Higher Education Institution (HEI) compared to the global average of 37% (UNHCR 2016). This reality suggests a crisis within the global refugee framework and a failure to provide higher educational opportunities in the face of increasing demand, exacerbating the likelihood of a “lost generation”. However, refugee access to higher education varies based on host country and country of origin (Ferede 2018). For example, 5% of Syrian refugees enrolled in universities in Lebanon, Jordan, Iraq, and Turkey, a rate which was five times higher than the average for refugees worldwide (UNHCR 2018a). In Egypt, nearly 40% of the Syrian refugees are young adults aged between 18 and 39 (Ayoub/Khallaf 2014). Additionally, there is a significant number of Yemeni students studying in universities in Egypt who are in vulnerable situations and unable to return home due to the ongoing war (Interview Yemini Student I).

The situation for female refugees from Syria and Yemen is far more precarious due to the lack of support and access to the formal job market in Egypt. In fact, research from the United States based Institute for International Education found that “displaced university-qualified Syrian males are three times more likely than females to resume their tertiary studies” (Damashke-Deitrick et al. 2019). However, “Syria used to be one of the most highly educated countries in the Arab world, and one of the earliest to achieve roughly equal gender parity in universities” (Locke 2017: 1). In Yemen, only 6% of Yemeni women were enrolled in HE compared to 14% of Yemeni men (UNESCO 2011).

The barriers to HE for refugees are well documented. The main challenges to entrance are identified as lack of documentation and credentials, information, language, discrimination, and finances (Damashke-Deitrick et al. 2019). The experiences of female refugees who enroll in these institutions

1 Ericka Galegher is an Independent Researcher in Egypt. Email: egalegher@gmail.com

as well as evidence-based individual and societal effects are less understood. Thus, the aim of this study was to examine the experiences of female refugees from Syria and Yemen in universities in Cairo, Egypt. Specifically, it sought to query the challenges and opportunities female refugees face in Egyptian universities. These experiences were then contextualized within Egypt's existing political framework for refugees and asylum seekers. The analysis found a significant decoupling between the government's agreement to international refugee frameworks and its capabilities on the national and local level given resource constraints and political instability. However, findings from the interviewed refugee women indicated that HE institutions could offer a pre-existing infrastructure to provide support by facilitating an identity as a student, cultivating long-term academic knowledge and language skills, encouraging the development of social and support networks, and empowering women. The study found these outcomes were applicable to all female interviewees across status indicators.

2. Higher education and refugee women

Research has consistently highlighted the increased vulnerability of girls and women refugees (Freedman 2016). Not only are they more vulnerable to gender-based violence in displacement, but they are more than twice as likely to be out of school and "90% more likely to be out of secondary school than their counterparts in countries not affected by conflict" (UNHCR 2015: 21). Increasing access to HE for refugee women is significantly important given the number of university-qualified females arriving in Egypt as well as the ability for education, HE in particular, to provide skills (Zeus 2011), encourage societal participation (Dryden-Peterson 2010), and create a normalizing effect (Mundy/Dryden-Peterson 2011).

Research has consistently highlighted the importance not only of opportunities to education along the continuum but also the global education movements' persistent neglect of funding at the level of HE (Avery/Said 2017; Barakat/Milton 2015; Dryden-Peterson/Giles 2010). This is due in part to the misconception that funding for HE may reinforce inequality within displaced communities (Dryden-Peterson 2010). These concerns, however, suggest that only a small proportion of refugees are university-qualified when in fact a significant number of current refugees are university-qualified and likely to desire to continue their education. In 2016 alone, an estimated 100,000 to 200,000 Syrian refugees were university qualified yet without access to HE (Institute for International Education 2016).

HE can also be an important path for female empowerment, specifically for those most marginalized (Damaschke-Deitrick et al. 2019). Therefore,

HEIs have an important role to play in providing not only short-term support for refugees but also in cultivating long-term skills and human capital (Stanton 2015; Streitwieser et al. 2017). Focusing on the positive role HEIs can play in refugee crises is a necessary component to the on-going shift in refugee frameworks and discourse from short-term relief to long-term, durable solutions (Dryden-Peterson 2010).

3. Theoretical framework and methods

This study draws on sociological neo-institutionalism to situate the experiences of female refugees from Syria and Yemen in Cairo's universities within the broader institutional framework for refugees in Egypt. Sociological neo-institutionalism emphasizes legitimacy-seeking and normative behavior of societal institutions (Jepperson 2001; Meyer/Boli/Thomas/Ramirez 1997). The goal of utilizing this framework was to examine how Egypt continued to emphasize its normative role on the international level and why national level limitations constrained these goals in practice.

First, the experiences of female refugees from Yemen and Syria were explored. The interviews were centered around the women's experiences and their perceptions of the challenges and opportunities within their HE experiences. These experiences were then contextualized highlighting the disconnection between the aspirational goals of the Egyptian government at the international level and the government's limited capabilities within the national context, exacerbated by internal political instability and economic constraints. Within this framework, the experiences of refugee women enrolled in HE provided important insight into how HEIs offer an infrastructure to support aspirational goals despite internal constraints.

The analysis consisted of primary and secondary source data. Primary data was gathered through individual and group interviews with female refugees in universities in Egypt. Interviews were conducted in English and/or Arabic depending on the choice of the interviewee. All interviews were conducted, translated, and transcribed by the author. This data was gathered between 2017 and 2018. Secondary sources such as UN and government documents as well as published scholarly articles were also analyzed to situate the primary source data and to inform about Egypt's political and educational context. After the interviews were transcribed, a coding system was developed using an interrater reliability coder to ensure reliability of the code system. This was done by utilizing three coders to ensure consistency and consensus in coding the data. This system was developed deductively using existing literature and the research question as well as inductively through the interviews.

All seven women interviewed were enrolled in or had recently graduated from a public or private university in Egypt, and all were refugees or asylum seekers according to the UNHCR definition. Individual interviews were held with four women, two from Syria and two from Yemen. One focus group discussion consisting of three women was also conducted; the three women were from Syria and attended a public Egyptian university (see Table 1). Additional information was gathered through communications with staff from UNHCR, refugee NGOs working in Egypt, and a refugee start-up providing information for refugee students. Participants were recruited through organizations and programs working with refugees in Egypt. Although identifying the socio-economic background of the female refugees was not an initial goal of this study, information gathered through the interviews did provide important indicators of background. Together this background information provided important insight into the status indicators of the women in this study which highlights the empowering effect of HE for females.

Table 1. Background Information

	University Type	Parent's Education		Language of University Instruction	Degree Pursued in Egypt
		Mother	Father		
Salma (Syria)	Private	Secondary	Primary	English	MA
Farida (Syria)	Private	Secondary	Secondary	English	MA
Sherine (Syria)	Public	Primary	Primary	Arabic	BA
Nadia (Syria)	Public	Preparatory	Secondary	Arabic	BA
Farah (Syria)	Public	Primary	Primary	Arabic	BA
Alia (Yemen)	Public	Illiterate	Secondary	Arabic/English	MA
Nadine (Yemen)	Private	Secondary	University	English	MA

Note. Pseudonyms are used to protect the confidentiality of the participants.

Source: Data collected by the author

4. Syrian and Yemeni refugees in Egypt

It was difficult to substantiate the exact number of refugees from Yemen and Syria as many had lived in Egypt prior to the wars or simply did not register with UNHCR. Prior to the war, estimates placed the number of Yemeni in Egypt at 30,000 (Espanol 2018). As a result of Yemen's civil war, which began in 2015, the number increased dramatically with estimates between 300,000 to 700,000 Yemenis (Espanol 2018). However, the number of Yemenis registered with UNHCR remained low at only 7,781 (UNHCR 2018b). The number of Syrians registered with UNHCR also remained somewhat low with 132,029 Syrians out of an estimated 500,000 Syrians living in Egypt (UNHCR 2018b). Reasons for not registering were varied. Many Yemenis (Espanol 2018) and Syrians (Ayoub 2017) reported seeing little benefit in registering with UNHCR and often stated the services they provided were difficult to access and insufficient. These claims were supported by the interviewees in this study. According to interviewees, many Syrians did not want to register with UNHCR as it restricted their ability to travel, viewed their stay as only temporary, or they feared retribution from their home government if they returned with a UNHCR stamp or documentation (see Table 2: Interview UN Specialist; Syrian Student I; Syrian Student II; Syrian Student IV; Refugee Entrepreneur). Others, largely from a higher social class, simply did not identify themselves as being a refugee (Ayoub/Khallaf 2014). Similarly, while recruiting Syrians for this study, those capable of paying the expensive international tuition rates did not wish to participate because they did not identify as refugees despite being unable to return to Syria. Similar findings are reported from universities in Lebanon where experiences and identification as a refugee varied significantly across social class (Watenpaugh et al. 2014).

Table 2. List of Interview Partners

Date	Institution	Interview Title	Pseudo-nym	Transcription
19.12.2017	University	Refugee Entrepreneur	Adam	Transcribed
31.01.2018	Public University	Yemeni Student I	Alia	Transcribed
20.06.2018	Public University	Syrian Student V	Farah	Transcribed
07.11.2017	Private University	Syrian Student II	Farida	Transcribed

Date	Institution	Interview Title	Pseudo- nym	Transcrip- tion
13.02.2018	UNHCR	UN Specialist	May	Transcribed
20.06.2018	Public University	Syrian Student IV	Nadia	Transcribed
14.11.2017	Private University	Yemeni Student II	Nadine	Transcribed
05.11.2017	Private University	Syrian Student I	Salma	Transcribed
20.06.2018	Public University	Syrian Student III	Sherine	Transcribed
31.03.2019	Non-govern- mental Organi- zation	NGO education specialist	Yasser	Personal communi- cation

Source: Data collected by the author

5. Experiences of refugee women in universities

Egypt has a vast higher education system with 2,624,705 registered students (EACEA 2017). Approximately 72% of students enroll in one of twenty-four public universities compared to 4.8% of students who enroll in one of nineteen private universities (CAPMAS 2015). Additional HEIs include public technical colleges, private higher institutes, and public middle institutes (Barsoum 2014). The number of refugees enrolled in HEIs is unknown, but more than 4,300 Syrians were attending public HEIs in 2016 (UNHCR 2017a).

A window of opportunity was granted to many Syrian refugees during the presidency of Mohamed Morsi from 2012 to 2013 when he announced that all Syrians would have the same access to HE as locals largely free of charge. It was during this time that the three Syrian women interviewed were able to enroll in public universities. The remaining four women accessed HE through scholarships, one through a scholarship provided by the Yemeni government and the remaining three through third party- and refugee-scholarships from universities. The following findings highlight the life-changing opportunity that access to HE can provide. Centered around the women's stories the author first highlights the challenges they faced and finally the resulting opportunities.

6. Challenges

The following discussion focuses on challenges the interviewees faced in their HE experiences. The interviewees most often indicated challenges related to residency status, finances, and institutional context. Each topic is now discussed in turn.

Acquiring and retaining legal residency status was one of the most serious challenges women and their families faced. However, Nadine a refugee from Yemen stated that she initially came to Egypt because the Egyptian government was “flexible” with Yemenis, and she could initially enter without a visa. This changed after 2013 due to political changes and security concerns. A few of the Syrian women were unable to leave Egypt while visiting family due to escalated violence back home. Additionally, the process to renew visas and paperwork took significant amounts of time and money. However, access to education was one path to acquiring residency for students, and most often, for their families.

The opportunity to apply for a refugee scholarship along with an increasing likelihood that their stay would not be temporary led many of the women to finally register with UNHCR. Most of the women did not feel stable in Egypt because of the persistent changes in government policies regarding status and access to services as well as lack of access to the formal labor market and fear of exploitation. At times HE was the only means of acquiring temporary legal status, as one woman, Salma, explained:

I signed up in the Cairo University for the Business School because I wanted residency. So I didn't study there but I signed up and I was a student there but I never attended classes. I knew that would give me 3 years residency. If I postpone the first year and then the next year I can just fail and then they will ask me to just cancel.

Finally, Alia often expressed frustration with what she viewed as unequal treatment of Yemeni refugees in comparison to Syrians or other refugee groups. In her point of view, Syrians have many advantages such as not needing the same security papers, paying lower fees, scholarships being offered to Syrians only and other free services. Regarding Yemenis she stated, “but us no, we are between. Not refugees and not normal. We are in the middle.”

Without access to Egypt's formal labor market, finances were a major obstacle for all interviewed women and the women in public universities in particular. The financial barrier for many was removed due to the Morsi-era policy allowing all Syrians, regardless of where they received their secondary school certificate, to enter Egyptian universities largely free of charge. However, the successor administration began restricting this policy in 2016, and only Syrians who graduated with an Egyptian secondary diploma could access HE like Egyptians (Interview Syrian Student I; NGO Education

Specialist). For the four women in public universities, university was difficult. As Sherine explains: “I did not get the chance to live like a normal student because of my work. It turned into a certificate that I want to get and that is it.” Due to financial hardships, these women had to work informally alongside studying with often long commutes to classes. They were also expected to help support their families financially and act as caretakers.

Alia faced significant financial problems related to the failure of the Yemeni government to provide the money promised for her scholarship. At the time of the interview in early 2018, she had not received her scholarship money in more than six months and she was expected to support her son in Egypt and family in Yemen. The university also had many additional, often hidden expenses, such as paying for books, labs, chemicals for experiments and even paperwork.

The three women in private universities did not cite such financial problems. This is partially because they did not have similar financial obligations to their families and also because the university provided sufficient funding for the students or opportunities to earn pocket money through work-study programs. The women were thus able to invest more time and focus on their studies. Nevertheless, most of the women were unsure of what would happen in the future upon completion of their studies, since they lacked access to the formal labor market.

With regard to the institutional context there was a stark difference between the public and private university students’ experiences. The institutional challenges faced by the women in public universities included challenges they recognized even Egyptians faced such as overcrowded classes and professors’ lack of time. However, they also described discrimination in the form of negative comments by Egyptian students and staff, lack of support or services and discrimination in administrative procedures, accessing necessary university materials for labs and class, and the electronic library, for one interviewee in particular. Alia, the Yemeni student in a public university, described feeling unwelcome and being treated unfairly by administrative regulations. Alia explains:

Many things we have the right to do it, but they didn’t give us. On the other side they raise the fees. If I want to have a paper [...] that I am registered there in order to renew the residence visa, they give us this paper by 300 pounds, 300 just a paper, to renew the residence visa from the university that says I am a student. Because we are Yemeni people and so we are foreigners, so foreigners pay 300 pounds. Before it was fifty pounds, but now 300 pounds.

Most of the discrimination women described occurred within the public university settings. The increased occurrence of discrimination may also be attributed to the fact that prior to entering university, many of the interviewees, the Syrian public university students in particular, had little interaction with Egyptians and stayed within their Syrian communities.

Although discrimination by government employees was acknowledged, remarks made by Egyptian students and apathy from faculty and staff regarding the additional hardships experienced by a refugee in Egypt were most often identified as problems they faced at university. The women were very clear that they did not want any sympathy or special treatment and often hid the fact that they were refugees, so it is unclear whether this discrimination or mistreatment was because they were refugees or foreigners. The women hoped that the universities could simply ease the bureaucratic process required for registration and make the rules and regulations more transparent.

Private university students most often described difficulties with the high level of academic English required in class for reading and writing as well as the change in learning environment which stressed critical thinking skills rather than memorization. To overcome these challenges, the women spent all their free time studying, and stated they had very little time to take part in other student activities available at their universities. These interviewees did not describe discrimination or apathy occurring within their private universities or unnecessarily complex university bureaucracy.

7. Opportunities

The following discussion focuses on opportunities the interviewees faced in their HE experiences. The interviewees most often indicated opportunities related to societal context, institutional context, social networks, and how their HE experiences cultivated a new identity outside of being a refugee, and feelings of empowerment. Each topic is discussed in turn.

Cultural, religious, and linguistic similarities made the transition to both Egypt and university life easier for the interviewees. Some women stated that wearing the veil and praying between classes was not a problem in Egypt, as they assumed it may be for refugees in Europe. Additionally, although the women in public universities stated English language skills would be very advantageous, they largely relied on their Arabic skills and did not need English to enter university. Language skills are often cited as a significant barrier to accessing HE for many refugees. Without this barrier, the Syrian women, arguably from more marginalized positions in society, were able to access HE and had particularly transformative experiences as a result.

The differing experiences with Egyptians reflect findings from Ayoub (2017) that experiences are largely dependent upon social class. The women in the private universities who also lived in more socio-economically advantaged neighborhoods in Cairo often stated Egyptians were friendly and helpful. Conversely, those in the public universities who lived in the Faisal neighborhood more often described less friendly, at times hostile, encounters.

One woman stated that Egyptians were “saturated” with their own problems.

All women stressed their gratitude for being granted the opportunity to continue their studies at the university level. The experience was transformative and empowering for all women interviewed. Additionally, their university experiences facilitated integration and further understanding of Egyptians, which reflect similar results regarding Sudanese and South Sudanese refugees in Egypt (Feinstein International Center 2012). For interviewed women in public universities, their experiences exposed them to the Egyptian community and provided an opportunity to integrate and learn the small nuances in Egyptian culture. For example, Nadia described how her Egyptian classmate taught her to stand up for herself. “For instance, in Syria, if a guy flirts with a girl it is inappropriate to reply to him, however in Egypt if this happened, the girl can easily go beat him. So we learned that if anyone flirted with you, you have to reply back.” In contrast, the women in private universities stated that their experiences encouraged intercultural exchanges with the other international students rather than understanding Egyptians specifically.

Finally, the universities provided important long-term skills and cultivated capabilities that the women hoped to use in their future whether in work or to continue their education. The private universities provided the women with access to services such as counseling, student clubs, sports facilities, and writing and language support. Despite limited time to utilize all these services, the women were all aware that they were available for them free of charge. Additionally, they stated that classmates, colleagues, and professors provided a significant amount of support. Such services were not mentioned by the interviewed women as being available in the public universities – here, they described the faculty as lacking the time to provide additional help. However, the women were well aware that lack of such services, overcrowding, and limited university resources were problems that even Egyptians faced.

Although the universities made little effort to connect refugee students, the interviewed Syrian students were quite proactive in developing these connections. In fact, one refugee male student single-handedly contacted all other refugee students the interviewee discovered were on campus and created a network amongst this student population (Interview Refugee Entrepreneur). Lack of information was one barrier to enrollment and scholarships. However, the women largely relied on their own community networks, word-of-mouth, and refugee-initiated online platforms on *Facebook* to access information.

Social media is an important networking tool where platforms like *Start-ups without Borders* connect refugees with an Egyptian counterpart to create a partnership for start-up companies. For the interviewed women, social media was a very important source for accessing information for registration

and finding financial aid and scholarships. The refugees themselves were very persistent in supporting their own community, when services and information were inadequate, and found their own durable solutions. For example, all seven interviewed women sought academic advice and support within the Syrian or Yemeni communities; and the interviewed Syrian women, in particular, through refugee initiated social media platforms which provides refugees with information and support to study in HEIs in Egypt. These results highlight the resilience of refugees and the support they found within their own communities in spite of the lack of external support in Egypt and internationally.

The most frequently emphasized effect of the women's university experiences was the facilitation of a new identity as a student which allowed them the space to shed the stigma associated with being a refugee, as well as feelings of empowerment and freedom. Two Syrian women described their experiences in the following way:

Sherine: "I liked that I managed to achieve something [...]. And the idea to be free here and move normally is good because in Syria there was always control."

Farah: "I liked that even though I am not young, I can think and act. There is freedom."

Not only did university provide an alternative identity, freedom, as well as normalcy after fleeing war, but their experiences and desire to pursue HE changed the perceptions of many of their family members. All of the interviewed women who studied at private universities described the support and encouragement they received from their families to pursue university. In contrast, the Syrian women who studied at public universities faced resistance from their families. They stated that many families were afraid to let their daughters study and believed only men should study and work. However, their families saw the effects of war on their daughters and agreed eventually that "the best thing for us is to study and work" (Interview, Nadia). Nadia continues:

All the mentalities have changed. There they would not agree that a girl proceed with her studies, but when we came to Egypt this idea has changed. [...] We learned here that girls are like boys. Imagine that our brothers are not here, so if we went back to Syria and our brothers are not there, how would we be able to develop it? The war has destroyed communities.

Despite the hardships of leaving their homes and the psychological toll many described from constantly worrying about family and friends still in their war-torn countries, Egypt provided many of the women with educational opportunities they would otherwise lack in Syria and Yemen. For the women at the private universities, they were transformed by the high-quality academic environment, international students, and improvement in English language skills. For one Syrian in particular, Egypt provided the opportunity to finish

her preparatory and secondary schooling before entering university. Farah was married at 14 and did not finish school. In Egypt, she was able to finish her pre-university studies and graduate from university.

Finally, all of the women stated hopes of returning to their countries to help rebuild or continue their studies. Nadine stated, “my plans for the future are to pursue my studies and to help my country to do something remarkable, to achieve things, and to be successful.” Salma plans to work with refugees and share her newly acquired knowledge and work experience if she returns to Syria. Nadia wants to study media. “At the end, it is the media who has affected the picture of Syria, to have an honest media.” Farah stated that she wanted to study sociology to “return to Syria and help my country [...]. The problem is mainly in the society, so the solution is in the hands of the social researcher and that is the reason why I chose this subject.”

8. Conclusion

Egypt is a signatory to a number of international agreements regarding refugees including the 1951 Geneva Convention relating to the Status of Refugees and its 1967 Protocol as well as the 1969 Organisation of African Unity Convention. Despite being one of two non-Western members of the drafting committee, Egypt has relegated the responsibility of registering refugees and asylum seekers to the UNHCR and has reservations on personal status, rationing, access to public education and relief as well as access to the labor market and social security (Al-Sharmani 2014). As a result, Egypt lacks both the national level legislation (Ayoub 2017) as well as the resources to provide durable support or the cultivation of livelihoods for these vulnerable populations (see Grabska 2006). Regarding education, Syrian and Yemeni refugees are granted access to public primary and secondary education free of charge like local Egyptians. However, significant challenges to accessing public schools remain, including overcrowding, physical abuse by teachers and students, low quality, and private tutoring fees (Ayoub 2017).

The government’s failure to provide both a legislative framework and financially support such a framework is further exacerbated by the current economic hardships and politicization of security concerns (Ayoub/Khallaf 2014). Egypt has a significantly high level of unemployment particularly with regards to its youth, continuing to protect its formal job market and access to services for nationals. This is problematic because although the Egyptian government and the refugees themselves often perceive their stay in Egypt as only temporary, most find themselves in protracted situations (Al-Sharmani 2014). As UNHCR states, “the promotion of self-reliance in Egypt’s urban refugee situation is hampered by the lack of a legal asylum framework, high

unemployment and limited opportunities for refugees in the informal sector” (UNHCR 2013: 136). Despite the limitations embedded in the national level framework, long-term advantages for both individuals and society as a result of HE can be transformative.

Insights into these refugee women’s experiences in universities suggest that despite challenges related to status, finances, and institutional context, the transformative power of their university experience was felt by all women and their families. For some, these experiences challenged familial resistance to HE and changed their outlook on women’s capabilities to work and study. HE provided women with an alternative identity as a student and a normalcy that many yearned for after the trauma of war. Findings showed that cultural and linguistic similarities along with universities’ pre-existing infrastructure significantly eased transitions and provided greater access to non-English speaking refugees, often the most marginalized.

Although significant differences existed between experiences in public versus private universities, all women expressed the opportunity to attend university as life-changing and empowering. As a result, HEIs in the Middle East must be acknowledged and utilized as an investment in long-term durable solutions for refugees. Within the larger refugee framework in Egypt, HE can provide an important path forward, cultivate human capital, and reignite hope for refugee women. Long-term durable solutions with the support of the international community are still needed. However, HE in Egypt and the Middle East and North Africa region (MENA) is unique in that traditional barriers to HE for refugees, such as language, are more easily overcome.

In conclusion, Egypt’s support for international agreements and refugee frameworks can be viewed as a normative commitment constrained by Egypt’s inability and unwillingness to fulfill the obligations required by these agreements or create a comprehensive legal framework for refugees within the national context (Buckner/Nofal 2019; Sadek 2016). Additionally, the lack of international funding to support countries which host large numbers of refugees further diminishes the significance of finding durable solutions. Only 4% of the funds needed by UNHCR Egypt to fulfill their obligations to refugees have been met (UNHCR 2019). The fragmented commitment both internationally and within Egypt’s domestic policies (Sadek 2016) intensifies the vulnerability and lack of durable solutions for refugees in Egypt.

Despite these international and national level constraints, pre-existing infrastructures present in institutions like HE can provide vital short-term and long-term opportunities and skills for both individuals and societies. For the interviewed women, not only did HE provide short-term relief from the instability in their lives but findings also support claims that HE is vital to cultivating skills for post-war reconstruction and rebuilding (Avery/Said 2017; Barakat/Milton 2015). Research consistently focuses on the importance

of HE for refugees, and these interviews provide further evidence that policymakers and donors alike need to prioritize HE within the global response to refugee crises. However, 85% or approximately 16.9 million displaced persons are hosted by developing regions in already resource-constrained countries (UNHCR 2017b). The international community must increase support and opportunities to access HE in these host countries. As Nadine states, “any Yemeni woman given the opportunity to study would take it.” This is the crux of the problem that the desire, perseverance and commitment to HE for refugee women are met with sparse opportunities.

References

- Al-Sharmani, Mulki (2014): Refugee migration to Egypt: Settlement or transit? In: Düvell, Frank/Molodikova, Irinia/Collyer, Michael (eds.): *Transit migration in Europe*. Amsterdam, Netherlands: Amsterdam University Press, pp. 55-77.
- Avery, Helen/Said, Salam (2017): Higher education for refugees: The case of Syria. In: *Policy & Practice: A Development Education Review*, 24, pp. 104-125.
- Ayoub, Maysa (2017): Gender, Social Class and Exile: The case of Syrian women in Cairo. In: Freedman, Jane/Kivilcim, Zeynep/Ozgur Baklacioğlu, Nurcan (eds.): *A Gendered Approach to the Syrian Refugee Crisis: Routledge Studies in Development, Mobilities, and Migration*. Taylor & Francis.
- Ayoub, Maysa/Khallaf, Shaden (2014): Syrian refugees in Egypt: Challenges of a politically changing environment. *Cairo Studies on Migration and Refugees*, Report no. 7. Cairo, Egypt: The American University in Cairo.
- Barakat, Sultan/Milton, Sansom (2015): Houses of wisdom matter: The responsibility to protect and rebuild higher education in the Arab world. Brookings Doha Center.
- Barsoum, Ghada (2014): The challenges of private higher education in Egypt. Working Paper no. 833. Cairo, Egypt: Economic Research Forum.
- Buckner, Elizabeth/Nofal, Mozyah (2019): The educational response to Syrian displacement: A professionalizing field in a politicized environment. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka L./Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education*. London, UK: Routledge Publishers.
- Central Agency for Public Mobilization and Statistics (CAPMAS) (2015): 2015 Higher Education Annual Bulletin 2013/2014. Cairo, Egypt: CAPMAS.
- Damaschke-Deitrick, Lisa/Galegher, Ericka/Park, Maureen F. (2019): Female refugees' transitions into higher education: Comparative perspectives from Germany, Egypt and Kyrgyzstan. In: Wiseman, Alexander W./Damaschke-Deitrick, Lisa/Galegher, Ericka L./Park, Maureen F. (eds.): *Comparative Perspectives on Refugee Youth Education*. London, UK: Routledge Publishers.

- Dryden-Peterson, Sarah (2010): The politics of higher education for refugees in a global movement for primary education. In: *Refuge: Canada's Journal on Refugees*, 27, 2, pp. 10-18.
- Dryden-Peterson, Sarah/Giles, Werona (2010): Introduction: Higher education for refugees. In: *Refuge: Canada's Journal on Refugees*, 27, 2, pp. 3-9.
- Education, Audiovisual and Culture Executive Agency (EACEA) (2017): Overview of the higher education system: Egypt. Brussels, Belgium: European Union. https://eacea.ec.europa.eu/sites/eacea-site/files/countryfiches_egypt_2017.pdf
- Espanol, Marc (2018): Yemenis in Egypt face another kind of hell. Cairo, Egypt: Al-Monitor. <https://www.al-monitor.com/pulse/originals/2018/12/yemen-yemenis-egypt-displaced-war-harassment-injustice-safe.html>
- Feinstein International Center (2012): Refugee livelihoods in urban areas: Identifying program opportunities. Case study Egypt. Boston, MA: Feinstein International Center.
- Ferede, Martha K. (2018): Higher education for refugees. Paper commissioned for the 2019 Global Education Monitoring Report Migration, displacement and education: Building bridges, not walls. Paris, France: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000266075/PDF/266075eng.pdf.multi>
- Freedman, Jane (2016): Sexual and gender-based violence against refugee women: A hidden aspect of the refugee "crisis". In: *Reproductive Health Matters*, 24, 47, pp. 18-26.
- Grabska, Katarzyna (2006): Marginalization in Urban Spaces of the Global South: Urban Refugees in Cairo. In: *Journal of Refugee Studies*, 19, 3, pp. 287-307.
- Institute of International Education (2016): Jusoor, IIE launch scholarships for Syrian university women. New York, NY: IIE. <https://www.iie.org/Why-IIE/Announcements/2016/09/2016-09-14-Jusoor-IIE-Partnership-Brings-18-Syrian-Women-to-US-Canadian-Universities>
- Jepperson, Ronald L. (2001): The development and application of sociological neoinstitutionalism. Working Paper 2001/5. Florence, Italy: Robert Schuman Centre, European University Institute. <https://worldpolity.files.wordpress.com/2010/05/jepperson-sociological-institutionalism-eui-5-2001.pdf>
- Locke, Alexandra (2017): The education information gap facing Syrian women outside Syria. The Tahrir Institute for Middle East Policy. <https://www.newsdeeply.com/syria/articles/2017/08/07/the-education-information-gap-facing-syrian-women-outside-syria>
- Meyer, John W./Boli, John/Thomas, George M./Ramirez, Francisco O. (1997): World society and the nation-state. In: *American Journal of Sociology*, 103, 1, pp. 144-181.
- Mundy, Karen/Dryden-Peterson, Sarah (eds.) (2011): *Educating children in conflict zones: Research, policy and practice for systemic change, a tribute to Jackie Kirk*. New York: Teachers College Press.
- Sadek, George (2016): *Refugee Law and Policy: Egypt*. US Library of Congress. https://www.loc.gov/law/help/refugee-law/egypt.php#_ftn10
- Stanton, Andrea (2016): Best practices: Making use of existing university resources to welcome and integrate refugee students. In: *Supporting displaced and refugee students in higher education: Principles and best practices*. New York, NY: Institute of International Education.

- Streitwieser, Bernhard/Miller-Idriss, Cynthia/De Wit, Hans (2017): Higher Education's Response to the European Refugee Crisis. In: Gacel-Avila, Jocelyne/Jones, Elspeth/Jooste, Nico/de Wit, Hans (eds.): *The Globalization of Internationalization: Emerging Voices and Perspectives*. London, UK: Routledge, pp. 29-39.
- UNESCO (2011): Yemen: Country factsheet. <http://uis.unesco.org/country/YE>
- UNESCO (2015): Education for all global monitoring report. Policy Paper 21. <https://unesdoc.unesco.org/ark:/48223/pf0000233557>
- UNHCR (2013): UNHCR global appeal 2013 update: Egypt. <https://www.unhcr.org/50a9f826b.pdf>
- UNHCR (2016b): Aiming higher – The other one percent: UNHCR education report 2016. <https://www.unhcr.org/aiming-higher.html>
- UNHCR (2017a): Egypt: Factsheet. Cairo, Egypt: United Nations High Commission for Refugees. <http://reporting.unhcr.org/sites/default/files/UNHCR%20Egypt%20Factsheet%20-%20September%202017.pdf>
- UNCHR (2017b): Global Trends: Forced Displacement in 2017. <https://www.unhcr.org/5b27be547.pdf>
- UNHCR (2018a): Turn the tide: Refugee education in crisis. <https://www.unhcr.org/turnthetide/>
- UNHCR (2018b): Egypt: Country factsheet. <https://view.joomag.com/mag/0668994001544542553?feature=archive>
- UNHCR (2019): Egypt: Country factsheet. http://reporting.unhcr.org/sites/default/files/UNHCR%20Egypt%20Fact%20Sheet%20-%20February%202019_0.pdf
- Watenpaugh, Keith David/Fricke, Adrienne L./King, James R. (2014): *The war follows them. Syrian university students and scholars in Lebanon*. Institute of International Education and UC Davis. www.iie.org/Research-and-Publications/Publications-and-Reports/IIE-Bookstore/The-War-Follows-Them-Syrian-University-Students-And-Scholars-In-Lebanon#.V567XEOJD8
- Zeus, Barbara (2011): Exploring barriers to higher education in protracted refugee situations: The case of Burmese refugees in Thailand. In: *Journal of Refugee Studies*, 24, 2, pp. 256-276.

III. International Large-Scale Assessments and Education Policy

Section Editors:

Nina Jude, Heidelberg University

Janna Teltemann, University of Hildesheim

International Large-Scale Assessments – (How) Do They Influence Educational Policies and Practices?

Nina Jude¹ and Janna Teltemann²

1. Introduction to the section

This section includes four papers focusing on the interplay of Large-Scale Assessments and Education Policy in Europe and the US. They summarize a discussion that was initiated by several roundtable presentations at the conferences of the American Educational Research Association (AERA) since 2016. The following papers mainly focus on the OECD'S Programme for International Student Assessment (PISA) and the IEA's Trends in Mathematics and Science Study (TIMSS) as probably most well-known Large-Scale Assessments. They describe the latest developments in the area of accountability taking into account the respective views of different stakeholders in education.

Nina Jude and Janna Teltemann analyze the developments in assessment and accountability practices in Germany based on data from the PISA school questionnaires. Focusing on the changes in relevant indicators, they try to relate changes in accountability on state and school level to policy developments over the course of 20 years.

Kerstin Martens and Dennis Niemann describe further policy reactions in Germany, focusing on the debate at the level of the municipality. They take a closer look at the implementation of new standard-based assessment in the classrooms and the respective potential curricular change influenced by these educational reform processes.

Lluís Parcerisa, Clara Fontdevila and Antoni Verger analyze policy transfer mechanisms in different European countries to understand the potential link between PISA and national educational policies. They focus on accountability and assessment policies as the most influential component in domestic policy-making processes.

David C. Miller and Frank T. Fonseca elaborate on the changes in TIMSS results over time. They argue that while mean values and league tables usually get the most attention, countries should carefully analyze the

1 Nina Jude is full Professor of Educational Science at the University of Heidelberg. Email: jude@ibw.uni-heidelberg.

2 Janna Teltemann is Professor of Sociology at the Institute for Social Sciences at the University of Hildesheim. Email: janna.teltemann@uni-hildesheim.de

variation and range of student performance to identify achievement gaps in relation to fostering equity in educational systems.

All four papers open up a broad perspective on the topic of the accountability function of large-scale assessment for educational policy. They summarize current research findings, report latest results based on secondary analysis on different levels of the educational systems and highlight current aspects that should be considered when designing future studies addressing accountability on an international scale.

2. International large-scale assessments – (how) do they influence educational policies and practices?

International Large-Scale Assessments (ILSAs) have played an essential part in national educational monitoring for a long time. A substantial body of literature demonstrates the impact of international school assessments, most importantly the OECD's Programme for International Student Assessment (PISA), on national reform projects in education (Breakspear 2012; Dobbins/Martens 2012; Egelund 2008; Ertl 2006; Grek 2009; Knodel et al. 2013, Takayama 2008). However, the effects on policies are complex and often mediated through cultural, institutional and organizational path dependencies. Evidence also suggests that ILSAs have affected the justification and the design of national assessment and evaluation approaches (see for example Best et al. 2013; Lietz/Tobin 2016).

So far, little research exists as to whether ILSA-related educational reform projects have led to changes in educational outcomes – which could then in turn be monitored by international testing projects. As cross-sectional data from ILSAs does not allow for an analysis of causal relationships between antecedents and outcomes, it is not possible to assign changes in outcomes over time to changes of policies. Recently, several papers have addressed the topic of causal analyses with data from ILSAs, for example due to the assessment design and methodological challenges these studies face (Chmielewski 2017; Kaplan 2016; Kuger et al. 2016; Rutkowski 2016).

However, given the limited validity of causal analyses with data from ILSAs, and the fact that countries often interpret the results of ILSAs in their own interest and in order to justify previously intended reforms (Feniger/Lefstein 2014; Heyneman/Lee 2014; Lingard/Lewis 2016, Ozga 2013; Sellar/Lingard 2013) there is still limited knowledge about the associations between assessments, their aims, educational reform, and educational outcomes. More evidence in this respect could help to balance concerns and doubts about the value of ILSAs for fostering quality education.

ILSAs have raised a lot of criticism, such as well-founded skepticism about data comparability between national contexts, but also about the legitimacy of the power some of these studies exert. The implicit alignment of PISA with the New Public Management Paradigm (see for example Mons 2009) constantly feeds into debates about the incompatibility of economic efficiency and holistic and equal education.

The debate whether and how ILSAs have influenced educational policies and practices is ongoing especially in Germany (Grek 2009; Ringarp/Rothland 2010; Niemann/Martens 2015). It has to be noted that ILSAs have become prominent and were strategically implemented in Germany only over the last 20 years, even though some of the western German states participated in selected ILSA since 1965 (van Ackeren 2002). It was the publication of the Third International Mathematics and Science Study (TIMSS) results for a reunified Germany in 1997 that led to first policy reactions (KMK 1997). These included the decision to further participate regularly in ILSAs, to implement quality assurance measures on the school level and to support competition between the German federal states (*Länder*). When the so called PISA-shock in 2001 showed again alarming results for Germany, education became a publicly debated topic over the subsequent decades. As a result of these debates, different national policies focusing on assessment and evaluation have since been implemented, revisited, and revised.

This chapter seeks to discuss whether impact of these policies in Germany can in return be observed in the PISA data. PISA as one of the most prominent ILSAs assesses context indicators of learning as well as students' competencies in different domains. It delivers information to policy makers every three years and is currently implemented in 80 countries. Germany has participated since the first round of PISA in the year 2000. We will analyze selected PISA indicators addressing national evaluation and assessment practices to estimate the changes visible in these indicators since PISA 2000 and discuss their potential in relation to national policies.

3. Assessment and evaluation practices in Germany – evidence from PISA

International large-scale assessments are designed to collect comparable data on student performance and context information on teaching and learning repeatedly over time, enabling a trend analysis of educational systems and their performance. Moreover, they “attempt to relate those trends to changes in policies, practices, and student populations” (OECD 2009: 150). However,

longitudinal datasets from these studies are hard to come by. To date, no comprehensive overview of trend indicators in studies like PISA or TIMSS exists (Jude/Kuger 2017). Furthermore, constructs and indicators might change over time based on refined theoretical frameworks for these studies (Jude 2016; van de Vijver/Jude/Kuger 2019). Hence, secondary analysis needs to carefully research and scrutinize the indicators in question (Jerrim et al. 2017; Rutkowski/Rutkowski 2016).

In our study focusing on assessment and accountability practices in secondary education (Teltemann/Jude 2019), we analyzed items included in the PISA school questionnaires since 2000, describing change over time and differences between countries in the implementation of these indicators. As not all indicators were available for all cycles, the analyses included different time spans for different indicators.

Based on a cluster analysis, we identified four groups of countries which differ in their assessment and accountability practices and show similar patterns of prevalence of the respective policies and practices within their group. In our analyses, Germany belongs to a cluster of countries which includes several continental welfare states (Austria, Belgium, Switzerland, Finland, Greece, and Italy). All countries in this group can be classified by comparably low average values on assessment practices, yet comparably higher values for school evaluation. Still, these countries have also experienced increases in assessment and accountability practices over time.

In this chapter we will further elaborate on the results for Germany, summarizing policy intentions and interventions that might have resulted in the pattern of assessment and accountability practices that can be observed in PISA data. By looking at the data collected by PISA over time, our analyses revealed the following findings for Germany (see Table 1)

- An increase in assessment practices intended to compare schools with regional and national performance can be found between 2000 and 2015: the number of students attending those schools developed from 12 percent to 44 percent. Among the 20 OECD countries under study this value remains still the third lowest in the international comparison.
- School achievement data is hardly ever *publicly accessible* (fifth lowest rank of 20 countries) – this practice was almost non-existent in 2006, and it has not increased since then: 14 percent of students attend schools which publish their assessment data.
- School accountability practices through *monitoring* of school achievement data by *educational authorities* have decreased since 2006, with Germany showing the second lowest value (38 percent).
- In 2000, Germany showed the lowest values when it comes to using assessment data for the purpose of *teacher accountability*. The data reveals a slight increase from 12 to 18 percent of students in

Germany attending schools using assessment data to judge teacher effectiveness.

- Regarding *External evaluation* practices in schools, Germany ranks in the middle with an increase between 2012 and 2015.
- *Teacher peer review* as part of internal evaluation practices is still not very common, even though values for Germany have almost doubled since 2003 from 25 to 45 percent compared to an international mean of 61 percent in 2015.
- Only for one out of 10 indicators (*Principal or senior staff observations of lessons to monitor teacher practice*) Germany shows values above the average of the 20 OECD countries under study.

Table 1. Assessment and Evaluation Practices in Germany and 20 OECD countries

	Item	Year	Value Germany *	OECD mean **	Rank (low-high, of 20)
Standardization/ Benchmarking	Use of standardized assessments (1-2 times a year)	2000	19	46	16
		2003	37	54	14
		2009	39	52	15
		2015	53	63	15
	Assessments used to compare school to district/national performance	2000	12	37	16
		2003	21	44	13
		2009	33	50	15
		2012	43	62	16
		2015	44	73	18
	Assessments used to compare the school with other schools.	2003	17	38	14
		2009	22	42	18
		2012	28	51	16
		2015	30	63	18
Marketization	Achievement data are posted publicly	2006	14	33	15
		2009	11	32	15
		2012	10	37	16
		2015	14	37	16

	Item	Year	Value Germany *	OECD mean **	Rank (low-high, of 20)
School Accountability	Assessments used to monitor schools' progress from year to year	2000	51	62	13
		2003	44	64	15
		2009	58	72	14
		2012	57	79	17
		2015	64	83	18
	Achievement data are tracked by an authority	2006	55	62	12
		2009	29	62	19
		2012	36	68	19
		2015	38	68	19
Teacher Accountability	Assessments used to make judgements about teacher effectiveness	2000	12	37	17
		2003	12	38	19
		2009	22	40	17
		2012	24	46	17
		2015	18	50	20
	Principal or senior staff observations of lessons to monitor the practice of teachers	2003	70	54	8
		2009	72	61	9
		2012	67	61	10
		2015	88	74	10
Evaluation	Quality Assurance: External Evaluation	2012	60	63	10
		2015	70	74	11
	Teacher evaluation through teacher peer review	2003	25	49	16
		2009	22	55	17
		2012	45	58	14
		2015	45	61	15

Source: OECD Databases 2000, 2003, 2006, 2009, 2012, 2015, own calculations. * Value reads as: xx percent of students in Germany attend schools having implemented a respective practice. ** Average value of 20 OECD countries (Austria, Belgium, Denmark, Finland, Germany, Greece, Hungary, Iceland, Ireland, Italy, Korea, Luxembourg, Mexico, Poland, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States)

Taking these indicators and their changes as a starting point, the question of whether ILSAs have influenced assessment and evaluation practices after PISA 2000 leads to inconclusive results for Germany. For most indicators, Germany shows comparably low values and little change over time. High-stakes evaluation like regional comparisons or the publication of performance results of single schools can rarely be found. This is also true for teacher accountability while external school evaluation is comparably more common.

In the following, we will explore possible policy-guided changes in the German educational system that might be reflected in the reported findings from PISA.

4. International LSA and policy intentions in Germany

In the case of Germany, rapid development in educational policy can be traced back to the publication of ILSA results at the turn of the century, namely TIMSS 1997 and PISA 2000 (see for example Waldow 2009; Niemann 2010; Martens/Niemann 2013; Lawn/Normand 2014; Niemann 2015). Both studies revealed i) a large share of students at low competence levels, ii) a huge gap in test scores between students with and without an immigrant background as well as iii) the highest correlation between students' socio-economic backgrounds and performance compared to all other participating countries, thus marking the German education system as unjust regarding equity as well as rather poor-performing. A national extension study comparing the 16 federal states showed rather large differences in students' performance outcomes across the federal states (Baumert et al. 2002).

These results had not been expected by the German public and have been known since then as the so-called "PISA-shock". Why did the findings cause such a shock? One reason might have been the fact that standard-based assessment, or even internationally comparable outcome measures, were not part of the German educational monitoring approach until the late 1990s (van Ackeren 2002; Lundahl/Waldow 2009). Moreover, one could argue that there had not been any monitoring in place at all, other than data from federal statistics focusing on input criteria like financing and resources. Accordingly, these new findings on the seemingly rather poor output of the educational system struck the policy makers rather hard.

So how has the shock influenced education policy-making in Germany until today? The PISA-shock can be seen as a key event that triggered educational policy discussions on different levels of the education system. These included discussions on comparable educational standards as well as questioning the tracking into different school types across all federal states

(Tillmann et al. 2008). In some cases, the reaction to the PISA-shock led to strengthening already existing reform plans. One example was the political debate on all-day schooling in Germany which was often justified with the first PISA results, even though no evidence could be based on the PISA data (Wolff 2003). On the state level, several joint policies emerged over the years. In 1997, following the publication of TIMSS, German educational policy-makers developed a first strategy for educational monitoring which included participation in national and international large-scale assessments, quality assurance and development as well as competition between the German *Länder* (KMK 1997). In 2002, seven areas of focus were identified, including additional support for students from low-income backgrounds as well as immigrants, and the development of comparable national standards and quality assurance through school evaluation (KMK 2002).

Consequently, a discussion on the development and implementation of national education standards for main curriculum subjects emerged alongside the need for an adequate assessment system (Klieme et al. 2003; Ertl 2006). Since then, a standard-based comparison between the federal states is a key indicator in the German educational monitoring system. It includes accountability based on standardized assessments on two levels: In order to compare the German federal states, a national large-scale assessment is conducted every three (five) years to evaluate the educational standards for grades four (nine) for a sample of students. Tests include the areas of mathematics, science, and languages. Results are used to monitor the implementation of the federal educational standards (Stanat et al. 2017). For accountability on school level, so-called “written comparison tests” are implemented for all grade 3 and grade 9 students every year. They include one compulsory subject (either mathematics or language competencies) and are designed especially to raise teaching quality and school development (Richter et al. 2014; Maag Merki/Oerke 2017).

Until today, Germany showed the lowest score on using mandatory standardized tests in schools of all OECD countries (OECD 2016). Lately, performance in these tests has been shown to lead to restructuring federal teacher education and new approaches in the area of school development (see below). Subject specific programs have thus been launched alongside so called “professional schools of education” which reform and professionalize teacher training (Sälzer/Prenzel 2018).

A new strategy for nation-wide educational monitoring was introduced in 2006, when the *Länder* and the federal government agreed on comprehensive, bi-annual reporting (*Autorengruppe Bildungsberichterstattung* 2018). The report focuses on input indicators, but also on outcome measures assessed in national and international assessments. This can be seen as the first systematic national approach to standard-based assessment and reporting, along with new measures for quality assurance in teaching and instruction (KMK 2015).

This joint overall strategy for educational monitoring was evaluated in 2015. The approach was updated and ever since also includes the aim to examine causes of trends over time and differences between federal states. Based on these results, steering mechanisms are envisioned to be implemented to ensure higher quality and equity in schools (KMK 2015).

As the aforementioned PISA indicators showed, schools have also become a target of accountability procedures. This specifically included approaches of external evaluation and assessments and sparked a discussion on evidence-informed school development, including measures like school inspections and evaluations as well as additional teacher training (Huber/Gördel 2006).

5. Evidence-informed school development and accountability in Germany

School autonomy has been discussed as an indicator showing strong relationships to performance outcomes in many educational systems (OECD 2011; Wößmann 2004). School autonomy is a broad concept, which captures the authority and ability of schools to make autonomous decisions about their operative processes. This includes for example decision-making processes in the allocation of human and physical resources, curriculum implementation and collaboration with other schools (Welsh/McGinn 1999). School autonomy is usually related to the *implementation* of rules and less about their *definition*. For example, schools may have the ability to decide *how* to achieve a goal that is defined externally (Teltemann/Windzio 2018). However, school accountability can be seen as a necessary prerequisite where school autonomy is high (Hanushek et al. 2012). On an international scale, Germany is among those countries with the lowest school autonomy (OECD 2013), although in 2015 a larger share of students attended schools which held at least some responsibility for school governance (OECD 2016).

In recent years, a rising number of publications have drawn on PISA data to assess mechanisms of autonomy and accountability across countries (Teltemann/Windzio 2018). Although school autonomy is not necessarily related to higher student performance, differential effects can be found depending on the overall level of economic development or school type and funding (Hanushek et al. 2012; Benton 2014). For Germany, Füssel (2002) describes that accountability on the level of schools or individual teachers – in the sense that a right to good quality education could legally be enforced – “has hitherto not yet fully developed“ (Füssel 2002: 131). Our empirical findings described above however indicate that there is a trend towards more

school and teacher accountability. Still, schools and teachers work with children – who bring very different presuppositions to school and whose interactions again create further conditions for learning. Holding schools and teachers accountable would mean to take these marginal conditions into account, which requires detailed information about schools and their students.

School inspections are an example for more in-depth evaluations of schools and teachers and consequently represent one aspect of accountability that has been implemented in the German federal states after PISA 2000. Data from school inspections is supposed to be analyzed by regional institutes for quality in education and should then feed back into the schools (Rürup 2014). In his overview of educational monitoring, Maritzen (2008) relates the German approach to a model of evidence-based school and system development. It includes feedback on the quality of both processes and products as characteristic of successful schools. He states

[f]or purposes of external accountability, or (in a more deregulated system) for the accreditation of institutions, it may suffice merely to assess the products of a system. As a 'learning organization', however, a school must know which processes offer points of intervention for maintaining or improving those products (Maritzen 2008: 55).

In their longitudinal study, Bischof et al. (2014) analyzed effects of internal and external evaluation in German schools over time. By re-assessing schools that had participated in the first PISA cycle in 2000 and again in the year 2009, they were able to track developments on the school level. They reported an increase in both internal and external evaluation programs along with a positive impact of internal evaluation on students' cognitive outcomes and well-being in school over time.

It can be concluded that evidence-based educational policy on the school level has become an essential goal of the German approach to educational governance (Dedering 2009; Maritzen 2015). Still, it has to be taken into account that the approaches and also accountability measures vary widely between federal states. Moreover, mechanisms of implementing accountability on the school level, including leadership decisions, can hardly be traced back to the influence of educational administration or even policy decisions (Brauckmann 2012). Research approaches summarizing the impact that school inspection might have on school improvement show that inspections bear the potential for a positive impact on the schools' quality development processes (Dedering/Müller 2011).

6. Conclusion

We can conclude that developments in evaluation and accountability practices in Germany can be tracked by using data from ILSAs. However, assessing policy consequences based on ILSA data can still be seen as a rather difficult task. Klieme (2020) discusses the use of so-called “soft” versus “strong” accountability processes that may be identified by analyzing PISA data over time. He advises to interpret results of ILSA with caution, as the implementation of assessment measures can vary greatly even within countries, and *effects* of specific policies and practices on student outcomes can hardly be derived using existing data.

The aforementioned overall strategy for educational monitoring in Germany which has been in place for over a decade now explicitly states the need for further knowledge on the impact of educational governance processes on all levels of the educational system. Analyzing data from ILSAs in this respect can be seen as a first step. Further in-depth analyses are required when it comes to practices in schools and their effects on students’ outcomes. Further analyses also have to take into account that the implementation of policies needs ample time. With respect to accountability at the school level, we are not yet able to draw causal conclusions with the data at hand. Using ILSA data to identify valid indicators can be seen as a first step that needs to be accompanied by country specific studies on the impact of educational governance over time.

References

- Autorengruppe Bildungsberichterstattung (2018): Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung. Bielefeld: wbv.
- Baumert, Jürgen/Lehmann, Rainer (1997): TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske + Budrich.
- Baumert, Jürgen/Deutsches PISA-Konsortium (2002): PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen: Leske & Budrich.
- Bayer, Sonja/Klieme, Eckhard/Jude, Nina (2016): Assessment and Evaluation in Educational Contexts. In: Kuger, Susanne/Klieme, Eckhard/Jude, Nina/Kaplan, David (eds.): Assessing contexts of learning. An international perspective. New York: Springer International Publishing, pp. 469-488.
- Benton, Tom (2014): A re-evaluation of the link between autonomy, accountability and achievement in PISA 2009. Cambridge: Cambridge Assessment Papers.

- Best, Maura/Knight, Pat/Lietz, Petra/Lockwood, Craig/Nugroho, Dita/Tobin, Mollie (2013): The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Final report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Bischof, Linda Marie/Hochweber, Jan/Hartig, Johannes/Klieme, Eckhard (2013): Schulentwicklung im Verlauf eines Jahrzehnts. Erste Ergebnisse des PISA-Schulpanels. In: Jude, Nina/Klieme, Eckhard (eds.): PISA 2009 – Impulse für die Schul- und Unterrichtsforschung. Weinheim u.a.: Beltz, pp. 172-199.
- Brauckmann, Stefan (2012): Schulleitungshandeln zwischen deconcentration, devolution and delegation (3D) – empirische Annäherungen aus internationaler Perspektive. In: Empirische Pädagogik, 26,1, pp. 78-102.
- Breakspear, Simon (2012): The policy impact of PISA: An exploration of the Normative effects of international Benchmarking in school system performance. OECD Education Working Paper No. 71. Paris: OECD.
- Chmielewski, Anna/Dhuey, Elizabeth (2017): The analysis of international large-scale assessments to address causal questions in education policy. https://www.utsc.utoronto.ca/people/dhuey/wp-content/uploads/sites/30/2017/05/ChmielewskiDhuey_2017.pdf
- Dedering, Kathrin (2009): Evidence-based education policy: Lip service or common practice? Empirical findings from Germany. In: European Educational Research Journal, 8, 4, pp. 484-496.
- Dedering, Kathrin/Müller, Sabine (2011): School improvement through inspections? First empirical insights from Germany. In: Journal of Educational Change, 12, pp. 301-322.
- Dobbins, Michael/Martens, Kerstin (2012): Towards an education approach à la finlandaise? French education policy after PISA. In: Journal of Education Policy, 27, 1, pp. 23-43.
- Ebbinghaus, Bernhard (2012): Comparing welfare state regimes: Are typologies an ideal or realistic strategy? In: Draft Paper presented at European Social Policy Analysis Network, ESPAnet Conference. Edinburgh, UK, pp. 6-8.
- Egelund, Niels (2008): The value of international comparative studies of achievement – a Danish perspective. In: Assessment in Education: Principles, Policy & Practice, 15, 3, pp. 245-251.
- Ertl, Hubert (2006): Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. In: Oxford Review of Education, 32, 5, pp. 619-634.
- Feniger, Yariv/Lefstein, Adam (2014): How not to reason with PISA data: an ironic investigation. In: Journal of Education Policy, 29, 6, pp. 845-855. doi: 10.1080/02680939.2014.892156
- Füssel, Hans-Peter (2002): Accountability in education: the case of Germany. In: Education and the Law, 14, 1-2, pp. 127-132.
- Grek, Sotiria (2009): Governing by numbers: The PISA ‘effect’ in Europe. In: Journal of Education Policy, 24, 1, pp. 23-37.
- Hanushek, Eric A./Link, Susanne/Wößmann, Ludger (2012): Does school autonomy make sense everywhere? Panel Estimates from PISA. IZA DP No. 6185. Bonn: Institute for the Study of Labor.

- Heyneman, Stephen P./Lee, Bommi (2014): The impact of international studies of academic achievement on policy and research. In: Rutkowski, Leslie/von Davier, Matthias/Rutkowski, David (eds.): *Handbook of International Large-Scale Assessment*. Boca Raton: CRC Press, pp. 37-75.
- Huber, Stephan Gerhard/Gördel, Bettina (2006): Quality assurance in the German school system. In: *European Educational Research Journal*, 5, 3-4, pp. 196-209.
- Jerrim, John/Lopez-Agudo, Luis Alejandro/Marcenaro-Gutierrez, Oscar D./Shure, Nikki (2017): What happens when econometrics and psychometrics collide? An example using the PISA data (IZA Discussion Paper No. 10847). Bonn.
- Jude, Nina (2016): The assessment of learning contexts in PISA. In: Kuger, Susanne/Klieme, Eckhard/Jude, Nina/Kaplan, David (eds.): *Assessing contexts of learning. An international perspective*. New York: Springer International Publishing, pp. 39-51.
- Jude, Nina/Kuger, Susanne (2017): Questionnaire development and design for international large-scale assessments (ILSAs): Current practice, challenges, and recommendations. Workshop series on methods and policy uses of international large-scale assessments (ILSA). http://naeducation.org/wp-content/uploads/2018/02/2018-Questionnaire-Design-for-ILSA_v02-1.pdf
- Kaplan, David (2016): Causal inference with large-scale assessments in education from a Bayesian perspective: A review and synthesis. *Large-Scale Assessments in Education*, 4, 1, p. 7.
- KMK (1997): Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland. Konstanzer Beschluss. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1997/1997_10_24-Konstanzer-Beschluss.pdf
- KMK (2002): PISA 2000 – zentrale Handlungsfelder. Zusammenfassende Darstellung der laufenden und geplanten Maßnahmen in den Ländern. Berlin: KMK. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2002/2002_10_07-Pisa-2000-Zentrale-Handlungsfelder.pdf
- KMK (2015): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. https://www.kmk.org/fileadmin/Dateien/pdf/Themen/Schule/Qualitaetssicherung_Schulen/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf
- Klieme, Eckhard/Jude, Nina/Baumert, Jürgen/Prenzel, Manfred (2010): PISA 2000-2009. Bilanz der Veränderungen im Schulsystem [School system changes – a summary]. In: Klieme, Eckhard/Artelt, Cordula/Hartig, Johannes/Jude, Nina/Köller, Olaf/Prenzel, Manfred/Schneider, Wolfgang/Stanat, Petra (eds.): *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann, pp. 277-300.
- Klieme, Eckard/Avenarius, Hermann/Blum, Werner/Döbrich, Peter/Gruber, Hans/Prenzel, Manfred/Reiss, Kristina/Riquarts, Kurt/Rost, Jürgen/Tenorth, Heinz-Elmar/Vollmer, Helmut (2003): *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: BMBF.
- Klieme, Eckhard (2020 in press): Policies and practices of assessment – A showcase for the use (and misuse) of international large-scale assessments in educational effectiveness research. In: Hall, James/Sammons, Pamela/Lindorff, Ariel (eds.): *International Perspectives in Educational Effectiveness Research*. Springer.
- Knodel, Philipp/Martens, Kerstin/Niemann, Dennis (2013): PISA as an ideational roadmap for policy change: Exploring Germany and England in a comparative perspective. In: *Globalisation, Societies and Education*, 11, 3, pp. 421-442.

- Kuger, Susanne/Jude, Nina/Klieme, Eckhard/Kaplan, David (2016): An introduction to the PISA 2015 questionnaire field trial: Study design and analysis procedures. In: Kuger, Susanne/Klieme, Eckhard/Jude, Nina/Kaplan, David (eds.): *Assessing contexts of learning. An international perspective*. New York: Springer International Publishing, pp. 75-113.
- Lawn, Martin/Romuald, Normand (2014): Soft governance through PISA benchmarking: German reforms in secondary education. *Shaping of European education*. London: Routledge, pp. 62-77.
- Lietz, Petra/Tobin, Mollie (2016): The impact of large-scale assessments in education on education policy: evidence from around the world. In: *Research Papers in Education*, 31, 5, pp. 499-501. doi: 10.1080/02671522.2016.1225918.
- Lundahl, Christian/Waldow, Florian (2009): Standardization and 'quick languages: The shape-shifting of standardized measurement of pupil achievement in Sweden and Germany. In: *Comparative Education*, 45, 3, pp. 365-385.
- Maag Merki, Katharina/Oerke, Britta (2017): Long-term effects of the implementation of state-wide exit exams: a multilevel regression analysis of mediation effects of teaching practices on students' motivational orientations. In: *Educational Assessment, Evaluation and Accountability*, 29, 1, pp. 23-54. <https://doi.org/10.1007/s11092-016-9244-y>
- Martens, Kerstin/Niemann, Dennis/Teltemann, Janna (2017): PISA and its consequences: Shaping education policies through international comparisons. In: *European Journal of Education*, 52, 2, pp. 175-183. <https://doi.org/10.1111/ejed.12220>
- Maritzen, Norbert (2008): Educational Monitoring in Germany – A System Innovation to Safeguard Quality Standards. In: Ambühl, Hans/Dobart, Anton/Hofmann, Jan (eds.): *Educational Monitoring, Comparative Studies, and Innovation. From evidence-based governance to practice*. Berlin: Berliner Wissenschaft-Verlag, pp. 52-61.
- Maritzen, Norbert (2015): Schulinspektion. Aspekte der Transformation von Governance-Strukturen. In: Pietsch, Marcus/Scholand, Barbara/Schulte, Klaudia (ed.): *Schulinspektion in Hamburg. Der erste Zyklus 2007-2013: Grundlagen, Befunde und Perspektiven*. Münster u.a.: Waxmann, pp. 13-36.
- Martens, Kerstin/Niemann, Dennis (2013): When do numbers count? The differential impact of the PISA rating and ranking on education policy in Germany and the US. In: *German Politics*, 22, 3, pp. 314-332.
- Mons, Nathalie (2009): Theoretical and Real Effects of Standardised Assessment: Background Paper to the study National Testing of Pupils in Europe: Objectives, Organisation and Use of Results. Eurydice Network. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.174.4813&rep=rep1&type=pdf>
- Niemann, Dennis (2010): Deutschland – Im Zentrum des PISA-Sturms. In Knodel, Philipp/Martens, Kerstin/de Olano, Daniel/Popp, Marie (eds.): *Das PISA-Echo: Internationale Reaktionen auf die Bildungsstudie*. Vol. 16. Frankfurt am Main: Campus Verlag.
- Niemann, Dennis (2015): PISA in Deutschland: Effekte auf Politikgestaltung und Organisation. In: *DDS – Die Deutsche Schule*, 107, 2, pp. 141-157.
- Niemann, Dennis/Martens, Kerstin (2015): Monitoring standards of education worldwide: PISA and its consequences. In: Hayden, Mary/Levy, Jack/Thompson,

- John Jeff (eds.): The SAGE Handbook of Research in International Education. Los Angeles: SAGE, pp. 488-497.
- OECD (2009): PISA 2009 Assessment framework: Key competencies in reading, mathematics and science. Paris: OECD.
- OECD (2011): School autonomy and accountability: Are they related to student performance? PISA IN FOCUS 2011/9. Paris: OECD.
- OECD (2013): PISA 2012 results: What makes schools successful? Resources, policies and practices (Volume IV). Paris: OECD Publishing.
- OECD (2016): PISA 2015 results (Volume II). Policies and practices for successful schools. Paris: OECD Publishing.
- Ozga, Jenny (2013): Accountability as a policy technology: Accounting for education performance in Europe. In: *International Review of Administrative Sciences*, 79, 2, pp. 292-309. doi: 10.1177/0020852313477763
- Richter, Dirk/Böhme, Kathrin/Becker, Michael/Pant, Hans Anand/Stanat, Petra (2014): Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. Zusammenhänge zu Veränderungen im Unterricht und den Kompetenzen von Schülerinnen und Schülern. In: *Zeitschrift für Pädagogik*, 60, 2, pp. 225-244.
- Ringarp, Johanna/Rothland, Martin (2010): Is the grass always greener? The effect of the PISA results on education debates in Sweden and Germany. In: *European Educational Research Journal*, 9, 3, pp. 422-430.
- Rürup, Matthias (2014): Landesinstitute für Qualität als neue Akteure im deutschen Schulwesen? Governance-theoretische Annäherungen an eine Alltagstheorie/Are governmental institutes for quality in education key figures in the German educational system? Governance theoretical approaches to a common theory. In: *DDS – Die Deutsche Schule*. Münster: Waxmann, pp. 201-218.
- Rutkowski, Leslie (2016): Introduction to special issue on quasi-causal methods. In: *Large-scale Assessments in Education*, 4, 1, p. 8.
- Rutkowski, Leslie/Rutkowski, David (2016): A call for a more measured approach to reporting and interpreting PISA results. In: *Educational Researcher*, 45, 4, pp. 252-257.
- Sälzer, Christine/Prenzel, Manfred (2018): Policy implications of PISA in Germany: The case of teacher education. In: Volante, Luis (ed.): *The PISA effect on global educational governance*. New York: Routledge, pp. 109-126.
- Sellar, Sam/Lingard, Bob (2013): PISA and the expanding role of the OECD in global educational governance. In: Meyer, Hans-Dieter/Benavot, Aaron (eds.): *PISA, power, and policy: Oxford studies in comparative education* Oxford: Symposium Books, pp. 185-206.
- Stanat, Petra/Schipolowski, Stefan/Rjosk, Camilla/Weirich, Sebastian/Haag, Nicole (eds.) (2017): *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann.
- Takayama, Keita (2008): The politics of international league tables: PISA in Japan's achievement crisis debate. In: *Comparative Education*, 44, 4, pp. 387-407.
- Teltemann, Janna/Jude, Nina (2019): Assessments and accountability in secondary education: International trends. In: *Research in Comparative and International Education*, 14, 2, pp. 249-271.

- Teltemann, Janna/Windzio, Michael (2018): The impact of marketisation and spatial proximity on reading performance: International results from PISA 2012. In: *Compare: A Journal of Comparative and International Education*, 49, 5, pp. 777-794.
- Tillmann, Klaus-Jürgen/Dedering, Kathrin/Kneuper, Daniel/Kuhlmann, Christian/Nessel, Isa (2008): PISA als bildungspolitisches Ereignis. Fallstudien in vier Bundesländern. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Van Ackeren, Isabell (2002): Von FIMS und FISS bis TIMSS und PISA: Schulleistungen in Deutschland im historischen und internationalen Vergleich. In: *DDS – Die deutsche Schule*, 94, 2, pp. 157-175.
- Van de Vijver, Fons/Jude, Nina/Kuger, Susanne (2019): Challenges in International Large-Scale Educational Surveys. In: Sutter, Larry E./Smith, Emma/Denman, Brian D. (eds.): *The Sage Handbook on Comparative Studies in Education*. Thousand Oaks: Sage, pp. 81-100.
- Waldow, Florian (2009): What PISA did and did not do: Germany after the 'PISA-shock'. In: *European Educational Research Journal*, 8, 3, pp. 476-483.
- Welsh, Thomas/McGinn Noel (1999): Decentralization of education: Why, when, what and how? Paris: UNESCO, International Institute for Educational Planning.
- Wolff, Karin (2003): Folgerungen der aktuellen Vergleichsstudien für das Bundesland Hessen und seine Bildungspolitik. In: Hansel, Toni (ed.): *PISA – und die Folgen? Die Wirkung von Leistungsvergleichsstudien in der Schule*. Herbolzheim: Centaurus, pp. 128-152.
- Wößmann, Ludger (2004): How equal are educational opportunities? Family background and student achievement in Europe and the United States. CESifo Working Paper Series 1284. CESifo Group Munich.
https://ideas.repec.org/p/ces/ceswps/_1162.html

Lost in Translation? Local Governance and Policy Responses to International Large-Scale Assessments

Kerstin Martens¹ and Dennis Niemann²

1. Introduction

Today, the education systems of many countries can be characterized by having entered a post-PISA era. PISA, the Programme for International Student Assessment, of the Organisation for Economic Co-operation and Development (OECD), became a global phenomenon in education policy over the last two decades. Since its first installment in 2000, the OECD's encompassing study on students' skills has continuously spread around the globe and effectively influenced national education activities. PISA comparatively evaluates education systems worldwide triennially by testing the academic skills of 15-year-olds. It is the largest international survey, with test questions available in 82 languages³, and has surpassed earlier international Large-Scale Assessments (ILSAs), such as TIMSS (Trends in International Mathematics and Science Study) by the IEA (International Association for the Evaluation of Educational Achievement), in media and policy responses. PISA shapes education systems today, and it is directly and indirectly responsible for the reforms of many education systems worldwide.

PISA does not provide detailed recommendations on what exact reform measures states should introduce. Rather, the studies point to basic characteristics of successful education systems, which can be copied by others. The OECD calls PISA a "global survey" and claims that "countries are keen to learn from each other's successes". PISA is its "brainchild" and the "whole world can take" its test (<http://www.oecd.org/pisa/aboutpisa/>). Further analyses and reports by the OECD link PISA data to education policies and implicitly suggest possible reform areas for lagging states. For instance, the positive correlation of school autonomy and education outcomes (OECD 2008) and emphasizing early childhood education (OECD 2011) are highlighted.

-
- 1 Kerstin Martens is Professor for International Relations and Global Society at the Institute for Intercultural and International Studies at the University of Bremen. Email: martensk@uni-bremen.de
 - 2 Dennis Niemann is Doctoral Researcher at the Institute for Intercultural and International Studies at the University of Bremen. Email: dniemann@uni-bremen.de
 - 3 <https://www.oecd.org/pisa/test/other-languages/xandar-82-languages.htm> [last accessed May 15, 2019]

At the very core of PISA, the performance of national education systems is ultimately determined by measuring competencies of students. Measuring competencies is different from testing knowledge. Rather than reproducing memorized knowledge, the focus is on learning outcomes and skills application that students should have learned at the end of a certain education stage. Through this novel PISA approach, the OECD urges states indirectly to develop mechanisms to monitor the outcome dimensions of their respective education systems.

Taken together, the widely non-precise policy recommendations of the OECD and the direct impetus for implementing broader frameworks provide some leeway for domestic decision-makers when designing PISA-conforming education reforms. The reform impulse and the recommendations taken from PISA are moderated by national and local peculiarities and idiosyncrasies. Furthermore, decisions have to be translated to concrete measures on school and classroom level in order to make a difference. It is a long way from decisions at the level of education ministries down to teachers in class. At multiple junctures, be it at the level of municipalities or school types, the top-level decisions have to be processed and transposed to direct educational measures. Obviously, this long chain can easily lead to over-complexity or unintended consequences in implementing reforms.

In this contribution, we describe how impulses from the international sphere become visible on the local level in Germany. Being a federal state, the German *Länder* determine how education is organized and how grading is done and presented. Our analysis is also an example of how international soft governance exerted by the OECD through PISA (Niemann/Martens 2018) has led to a paradigm shift in German education policy. Focusing on the German *federal state* of Bremen, we show that education reforms, introduced with PISA in mind, resulted in new measures at the classroom level. Thus, we show how the translation of an international concept of competencies measurement has replaced the measurement of knowledge.

2. The PISA effect and Germany's response

Although the magnitude of PISA's impact may differ, a closer look to the literature reveals that countries can hardly ignore it when considering education reforms. In fact, many countries experienced their "PISA shock" in one way or another, each to different extents and at different times. While some countries responded with reform processes to unexpectedly bad results immediately after they had been released, other countries delayed reactions to the PISA study. The US, for example, did not score comparatively well in the first three PISA studies, but reacted to PISA only in 2010, when the Chinese

had outpaced all other participating countries (Martens/Niemann 2013). In 2012 some observers estimated that approximately 50% of participating countries had already initiated reforms in schools and education systems in response to PISA (Breakspear 2012). By now, we can be sure that the number of PISA-responding countries is much higher. In fact, some countries with good results in PISA reacted with reforms to make their education systems even better; Switzerland and Japan serve as two examples (Bieber 2010; Takayama 2013). Moreover, even countries that do not participate in PISA tests themselves are known to observe the survey in order to learn from what works best (Niemann/Martens 2018).

This overview also shows that PISA has evoked numerous studies in the social sciences. However, due to PISA, the existing literature has primarily focused on broader policy reforms on the state level or on national outcome variations after the implementation of reforms. This contribution will more closely examine the concrete measures undertaken at the municipality level in the name of wider PISA reforms.

Germany exhibited one of the earliest and most intense reactions. When the initial results were released in December 2001, an almost hysterical debate about education was triggered throughout the country. While Germany had long taken pride in its education system with its contributions to Western science and philosophy, the international comparative data empirically revealed that the expected superiority of the German education system appeared to be no more than mere mediocrity (Niemann 2010). There was only one issue that placed Germany in a top position in PISA: educational inequality. In no other country was educational success as much determined by students' socio-economic status as in Germany (Allmendinger/Leibfried 2003). In essence, what happened in response was a comprehensive reform initiative in and of the German secondary education system that had not been experienced since the 1960s (Tillmann/Dedering/Kneuper/Kuhlmann/Nessel 2008).

To give an example: as a response to its low PISA results in 2000, Germany introduced binding national education standards, strengthened early education, and all-day schools became the rule rather than the exception in most German *Länder*. A paradigm shift commenced that entailed the introduction of predefined and measurable education outcomes, emphasizing considerations concerning the efficacy and efficiency of the whole school system (Leschinsky 2005). This transformation is still ongoing. The *Länder* are keen on modifying their education systems in the light of new results taken from international and inner-German assessments.

It is a long way down from the decision-making level to the classroom level, and one has to take into account that it is not easy to introduce nationwide standards and educational projects when a country is federally organized, or at least has a federal education system. In fact, this is the case in

26 countries around the world, and Germany is one of them. As part of its constitution (the Basic Law, or “*Grundgesetz*”), it is laid down that education is organized on the level of the federal states (Article 30 Basic Law), and that the *Länder* have the authority to exercise governmental powers insofar as the Basic Law does not provide or allow for any other arrangement or confer legislative power to the federal government (Article 70 Basic Law).

The German *Länder* are autonomous state entities with their own constitutional provisions and are predominantly responsible for the legislation, administration and funding in the policy field of education. A certain degree of homogeneity between the 16 *Länder*’s education policies is primarily secured by the Standing Conference of the Ministers of Education and Cultural Affairs (*Kultusministerkonferenz*, KMK) which serves as a forum for coordination for the education ministries.

The federal government and the responsible ministry, the Federal Ministry of Education and Research (BMBF), in contrast, have almost no formal influence capacities on secondary education. Against this background, the *Länder* have made abundant use of their exclusive legislative competencies (Helbig/Nikolai 2015) and enacted concrete schooling legislations, which cover detailed rules and regulations for the secondary education sector of each *Land* (Hornberg/Parreira do Amaral 2012). In consequence, regulations regarding the introduction of reforms of monitoring education competences were eventually within the responsibility of the individual *Länder*.

With regards to *international* large-scale assessments, federally organized countries are in a different position than countries with centralized education systems. While ILSAs usually measure the whole country and do not discern federal states, policy responses take place on the subnational level. Thus, as regards responses to PISA from a German perspective, one could easily argue that PISA triggered 16 responses by the *Länder* plus one federal response. Furthermore, the units where educational success or failure is ultimately determined are located on a much lower level: schools. This means that findings and policy implications forms ILSA, such as PISA, ultimately have to be translated to local entities.

However, on account of the responsibility of the German *Länder* to legislate all matters concerning secondary education, implementing tangible evaluation procedures of education competences was far from being a unitary process. While the common framework of education standards was laid down in the joint decision of the *Länder* and the Federal Government, each *Land* was individually responsible for implementing the provisions. Since the overall agreement on the objectives was not a detailed concept with spelled out regulations, the *Länder* had some latitude for their own education systems.

Thus, the *Länder* set goals to be achieved by students in a specific subject at a specific point of time in a specific education program, aimed at

systemized and networked learning, and pointed out the expected performances in terms of ranges of requirements (KMK 2004). In this regard, normative expectations were defined. However, the standardization did not involve the standardization of teaching processes (Böttcher 2007), but rather the definition of aims in education. Education standards should contribute to improve the quality and outcomes of teaching and learning that were primarily understood in the context of competence development (KMK 2010). Since the KMK's education standards only formulated overarching expectations and provided basic orientations of general aims, more concrete guidelines for teaching outcomes were to be defined by the *Länder* (KMK 2004). While the framework education standards of the KMK are not planned to be applied directly at the school level, the elaborated standards developed in the *Länder*, in contrast, are directly applicable.

3. Bremen's translation of assessing education competences

From 2000 until 2009, the *Länder's* education performances were compared on the basis of a national standardized assessment (PISA-E), which was directly informed by the international PISA study but has a considerably larger sample size.⁴ Since then, tests of the education standards are used to evaluate and compare the education performance of the *Länder* and established procedures for reviewing individual schools by external expertise of education monitoring (Döbert/Klieme 2009). Furthermore, by using materials for competence assessment developed by the IQB (*Institut zur Qualitätsentwicklung im Bildungswesen*, Institute for Educational Quality Improvement), schools were urged to conduct internal evaluation of their own performance.

In fact, compared to the other 15 German *Länder*, Bremen, the smallest German *Bundesland* (federal state), scored particularly poor on PISA-E, and in any of the following education surveys. Out of 16 *Länder*, Bremen was ranked by far the last in the first PISA-E, and this trend continued in almost every education survey or testing in which education performances of the *Länder* were compared. Bremen is consistently the last or one of the lowest scoring. One of the latest examples is the nationwide *Bildungsmonitor* 2018, in which Bremen also scored the lowest.⁵ Several structural explanations may be provided for these results: Bremen is one of the poorest *Länder*; as a city

4 In the first PISA-E around 34.000 students in 1.460 schools were tested while in the international PISA study the sample of German students was approximately 5.000 of 219 schools (<https://www.mpib-berlin.mpg.de/Pisa/faq.htm>, [last accessed February, 8, 2019].

5 <https://www.insm-bildungsmonitor.de/> [last accessed May 15, 2019]

state it has the highest rate of families receiving social benefits, and it has a high percentage of children who have a migration background where German is not the main language spoken at home.

Thus, according to the PISA data and the data from education surveys within Germany, Bremen had a great need for reforms, and huge political pressure arose to improve the *Land's* education system. Of course, and like in all other *Länder*, there were various ad-hoc measures taken immediately. However, long-term strategic decisions and institutional changes were also introduced. Most importantly, there was an encompassing school reform in 2009, when the traditional tripartite secondary school system was given up in favor of a bipartite system with the so-called *Oberschule* and *Gymnasium*. In both school types students could attain the *Abitur* (equivalent to an American high school diploma).

The transformation took place within two years, and parties agreed on a ten year so-called “school peace” (*Schulfrieden*), an agreement between all parties represented in the Bremen parliament to withhold institutional changes to these agreed reforms independently of who wins or loses the elections. In September 2018 all major parties in Bremen (with the exception of the *Freie Demokratische Partei*/Free Democratic Party, FDP) agreed to extend this *Schulfrieden* for another 10 years.

As part of this reform process there were also reforms in measuring learning achievements introduced in spring 2012. Starting with primary school, new school learning achievement reports were to be designed, which would later also be used by the *Oberschule*. Instead of giving out grades to students on a scale of 1 to 6 (1 being the best and 6 being the worst), or providing short individual texts about how the child is doing in school and what its strength or weaknesses are, children get a so-called *Kompetenzorientierte Leistungsrückmeldung/KompoLei* (competency-oriented feedback), thus measuring the acquiring of so-called “competencies”. It is an explicit aim of this new grading documentation system to focus on competencies in the process of learning achievements, just as PISA recommends.

One component of *KompoLei* is that teachers tick boxes to document what competencies a child acquires over the four years of elementary schooling. Only German and mathematics are categorized into 4 competencies, each of these competencies contains 2 to 3 “sub-competencies.” These start with a competence level of B for basic, followed by a scale from 1 to 10. A frame is supposed to tell parents what is expected in a particular year. The frame encompasses 4 boxes for a school year and moves by two boxes from year to year. Thus, in grade 1 the frame encompasses the boxes 1 to 4; in grade two it encompasses the boxes 3 to 6 and so on. All ticks are binary, thus the cross only documents that the competence is achieved. It does not indicate how well a child did on it or whether the child marginally passed.

Figure 1. Example of the Bremen certificate with one ticked competency of a second grader

Mathematics

Competence area form and shifting:	B	1	2	3	4	5	6	7	8	9	10
Is able to recognize, specify and illustrate figures											
Is able to recognize, specify and illustrate symmetries											

Competence area numbers and operations:	B	1	2	3	4	5	6	7	8	9	10
Is able to orientate him/herself in numerical ranges						x					
Is able to utilize calculation methods											
Is able to solve written maths problems											

Competence area sizes and measurement:	B	1	2	3	4	5	6	7	8	9	10
Has a conception of sizes											
Is able to deal with sizes in written exercises											

Competence area data and radoms:	B	1	2	3	4	5	6	7	8	9	10
Is able to determine, illustrate and evaluate data											
Is able to estimate probabilities											

Boxes 1 to 10 on the scale describe the development of competencies of the child from first to fourth grade.
Box no. 5 complies with the norm standards of the second grade.

Source: adapted for the example from <https://www.lis.bremen.de/fortbildung/grundschulen/kompolei-68225>

Take the example of a 2nd grader. A cross in the competence area “numbers and operations” in box 5 means that the child acquired this competence and that she/he is on the level of what should students have learned in this competence area in this grade. A cross in 6 would mean she/he knows more than is required for a 2nd grader, whereas a cross in 3 or 4 means she/he knows less but is still within the frame of that year.

To illustrate the complexity of the competence measures, behind each of the boxes there is a list of items assigned. For each of these, the teacher has to indicate for each child whether she/he acquired and showed this competence in the classroom. For each “sub-competence”, there are between 1 and 19 items to be checked. Taken altogether, there are 777 items to be checked for each child during the four years of elementary school. In a class with 25 children this amounts to 19,425 items a teacher has to tick for his or her class over the four years.

For example, for the single tick in box 5 in the competence area “numbers and operations” within the sub-competence “orientation in numerical ranges”, the list entails the following items: the child is able to interpret

determination, plotting and estimation of quantities until 100 within the numerical range; is able to count forwards and backwards until 100 within the numerical range; is able to count forwards and backwards in counts of 2, 5, and 10 until 100 within the numerical range; is able to read, write, illustrate and describe numbers until 100; recognizes the extension of 100 within the place-value system; is able to transfer the comprehension of structured illustrations of numbers until 100; is able to transfer the description of quantity comparisons and estimations until 100; is able to duplicate and to divide in half within the numerical range until 100; is able to describe characteristics of even and odd numbers; is able to estimate quantities and therefore uses quantity illustrations; is able to proceed arithmetic patterns. The teacher may tick the box if 50% of the individual items are reached.⁶

While the proclaimed goals of these new competencies measurements were more transparency for parents, teachers and children, the largest German teachers union GEW (German Education Union/*Gewerkschaft Erziehung und Wissenschaft*) criticized the overhasty introduction of the new scheme. During the school year 2013/14, *KompoLei* was tested in five elementary schools in Bremen. Involved teachers and the teachers union responded, amongst other issues, that the amount of work was very high with the new system as the competencies were not sufficiently described, no gradations for single competencies were possible, that the equal weight of all competencies was problematic, and that the new system was only understandable for parents and children after intense explanations.⁷ Teachers further complained about the massive bureaucracy this new system entailed. In order to handle this system adequately, a teacher would have to check for items for each child every day.⁸

Despite this, the Bremen local governments announced shortly after, and before the testing phase was finished and evaluated, that this new grading pattern would be introduced in all elementary schools in Bremen from the following year onwards.⁹ Thus the political determination for introducing competence measures was formative.

Obviously, the intentions when initially introducing these reports reflect the willingness to shift towards a competence monitoring model as proposed by PISA. A child can see progress, reaching further boxes over the years, instead of receiving the same bad mark in mathematics every year after the other. Also, children with learning disabilities can see progress and be

6 FAQ cf. <https://www.lis.bremen.de/fortbildung/grundschulen/kompolei-68225> [last accessed May 16, 2019]

7 GEW 2015, <https://www.gew-hb.de/aktuelles/detailseite/neuigkeiten/kompetenzraster-in-der-grundschule/> [last accessed May 16, 2019]

8 Weser Kurier 2015, https://www.weser-kurier.de/bremen/bremen-stadt_artikel,-grundschulen-viel-buerokratie-wenig-zeit-fuers-wesentliche-_arid,1673656.html [last accessed May 16, 2019]

9 GEW 2015, <https://www.gew-hb.de/aktuelles/detailseite/neuigkeiten/kompetenzraster-in-der-grundschule/> [last accessed May 16, 2019]

evaluated within the same scheme. However, the example of Bremen also shows how difficult it is to find decent measurements in learning achievements in education.

Thus, it is also an example of how valuable ideas get lost in translation from the international to the local level. While these Bremen style competency reports are supposed to show progress, they are an example of complex de-information. All the report says is that the child is within the *norm* (thus the frame). It does not define what the norm is, nor does it indicate what the child is good at, what she/he likes at school, or where she/he should apply more effort.

The example of Bremen's certificates also shows a resolute reaction and how to overdetermine the goal of evaluating. It further reminds us of the multi-level architecture of education policy. As a field, education is a policy field which is interesting from an international relations point of view, in particular because of international initiatives such as the PISA study and the Bologna process in the field of higher education. These ILSAs are an important reason for these developments. They draw the link between global, national, and, as we see here in this case, local level of policymaking in federal systems.

4. Conclusion

Bremen's case of education standards shows us how transnational impetus translates not only to the national level of policy-making, but also to the local level of policy implementation. Due to ILSAs, education is now seen from an output perspective, rather than from an input perspective. In other words, education measurement now concentrates on assessing "competencies" instead of testing knowledge or learning inputs. An often heard argument in this context is that national economies seek to make the best use of human capital in order stay efficient and productive in a global economy. The reform history after PISA mirrors the fact that education is today seen as the key resource of the 21st century. This connection between economy and education also explains the growing significance of ILSAs. They allow education systems to be quantified and to be compared across various levels, be it countries, regions or individual schools. These comparisons also allow weaknesses and strengths to be examined. ILSAs remind us that these paradigm shifts, or perhaps the more tangible PISA, pushed towards emphasizing the human capital approach in education policy, much more than any other education goal. Education reform processes, particularly standardization in education, are a response to problems of "effectiveness", "productivity" and "competitiveness" in a global marketplace. ILSAs also allow identifying

inequalities within education systems more precisely. PISA particularly highlighted the impact of the socio-economic background of students on their academic success. Children from more affluent families have much better chances of success in school than their age peers with less favorable backgrounds.

The example of Bremen reminds us of the idiosyncratic and slow-moving nature of education policy, despite regular ILSAs. It takes time before a reform process reaches the classroom, making significant, measurable differences to learning processes. Therefore, it is legitimate to ask if it is really useful that PISA is conducted every three years. Moreover, although it is a good thing to extract examples of “what works” or “best practices” out of ILSAs, very rarely one can implement them one-to-one in a new context. How existing institutional frameworks in education systems affect the transfer of ILSAs policies is, however, an open question for further research. This contribution aimed to demonstrate that unintended consequences can occur when internationally conceptualized education policies are transferred to the concrete school level.

References

- Allmendinger, Jutta/Leibfried, Stephan (2003): Education and the Welfare State: The Four Worlds of Competence Production. In: *Journal of European Social Policy*, 13,1, pp. 63-81.
- Bieber, Tonia (2010): Playing the Multilevel Game in Education – the PISA Study and the Bologna Process Triggering Swiss Harmonization. In: Martens, Kerstin/Nagel, Alexander-Kenneth/Windzio, Michael/Weymann, Ansgar (eds.): *Transformation of Education Policy*. Basingstoke: Palgrave, pp. 105-131.
- Böttcher, Wolfgang (2007): Zur Funktion staatlicher 'Inputs' in der dezentralisierten und outputorientierten Steuerung. In: Altrichter, Herbert/Brüsemeister, Thomas/Wissinger, Jochen (eds.): *Educational Governance. Handlungskoordination und Steuerung im Bildungssystem*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 185- 206.
- Breakspear, Simon (2012): The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. In: *OECD Education Working Papers*, 71, pp. 1-31. <https://doi.org/10.1787/5k9fdqffr28-en>
- Döbert, Hans/Klieme, Eckhard (2009): Indikatoren gestützte Bildungsberichterstattung. In: Tippelt, Rudolf/Schmidt-Hertha, Bernhard (eds.): *Handbuch Bildungsforschung*. Wiesbaden VS Verlag für Sozialwissenschaften, pp. 317-336.
- Helbig, Marcel/Nikolai, Rita (2015): *Die Unvergleichbaren: Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Verlag Julius Klinkhardt.

- Hornberg, Sabine/Parreira do Amaral, Marcelo (eds.) (2012): *Deregulierung im Bildungswesen*. Münster: Waxmann.
- KMK (Kultusministerkonferenz) (2004): *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung*. Neuwied: Luchterhand.
- KMK (Kultusministerkonferenz) (2010): *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Köln: Carl Link.
- Leschinsky, Achim (2005): Vom Bildungsrat (nach) zu PISA. Eine zeitgeschichtliche Studie zur deutschen Bildungspolitik. In: *Zeitschrift für Pädagogik*, 51, 6, pp. 818-839.
- Martens, Kerstin/Niemann, Dennis (2013): When Do Numbers Count? The Differential Impact of Ratings and Rankings on National Education Policy in Germany and the US. In: *German Politics*, 22, 3, pp. 314-332.
- Niemann, Dennis (2010): Turn of the Tide – New Horizons in German Education Policymaking through IO Influence. In: Martens, Kerstin/Nagel, Alexander-Kenneth/Windzio, Michael/Weymann, Ansgar (eds.): *Transformation of Education Policy*. Basingstoke: Palgrave, pp. 77-104.
- Niemann, Dennis/Martens, Kerstin (2018): Soft governance by hard fact? The OECD as a knowledge broker in education policy. In: *Global Social Policy*, 18, 3, pp. 267-283. doi: 10.1177/1468018118794076
- OECD (2008): *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*. Paris: OECD Publishing.
- OECD (2011): *Starting Strong III: A Quality Toolbox for Early Childhood Education and Care*. Paris: OECD Publishing.
- Takayama, Keita (2013): Untangling the Global-Distant-Local Knot: the Politics of National Academic Achievement Testing in Japan. In: *Journal of Education Policy*, 28, 5, pp. 657-675.
- Tillmann, Klaus-Jürgen/Dedering, Kathrin/Kneuper, Daniel/Kuhlmann, Christian/Nessel, Isa (2008): *PISA als bildungspolitisches Ereignis. Fallstudien in vier Bundesländern*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Understanding the PISA Influence on National Education Policies: A Focus on Policy Transfer Mechanisms¹

Lluís Parcerisa², Clara Fontdevila³ and Antoni Verger⁴

1. Introduction

Over the last decades, the Organisation for Economic Co-operation and Development (OECD) has acquired an increasingly relevant and authoritative role in the global governance of education. The influence of the OECD in education owes much to the greater focus of this international organization on the production of new sources of quantitative data, and to the comparative perspective through which these data is approached (Grek 2009; Martens/Jakobi 2010). This shift has been driven by different data-gathering initiatives, among which the Programme for International Student Assessment (PISA) stands out. Since its first edition in the year 2000, PISA has been administered every three years in an increasing number of countries. Nearly 80 countries have participated in the 2018 edition. According to different observers, PISA has represented a turning point for the OECD and has consolidated its leading role within the global education field (Niemann/Martens 2018). The success of PISA relies, on the one hand, on its capacity to commensurate complex educational processes, such as teaching and learning, in concrete numerical indicators and, on the other, on the country comparisons that result from this quantification exercise (Martens 2007; Grek 2009).

The impact of PISA on domestic policy-making processes has become a well-established and recurring theme within global education studies. While Breakspear noted in 2012 that research into the effects of PISA over national education reform was still limited, considerable progress has been achieved since then. There is mounting evidence of the influence of PISA at different stages of the policy cycle (see for instance Carvalho/Costa 2014; or Steiner-

1 This work has been supported by the European Research Council under the European Union's "Horizon 2020 Framework Programme for Research and Innovation" [grant number 680172 – REFORMED].

2 Lluís Parcerisa is Research Fellow in the Department of Sociology at the Universitat Autònoma de Barcelona. Email: Lluís.Parcerisa@uab.cat

3 Clara Fontdevilla is PhD Researcher in the Department of Sociology at the Universitat Autònoma de Barcelona. Email: clara.fontdevila@gmail.com

4 Antoni Verger is Associate Professor at the Faculty of Political Science and Sociology at the Universitat Autònoma de Barcelona. Email: antoni.verger@uab.cat

Khamsi/Waldow 2018). However, evidence on the influence of PISA remains fragmentary and privileges particularistic accounts and specific country-cases. Also, there is limited evidence on how or whether the influence of PISA on national policy-making results into some form of policy convergence – that is, to what extent country reactions to PISA share a common policy orientation.

This chapter aims at gaining a better understanding of the role of the OECD in the global dissemination of education policies through the PISA program. More specifically, it aims at identifying those mechanisms through which the PISA program shapes or influences processes of domestic education reform. To this purpose, we focus on PISA's role in transferring accountability and assessment policies in education. Accountability and assessment policies represent a potentially productive entry point to understand PISA influence for two different (albeit interconnected) reasons. First, as we have discussed elsewhere (Verger et al. 2019a; see also Gorur 2016; Meyer 2014), the accountability and assessment themes gained centrality within the OECD educational agenda in the mid-2000s; since then, they feature among the most recurrent policy recommendations found on OECD's policy guidance initiatives and research products. Second, according to a survey distributed in 2011 among national representatives in the PISA Governing Board, assessment and accountability constitute the area of PISA policy analysis that countries have judged as the most influential in domestic policy-making processes (Breakspear 2012)⁵.

2. Research framework

The international spread of policy models and policy instruments across countries is frequently explained through policy diffusion and policy transfer theories – that is, theories that emphasize transnational interdependence as a key driver of the dissemination and propagation of certain policies (Dobbin/Simmons/Garrett 2007; Gilardi 2012).

Most studies falling within this area of research tend to focus on bilateral relationships and to suffer from a form of state-centrism that neglects the role

5 A survey previously conducted by Hopkins et al. (2008) suggested similar trends – according to the key stakeholders surveyed, the development of national standards and the establishment of national institutes of evaluation were among the reforms most likely to be adopted in light of PISA results; also, the establishment or further development of accountability systems and increased autonomy for schools were listed as frequently reported changes in school practices and policies.

of international policy intermediaries (Stone 2012). However, more recently, there has been a growing reflection on the role played by non-state and transnational actors in policy diffusion and transfer processes.

Conventionally, three main mechanisms behind policy diffusion dynamics can be differentiated, namely competition, policy learning and emulation⁶. In the following lines, we describe briefly each of these mechanisms while highlighting the potential role of international organizations in activating them.

- a) *Competition* occurs when countries' decisions are motivated by the behavior of their competitors and a sense of a zero-sum game. Competition mechanisms are usually identified in the diffusion of economic policy – as the ultimate goal of such efforts is to secure a certain share of a limited resource, including global capital, access to global trade or export markets, etc. (Dobbin/Simmons/Garrett 2007). International organizations play a key role in the promotion of competition by providing the infrastructure for such dynamics to occur, such as the construction of investment indicators or the publication of country rankings (Doshi/Kelley/Simmons 2004).
- b) *Learning* (also known as lesson-drawing) refers to those cases in which a certain policy is adopted on the basis of its consequences and (perceived) success elsewhere (Magetti/Gilardi 2016; Shipan/Volden 2008). As noted by Marsh and Sharman (2009), learning can occur on a bilateral basis but can also be mediated or encouraged by international organizations, international policy networks or epistemic communities engaged in transnational problem solving.
- c) *Emulation* captures those instances in which a policy option is adopted for symbolic or normative reasons – including a desire for conformity or a quest for legitimacy. Meseguer (2004) notes that the legitimacy and reputational concerns behind emulation dynamics may have a domestic dimension (i.e. a government's need to legitimize its agenda in front of its citizens), but also a global one (countries' need to conform to global norms). Again, transnational actors can play a key role in the promotion of policy models, not only by constructing these models, but also by generating the legitimacy pressures that encourage countries to adopt them (cf. Holzinger/Knill 2005).

6 Some categorizations, including the seminal classification advanced by Dolowitz and Marsh (1996, 2000) consider a fourth mechanism – namely, coercion or coercive transfer. However, other authors exclude this mechanism from the diffusion mechanism category as, unlike learning, emulation and competition, coercion has a vertical or top-down nature and implies the existence of a central force coordinating policy spread (cf. Maggetti/Gilardi, 2016; Shipan/Volden 2008) – thus constituting a distinct category, difficult to reconcile with those approaches to policy diffusion emphasizing the notion of decentralized coordination (Busch/Jørgens 2007).

It should be noted, however, that the distinction between these three mechanisms is essentially analytical. In fact, in empirical situations, differentiating between emulation and learning dynamics represents a particularly challenging endeavor. As noted by different authors, such distinction ultimately depends upon the interpretation of the logics and reasoning guiding policy-makers, and is consequently mediated by one's theoretical lens (cf. Marsh/Shaman 2009). Some authors have proposed different approaches to differentiate learning from emulation. Shipan and Volden (2008), for instance, suggest that learning dynamics put the emphasis on successful policies, whereas emulation dynamics put the emphasis on successful countries. Gilardi (2012), in turn, observes that learning relies on the logic of consequences (that is, the evaluation of the outcomes of a given course of action or its alternatives), whereas emulation relies on the logic of appropriateness (which considers what social norms deemed more adequate or pertinent in relation to a given role, identity or situation).

Overall, policy diffusion literature represents a promising theoretical approach to understand the role of the OECD/PISA in the spread of assessment and accountability reforms across a wide spectrum of countries. Specifically, this chapter examines the role of PISA in facilitating or stimulating educational change through each of the above-mentioned mechanisms of policy diffusion. In terms of methodology, the chapter builds on the results of a document analysis of OECD publications with a focus on accountability policies, and the results of a systematic literature review on processes of policy adoption and policy instrumentation of accountability reforms, which is based on a total of 158 papers obtained through the SCOPUS database (cf. Verger et al. 2019b for an overview of the procedure). To elaborate this chapter, we rely on a subset of 33 papers with an explicit focus on the role of the OECD in the promotion and diffusion of accountability reforms.

3. Mechanisms of PISA policy influence

3.1 Competitive dynamics generated by PISA: Scandalizing countries by comparison

The policy influence exerted by PISA stems largely from the presentation of its results under the form of country rankings and league tables. As noted by Gilbert (2015), rankings bring reputation to the fore and contribute to the emergence of a hierarchical reputational economy. In this context, competition dynamics are likely to emerge as countries strive to escalate rankings or

to preserve a leading position in them. By altering the informational environment, rankings can increase social pressure among policy-makers and bureaucrats due to reputational concerns (Doshi/Kelley/Simmons 2004). We assume thus that the impact of PISA is largely explained by the competition dynamics it triggers.

The statistical data produced through PISA has indeed been found to trigger competition at different levels as a direct result of the “naming and shaming” dynamics and the audit culture that this international assessment, through its comparative approach, generates. As noted by Sellar, Thompson and Rutkowski (2017), PISA promotes the engagement of participant countries in a sort of “global education race” aimed at constantly improving students’ performance in a highly competitive and interdependent economic environment. This education race intensifies for political but also economic reasons since, in a globalizing economic environment, students’ knowledge and skills become a governmental asset to attract foreign investors and to aspire to generate more knowledge-intensive jobs. The US engagement with PISA results is quite illustrative of the competitive pressures brought about by PISA benchmarking. During the 2000s, US authorities did not pay much attention to the release of PISA reports, since the country results mainly confirmed the quality education concerns that had been present in the national debate for decades (Hursh 2007). Nevertheless, the US started to react to PISA results after the 2009 edition. In PISA 2009, China’s performance surpassed the US, and this overtaking was framed and interpreted in the US as a symbol of China’s economic superiority (Niemann et al. 2017).

Overall, competition dynamics have proved to be an effective form of framing and conditioning policy decisions in the context of the OECD (Marcussen 2004). Breakspear (2012) shows that the PISA Governing Board representatives consider the publication of league tables as one of the most persuasive aspects of PISA to advance policy change. The perception, anticipation or fear of damaged reputation or self-image appears thus to be a powerful catalyzer of policy reform.

The connection between reputational damage and policy change is frequently mediated by a change or disruption of domestic policies, and by changes in the terms of the public debate – for instance, through the creation of a narrative about a crisis that requires urgent action. In Norway, for example, the scandalization effect caused by both PISA 2000 and PISA 2003 results facilitated the crystallization of a political consensus around the need of further accountability and quality assurance in education (Hatch 2013; Camphuijsen/Skedsmo/Møller 2018). During the decade that followed, the country engaged in different reforms on accountability, testing and curriculum, portrayed as highly inspired by “the policy advice that emerged from the PISA studies” (Sjøberg 2016: 109). Comparable dynamics can be observed in Spain, where the PISA shock played a key role in the eventual

acceptation of the accountability and external evaluation agenda within the social-democratic party (the Spanish Socialist Workers' Party, PSOE) during the mid-2000s, and opened a phase of (relative) bipartisan convergence that enabled the adoption of performance evaluation arrangements and accountability-oriented policies (Dobbins/Christ 2019; Popp 2010). Similarly, in Denmark, disappointing PISA results played a key role in fostering a public debate that ultimately led to a major education reform in 2006 in which accountability through assessment featured prominently. Remarkably, the impact of PISA-triggered reputational concerns on Danish policy-making dynamics persisted over time – to the point that, in 2010, the Danish Prime Minister stated that the aim of the education system was to secure a position among the top five nations listed in the PISA report (Moos 2010).

More in general, there is evidence that the existence of a gap between national expectations and the results obtained in PISA has frequently favored the opening of a window of political opportunities for the introduction of certain educational reforms (Breakspear 2012; Martens/Niemann 2013). “PISA effects” or “PISA shocks” have been documented in countries such as Germany, Switzerland, England and Australia. In these countries, PISA results have fostered public debates leading to the adoption of assessment and external evaluation arrangements at some level (cf. Baxter/Clarke 2013; Gorur 2013; Niemann/Martens/Teltemann 2017; Sellar/Lingard 2013).

Overall, available evidence shows that PISA plays a crucial role in creating an appetite for reform among decision-makers and impacts agenda-setting dynamics at a domestic level. It is less obvious, however, how (or whether) these “PISA shocks” condition and shape the specific policy response – that is, the content of the policy reforms motivated by (or justified on the grounds of) PISA. As the examples above suggest, there is evidence that PISA induced crises have frequently led to the adoption of accountability and external assessment policies. There is however no obvious explanation for this. To a certain extent, it is possible to assume that the very participation in PISA may increase the legitimacy and social acceptance of rankings and external evaluation – both among policy circles and the public. It is also likely that PISA crises will increase the appeal of output-oriented governance models as a means to improve performance at the system level. However, the interpretation and translation of PISA results into some form of policy guidance has also become instrumental in processes of educational policy change. This is something that we explore in the section that follows.

3.2 *Learning and emulation: What PISA tells us about “what works” in education*

PISA data is customarily used by the OECD as a key source of evidence to support and disseminate policy recommendations, or to promote certain policy models. While this has been the case since the publication of the first PISA results, such dynamics intensified in the mid-2000s, when the OECD stopped outsourcing the elaboration of the PISA reports to external contractors. Specifically, since the 2006 PISA cycle, the final PISA products are produced in-house, what provides the organization with greater capacity to frame and control the message and policy lessons resulting from the data (Bloem 2015).

PISA data remains thus the most relevant source for policy development and policy dissemination activities of the OECD – it lies at the center of the normative work of the organization. The results of the assessment are translated into policy lessons and recommendations (Bloem 2015; Engel 2015) and advance through a wide range of knowledge products – including *PISA in Focus*, *Education Indicators in Focus* or the *Strong Performers and Successful reformers* video series. However, the translation of PISA results into education best practices does not rest exclusively with the OECD. As advanced by Waldo (2017), national and regional governments usually produce their own PISA reports, and local stakeholders and the media do frequently engage in the construction, depiction and promotion of PISA top-scorers as “reference societies”. These countries often serve as models worth imitating – or learning from.

Thus, by providing empirical foundations to the depiction of certain policy options as successful or superior, PISA is likely to trigger both learning and emulation dynamics. Hence, countries are likely to engage in education policy reform on the basis of certain perceptions of “what works” that build largely on PISA data, conveniently translated by the OECD.

The impact of the PISA-based analytic and normative work conducted by the OECD, as well as the resulting learning and emulation dynamics, are particularly evident in relation to the accountability and assessment debate. First, the OECD appears to have played a crucial role in articulating and disseminating accountability and assessment in education as a policy approach that is both effective and desirable. As we have discussed elsewhere (Verger et al. 2019a), accountability and assessment (along with other policies, including school autonomy) have occupied a prominent position within the organization’s agenda for nearly two decades, and a variety of publications (produced by the different units of the Directorate for Education and Skills) have promoted such policies as the solution to a wide variety of problems.

More specifically, publications such as *PISA in Focus* No. 9 or the working paper *School accountability, autonomy, choice, and the level of*

student achievement: International evidence from PISA 2003 (OECD 2011; Wößmann et al. 2007 respectively) which drew largely on PISA data, played a key role in positing the combination of accountability and autonomy as conducive to the improvement of student learning. The latter argued that pedagogic school autonomy (i.e. autonomy and responsibility over curricula, evaluation style and didactics) was positively associated with higher PISA scores, and that managerial autonomy (concerning staffing and resource-allocation decisions) worked in those systems with high levels of accountability – measured as the publication of schools’ results in national assessments. Although more recent initiatives have shifted away from the initial emphasis on market dynamics or high-stakes accountability, certain principles (including the culture of evaluation and assessment, transparency and a focus on outcomes) have consolidated as highly desirable and as a key component of modern education systems.

Second, recent episodes of education policy reforms are indicative of learning and emulation dynamics somehow influenced by PISA results – or by PISA-based advice. As noted above, distinguishing learning from emulation poses an interpretative challenge – as the ultimate motivations and reasoning guiding policy-makers cannot be directly observed. The reviewed cases suggest in fact that, generally speaking, PISA-data sparked a combination of them.

In the case of Spain, for instance, literature suggests that some education reforms at the regional level were partially informed by PISA findings. There is evidence that policy-makers’ perceptions on “what works” in Spain was partially informed by PISA-based policy guidance. This is for instance the case of Catalonia, where the perception of school autonomy and external assessment as desirable policy solutions, consolidated among certain policy circles since the mid 2000s, owes much to the dissemination of these ideas by the OECD through PISA and other products associating this policy option with better-performing education systems (Verger/Curran 2014). These processes can be interpreted as indicative of learning dynamics. They suggest a genuine belief in the potential of certain components of the accountability agenda – empirically substantiated by PISA. At the same time, there is also evidence that such learning was, in any case, partial and selective – and that references to PISA findings were also used with legitimizing purposes. As noted by Verger and Curran (2014), the attention to certain practices promoted by the OECD (including external assessment) among Catalan policy-makers contrasts with the neglect of other recommendations advanced by the same organization (for instance, the need to combine school-level reforms with system-level reforms). Similarly, certain recommendations have been re-interpreted and adopted in a selective, interested way. This is the case of OECD advice regarding school autonomy. While OECD products have tended to emphasize the potential of *pedagogic* autonomy (given its positive

association with school effectiveness), recent policy changes in the Catalan context have tended to focus on the devolution of managerial tasks to the school level, thus privileging the advance of *managerial* autonomy. Overall, this suggests that the recommendations deriving from PISA, as well as other sources of OECD policy advice, simultaneously serve learning and legitimization purposes.

The cases of Italy and Ireland, in turn, are illustrative for the emulation dynamics triggered by PISA-based OECD recommendations. According to the reviewed literature, the advance of accountability and assessment reforms in these contexts owes much to the role of the OECD in the promotion of an “evaluation culture” – and the need or interest of these countries to “comply with” such recommendations. The adoption of national assessments, evaluation and autonomy systems would not be driven by a logic of consequences (as it did not intend to address any particular problem) but rather by a logic of appropriateness (that is, by the symbolic or legitimizing power of such reforms). In the case of Italy, for instance, Grimaldi and Serpieri (2014) observe that international comparisons have favored the advance of education policies inspired by the logic of benchmarking, and that PISA results in particular played a key role in creating an appetite for a culture of evaluation. Such evaluation culture, however, would have long remained a rhetoric device before penetrating the level of practice – Italy is regarded as a late-adopter of standardized testing, and schools’ and teachers’ evaluation arrangements were not launched until 2010 under the form of pilot programs (see similar findings for the case of Ireland in McNamara/O’Hara/Boyle/Sullivan 2009).

4. Conclusion

PISA’s role in the international dissemination of policy ideas such as accountability and assessment in education is multifaceted. The most evident policy transfer mechanism through which PISA promotes changes in accountability and assessment policies at the country level is competition. Competition, “shame and blame” dynamics and performative pressures are powerful and particularly well-theorized triggers of policy change, although they do not suffice to explain how policy diffusion happens in the educational domain. Beyond competition, we have also observed how the OECD, through PISA and PISA-related initiatives, has been able to trigger the mechanisms of policy learning and emulation as well.

Despite the centrality of the competition mechanism to understand PISA’s influence, more research is necessary to gain further understanding of which countries are more likely to adopt a competitive mindset and behavior

in the context of education reform. For instance, shall we assume that poor-performers or those “lagging behind” face greater reform pressure? Or, would rather the impact of PISA among “mid-performers” (Germany, Denmark, and Norway) suggest that the gap between self-perception and PISA results are a more powerful trigger of policy change? Also, it would be interesting to gain insight into the pressures resulting from high performance in PISA, and the challenges that league leaders face to sustain the reputational capital that comes with outstanding PISA results.

Our findings do not take for granted that there is some form of intentionality behind the PISA program to influence countries’ policies. Despite the existing evidence of the policy effects of PISA, which in this chapter we have illustrated by focusing on accountability and assessment reforms, these effects cannot be exclusively attributed to PISA (not even to PISA-based advice). Instrumentalization dynamics on the reception side (i.e. countries), as well as the analytic work produced in other OECD divisions, might be of great(er) relevance to explain the international diffusion of the accountability agenda. Overall, we argue that PISA is useful in “making the case” for education reform, but that the content and approach of these reforms is more likely to be shaped by the policy work conducted in other OECD units and teams (i.e. not only through the “translation” of PISA data into policy advice, but also through a variety of products that are not necessarily based on PISA, or in which PISA results play a secondary or auxiliary role). Future research could delve into the micro-politics of the OECD in order to understand to what extent/whether there is a significant degree of coordination between different OECD operational units and governing boards, or to what extent the PISA governing board and the PISA staff are aware of the policy usages given to the assessment results, and whether they would prefer that PISA policy effects move in a different direction.

References

- Baxter, Jacqueline/Clarke, John (2013): Farewell to the tick box inspector? Ofsted and the changing regime of school inspection in England. In: *Oxford Review of Education*, 39, 5, pp. 702-718.
- Bloem, Simone (2015): The OECD directorate for education as an independent knowledge producer through PISA. In: Kotthoff, Hans Georg/Klerides, Eleftherios (eds.): *Governing Educational Spaces*. Rotterdam: Sense Publishers, pp. 169-185.
- Breakspear, Simon (2012): The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. OECD Education Working Papers, No. 71. Paris: OECD Publishing.

https://www.oecd-ilibrary.org/education/the-policy-impact-of-pisa_5k9fdfqffr28-en

- Busch, Per-Olof/Jørgens, Helge (2005): The international sources of policy convergence: explaining the spread of environmental policy innovations. In: *Journal of European Public Policy*, 12, 5, pp. 860-884.
- Camphuijsen, Marjolein/Skedsmo, Guri/Møller, Jorunn (2018): School autonomy with accountability as a global education reform: Its adoption and re-contextualization in the Norwegian context. Paper presented at the European Conference on Educational Research (ECER). Bolzano, Italy.
- Carvalho, Luís Miguel/Costa, Estela (2014): Seeing education with one's own eyes and through PISA lenses: considerations of the reception of PISA in European countries. In: *Discourse: Studies in the Cultural Politics of Education*, 36, 5, pp. 638-646.
- Dobbin, Frank/Simmons, Beth/Garrett, Geoffrey (2007): The global diffusion of public policies: Social construction, coercion, competition, or learning? In: *Annual Review of Sociology*, 33, pp. 449-472.
- Dobbins, Michael/Christ, Claudia (2019): Do they matter in education politics? The influence of political parties and teacher unions on school governance reforms in Spain. In: *Journal of Education Policy*, 34, 1, pp. 61-82.
- Dolowitz, David/Marsh, David (1996): Who Learns What from Whom: A Review of the Policy Transfer Literature. In: *Political Studies*, 44, 2, pp. 343-357.
- Dolowitz, David/Marsh, David (2000): Learning from abroad: The role of policy transfer in contemporary policy-making. In: *Governance*, 13, 1, pp. 5-23.
- Doshi, Rush/Kelley, Judith G./Simmons, Beth A. (2004): The Power of Ranking: The Ease of Doing Business Indicator and Global Regulatory Behavior. In: *International Organization*, 73, pp. 1-33.
- Engel, Laura C. (2015): Steering the national: exploring the education policy uses of PISA in Spain. In: *European Education*, 47, 2, pp. 100-116.
- Gilardi, Fabrizio (2012): Transnational Diffusion: Norms, Ideas, and Policies. In: Carlsnaes, Walter/Risse, Thomas/Simmons, Beth (eds.): *Handbook of International Relations* (2nd ed.). Zurich: SAGE Publications, pp. 453-477.
- Gilbert, Paul Robert (2015): Commentary: The ranking explosion. In: *Social Anthropology*, 23, 1, pp. 83-86.
- Gorur, Radhika (2013): My school, my market. In: *Discourse: Studies in the Cultural Politics of Education*, 34, 2, pp. 214-230.
- Gorur, Radhika (2016): Seeing like PISA: A cautionary tale about the performativity of international assessments. In: *European Educational Research Journal*, 15, 5, pp. 598-616. doi: 10.1177/1474904116658299
- Grek, Sotiria (2009): Governing by numbers: The PISA effect in Europe. In: *Journal of Education Policy*, 24, 1, pp. 23-37.
- Grimaldi, Emiliano/Serpieri, Roberto (2014): Italian education beyond hierarchy: Governance, evaluation and headship. In: *Educational Management Administration and Leadership*, 42, 4S, pp. 119-138.
- Hatch, Thomas (2013): Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. In: *Journal of Educational Change*, 14, 2, pp. 113-138.

- Holzinger, Katharina/Knill, Christoph (2005): Causes and conditions of cross-national policy convergence. In: *Journal of European Public Policy*, 12, 5, pp. 775-796.
- Hopkins, David/Pennock, Dianne/Ritzen, Jo/Ahtaridou, Elpida/Zimmer, Karin (2008): External Evaluation of the Policy Impact of PISA (EDU/PISA/GB(2008)35/REV1). Paris: OECD Publishing.
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB\(2008\)35/REV1&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB(2008)35/REV1&docLanguage=En)
- Hursh, David (2007): Assessing No Child Left Behind and the rise of neoliberal education policies. In: *American Educational Research Journal*, 44, 3, pp. 493-518.
- Maggetti, Martino/Gilardi, Fabrizio (2016): Problems (and solutions) in the measurement of policy diffusion mechanisms. In: *Journal of Public Policy*, 36, 1, pp. 87-107.
- Marcussen, Martin (2004): Multilateral surveillance and the OECD: Playing the idea game. In: Armington, Klaus/Beyeler, Michelle (eds.): *The OECD and European welfare states*. Cheltenham, UK/Northampton, MA: Edward Elgar Publishing, pp. 13-31.
- Marsh, David/Sharman, J. C. (2009): Policy diffusion and policy transfer. In: *Policy Studies*, 30, 3, pp. 269-288.
- Martens, Kerstin (2007): How to become an influential actor – The “comparative turn” in OECD education policy. In: Martens, Kerstin/Rusconi, Alessandra/Lutz, K. (eds.), In: *Transformations of the state and global governance*. London: Routledge, pp. 40-56.
- Martens, Kerstin/Jakobi, Anja P. (2010): Introduction: The OECD as an actor in international politics. In: Martens, Kerstin/Jakobi, Anja P. (eds.): *Mechanisms of OECD governance: International incentives for national policy-making?* Oxford: Oxford University Press, pp. 1-25.
- Martens, Kerstin/Niemann, Dennis (2013): When do numbers count? The differential impact of the PISA rating and ranking on education policy in Germany and the US. In: *German Politics*, 22, 3, pp. 314-332.
- McNamara, Gerry/O'Hara, Joe/Boyle, Richard/Sullivan, Conor (2009): Developing a culture of evaluation in the Irish public sector: the case of education. In: *Evaluation*, 15, 1, pp. 101-112.
- Meseguer, Covadonga (2004): What role for learning? The diffusion of privatisation in OECD and Latin American countries. In: *Journal of Public Policy*, 24, 3, pp. 299-325.
- Meyer, Heinz-Dieter (2014): The OECD as pivot of the emerging global educational accountability regime: How accountable are the accountants. In: *Teachers College Record*, 116, 9, pp. 1-20.
- Moos, Lejf (2014): Educational governance in Denmark. In: *Leadership and Policy in Schools*, 13, 4, pp. 424-443.
- Niemann, Dennis/Martens, Kerstin/Teltemann, Janna (2017): PISA and its consequences: Shaping education policies through international comparisons. In: *European Journal of Education*, 52, 2, pp. 175-183.
- Niemann, Dennis/Martens, Kerstin (2018): Soft governance by hard fact? The OECD as a knowledge broker in education policy. In: *Global Social Policy*, 18, 3, pp. 267-283.

- OECD (2011): PISA in Focus, No. 9. School autonomy and accountability: Are they related to student performance? Paris: OECD Publishing.
<https://www.oecd.org/pisa/pisaproducts/pisainfocus/48910490.pdf>
- Popp, Marie (2014): New Culture, Old System – Reactions to Internationalization in Spanish Education Policy. In: Martens, Kerstin/Knodel, Philipp/Windzio, Michael (eds.): Internationalization of Education Policy: A new constellation of statehood in education? London: Palgrave Macmillan, pp. 163-178.
- Sellar, Sam/Lingard, Bob (2013): The OECD and global governance in education. In: Journal of Education Policy, 28, 5, pp. 710-725.
- Sellar, Sam/Thompson, Greg/Rutkowski, David (2017): The global education race: Taking the measure of PISA and international testing. Edmonton, Canada: Brush Education.
- Shipan, Charles R./Volden, Craig (2008): The mechanisms of policy diffusion. In: American Journal of Political Science, 52, 4, pp. 840-857.
- Sjøberg, Sven (2016): OECD, PISA, and globalization: The influence of the international assessment regime. In: Tienken, Christopher H./Mullen, Carol A. (eds.): Education policy perils: Tackling the tough issues. New York: Routledge, pp. 114-145.
- Steiner-Khamsi, Gita/Waldow, Florian (2018): PISA for scandalisation, PISA for projection: the use of international large-scale assessments in education policy making – an introduction. In: Globalisation, Societies and Education, 16, 5, pp. 557-565.
- Stone, Diane (2012): Transfer and translation of policy. In: Policy Studies, 33, 6, pp. 483-499.
- Verger, Antoni/Fontdevila, Clara/Parcerisa, Lluís (2019a): Constructing school autonomy with accountability as a global policy model: A focus on OECD's governance mechanisms. In: Ydesen, Christian (ed.): The OECD's Historical Rise in Education: The Formation of a Global Governing Complex. Palgrave, pp. 219-243.
- Verger, Antoni/Fontdevila, Clara/Parcerisa, Lluís (2019b): Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. In: Discourse: Studies in the Cultural Politics of Education, 40, 2, pp. 248-270.
- Verger, Antoni/Curran, Marta (2014): New public management as a global education policy: Its adoption and re-contextualization in a Southern European setting. In: Critical Studies in Education, 55, 3, pp. 253-271.
- Waldow, Florian (2017): Projecting images of the 'good' and the 'bad school': Top scorers in educational large-scale assessments as reference societies. In: Compare: A Journal of Comparative and International Education, 47, 5, pp. 647-664.
- Wößmann, Ludger/Lüdemann, Elke/Schütz, Gabriela/West, Martin R. (2007): School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003. OECD Education Working Papers, No. 13. Paris: OECD Publishing. https://www.oecd-ilibrary.org/education/school-accountability-autonomy-choice-and-the-level-of-student-achievement_246402531617

An Average Is Just an Average: What Do We Know About Countries' Low- and High-Performing Students in Mathematics?

David C. Miller¹ and Frank T. Fonseca²

1. Introduction

With the recent release of results from the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA), once again country rankings based on average country performance dominated news headlines around the world (Coughlan 2016; Gurney-Read 2016; Anderson/Shendruk 2019; Lofgren 2016; Wright 2019; Omirgazy 2016). Unfortunately, country rankings and average student performance do not provide information about equity, which is a key factor in evaluating the quality of an education system (Scheerens/Hendriks 2004). These results provide insufficient information about a country's success in educating its low- and high-performing students, who need an appropriate and challenging education if they are to become contributing members of society (Badescu/D'Hombres/Villalba 2011; Barone/van de Werfhorst 2011).

Published reports from large-scale international assessments, including TIMSS and PISA, have included tables with percentiles of achievement that show, for example, how scores at the 10th and 90th percentiles compare across countries. However, there has been very little research systematically investigating and statistically testing these achievement gaps and examining whether they have narrowed or widened over time.

A plethora of research on the effects of schooling starting with the landmark release of the Coleman Report in the United States (Coleman et al. 1966) and the Plowden Report in the United Kingdom (Peaker 1971; Plowden 1967) has suggested that the majority of the variance in academic achievement could be explained by a student's experiences and socioeconomic background prior to entering school and that differences in the quality of schools and teachers has only a small positive impact on student outcomes. However, subsequent research by Heyneman and Loxley (1983) found that in low-income countries, school-level factors were more important

1 David C. Miller retired in May 2019 as Managing Researcher at the American Institutes for Research (AIR), Washington D.C. where he worked for 20 years. Email: dcfm1000@gmail.com

2 Frank T. Fonseca is a Research Associate at the American Institutes for Research (AIR). Email: ffonseca@air.org

than student-level characteristics such as family socioeconomic status in determining academic achievement. Stemming from this theoretical framework, this analysis uses country-level data to examine the relationship between income inequality and mathematics achievement gaps. Prior research has not specifically examined this relationship. We hypothesize that, at the country level, the more unequal the income distribution is, the larger the mathematics achievement gaps between low- and high-performing students. Furthermore, we hypothesize that the correlation between income inequality and mathematics achievement gaps will be stronger among industrialized OECD countries and weaker among less developed (non-OECD) countries. In sum, this paper will address the following research questions:

- What is the extent of the variation seen across education systems in the mathematics achievement of low- and high-performing students, especially relative to average performance within education systems?
- What is the extent of the variation seen across education systems in the size of students' within-country achievement gaps in mathematics and how is the size of these achievement gaps related to education systems' average performance?
- Across education systems, has the mathematics achievement of low- and high-performing students changed over time, and how does this compare with the change in education systems' average performance?
- Across education systems, has the size of the mathematics achievement gaps changed over time?
- Using country-level data, what is the relationship between income inequality and mathematics achievement gaps?

2. Method

2.1 Data source and procedure

Using fourth- and eighth-grade mathematics data from the 1995 and 2015 administrations of the Trends in International Mathematics and Science Study (TIMSS), this analysis examines cross-national differences in the achievement of low- and high-performing students, especially relative to average performance within education systems. It does so by examining average scores and the cut-point scores of each education system at the 10th and 25th percentiles (representing the low side of the achievement distribution) and the 75th and 90th percentiles (representing the high side of the achievement distribution). In these analyses, the achievement gap between

low- and high-performing students in each education system is represented by the difference between the 10th percentile and 90th percentile cut-point scores.

We chose to use data from just the first round of TIMSS (1995) and the latest round (2015) because we felt that including all other rounds of TIMSS data collection (e.g., 1999, 2003, etc.) would introduce too much data and complexity into the analyses, the graphs, and the results. Also, many education systems have not participated in all rounds of TIMSS data collection, and thus missing data can be a growing concern as more rounds of data collection are included. As it is, in this paper, the analyses with 2015-only data included the 48 education systems that participated in TIMSS at fourth grade and the 37 that participated at eighth grade in 2015, while analyses using data at the two time points were limited to the 17 education systems that participated in TIMSS at fourth grade and the 16 that participated at eighth grade in 1995 and 2015. Germany, for example, participated in TIMSS 2015 at grade 4 but not grade 8, and in 1995 Germany did not participate in TIMSS at grade 4. Thus, Germany is only included in analyses with the 48 education systems with 2015 data at grade 4. The respective international averages that appear included these 48, 37, 17, or 16 education systems, with each education system weighted equally. Benchmarking education systems that participated in TIMSS, such as U.S. states and Canadian provinces, were not included in the analyses.

TIMSS 2015 data is used to address the first research question by examining cross-national differences in average mathematics achievement and performance at the 10th, 25th, 75th, and 90th percentiles.

Another way to evaluate cross-national variations in the mathematics performance of low and high performers is to graphically plot scores at the 10th percentile (shown on the x-axis) and the 90th percentile (shown on the y-axis). When categorized in this way, education systems generally appear in one of four quadrants: (1) top right: both low and high performers scored high relative to the international average, (2) bottom left: both low and high performers scored low relative to the international average, (3) bottom right: low performers scored high and high performers scored low relative to the international average, and (4) top left: low performers scored low and high performers scored high relative to the international average.

TIMSS 2015 data is also used to address the second research question, in which we examine the size of the within-country performance gaps in mathematics. Doing so allows us to differentiate those education systems that have a more equitable distribution of student performance (i.e., a relatively small point difference between mathematics scores at the 10th and 90th percentiles) and those education systems that have a relatively large performance gap between low- and high-performing students.

The third research question is addressed using TIMSS data from two time points: 1995 and 2015. Across education systems, we examine changes over

time in average mathematics scores and in mathematics scores at the 10th, 25th, 75th, and 90th percentiles.

TIMSS 1995 and 2015 data is also used to address the fourth research question. Across education systems, we examine whether the point difference between mathematics scores at the 10th and 90th percentiles significantly narrowed or widened during the 20-year time period from 1995 to 2015. Thus, the achievement gap at each year is calculated by subtracting the cut-point score at the 10th percentile from the cut-point score at the 90th percentile, and the change in gaps is calculated by subtracting the 1995 gap from the 2015 gap.

In testing the fifth research question, income inequality is measured using the Gini coefficient. This is a measure of statistical dispersion that represents the income distribution of a country's residents. A value of 0 represents perfect income equality, while a value of 100 represents absolute income inequality. The Gini coefficient can be derived from various sources, and for this analysis the World Bank will serve as the primary data source.

2.2 Target population

The international target populations in TIMSS are defined in terms of the amount of schooling students have received. The 2011 International Standard Classification of Education (ISCED) by UNESCO provides an internationally accepted classification scheme for describing levels of schooling across countries (ISCED 2012). The ISCED system describes the complete range of schooling, from pre-primary (level 0) to the doctoral level (level 8). ISCED level 1 corresponds to primary education (or the first stage of basic education). “The boundary between ISCED level 0 and level 1 coincides with the transition point in an education system where systematic teaching and learning in reading, writing and mathematics begins” (UNESCO 2012: 30). In TIMSS at the lower grade, the international target population is defined as all students in their fourth year of formal schooling counting from the first year of ISCED level 1; at the upper grade it is defined as all students in their eighth year of formal schooling counting from the first year of ISCED level 1 (LaRoche/Joncas/Foy 2016). However, given the cognitive demands of the TIMSS assessments, an effort is made to avoid assessing very young students. Thus, TIMSS recommends assessing the next higher grade – i.e., fifth grade for fourth-grade TIMSS and ninth grade for eighth-grade TIMSS – if, for fourth graders, the average age at the time of testing would be less than 9.5 years and, for eighth graders, less than 13.5 years.

2.3 Data analysis

Most of the analyses in this paper were carried out using the TIMSS International Data Explorer (IDE), which is a free online tool for producing tables and doing statistical analyses with the TIMSS data (<http://nces.ed.gov/surveys/international/ide/>). Using the TIMSS IDE, estimates were produced from cross-tabulations of the data, and *t* tests were performed to test for differences between estimates. SPSS statistical software was used to compute correlation coefficients.

All of the estimates and comparisons that are discussed in this paper are statistically significant at the $p < .05$ level to ensure that they are larger than what might be expected due to sampling variation. No adjustments were made for multiple comparisons.

3. Results

The results are presented in five main sections in response to the research questions.

Research Question 1: *What is the extent of the variation seen across education systems in the mathematics achievement of low- and high-performing students, especially relative to average performance within education systems?*

There were considerable cross-national differences when examining average mathematics achievement and performance at the 10th, 25th, 75th, and 90th percentiles. To illustrate these differences, a few country comparisons using the most recent 2015 data are presented below. As a first example, the United States scored higher than Slovenia at grade 4 on average and at all points shown along the achievement distribution except for the 10th percentile (table 1a). Thus, although U.S. fourth-graders generally outperformed their peers in Slovenia, the lowest-performing students scored similar in both countries. At grade 8, the pattern is a little different: While the United States and Slovenia scored similar on average, high-performing students did better in the United States than in Slovenia and the lowest-performing students did better in Slovenia.

Table 1a. Differences in fourth- and eighth-grade students' average mathematics scores and cut-point scores at the 10th, 25th, 75th, and 90th percentiles in the United States and Slovenia: 2015

	Educa- tion system	Cut- point score at 10 th percen- tile	Cut- point score at 25 th percen- tile	Aver- age score	Cut- point score at 75 th percen- tile	Cut- point score at 90 th percen- tile
Grade 4	United States	432	▲485	▲539	▲596	▲640
	Slovenia	430	476	520	568	605
Grade 8	United States	408	461	518	▲577	▲624
	Slovenia	▲425	470	516	564	605

▲ Score is higher than the pairwise comparison score at $p < .05$.
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2015.

Next, we compare Hungary to the Netherlands at grade 4 and Hungary to Norway at grade 8. As shown in table 1b, Hungary is a country that scored similar, on average, to the Netherlands at grade 4 and Norway at grade 8. However, high-performing students did better in Hungary than in the Netherlands and Norway, and low-performing students did better in the Netherlands and Norway at these respective grade levels.

Table 1b. Differences in average mathematics scores and cut-point scores at the 10th, 25th, 75th, and 90th percentiles in Hungary and the Netherlands at grade 4 and Hungary and Norway at grade 8: 2015

	Education system	Cut-point score at 10 th percentile	Cut-point score at 25 th percentile	Average score	Cut-point score at 75 th percentile	Cut-point score at 90 th percentile
Grade 4	Hungary	412	475	529	▲591	▲635
	Netherlands	▲457	▲492	530	569	601
Grade 8	Hungary	390	452	514	▲582	▲632
	Norway	▲420	▲465	512	560	600

▲ Score is higher than the pairwise comparison score at $p < .05$.
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2015.

Turkey scored lower than the Netherlands at grade 4 on average and at all points shown along the achievement distribution except for the 90th percentile (table 1c). That is, despite Turkey scoring almost 50 points lower than the Netherlands on average, there was no measurable difference between the highest-performing fourth-graders in these two countries. Comparing Turkey to Chile at grade 8, the results show that students in Turkey outperformed students in Chile on average and on the high side of the achievement distribution. There were no measurable differences between Turkey and Chile on the low side of the achievement distribution.

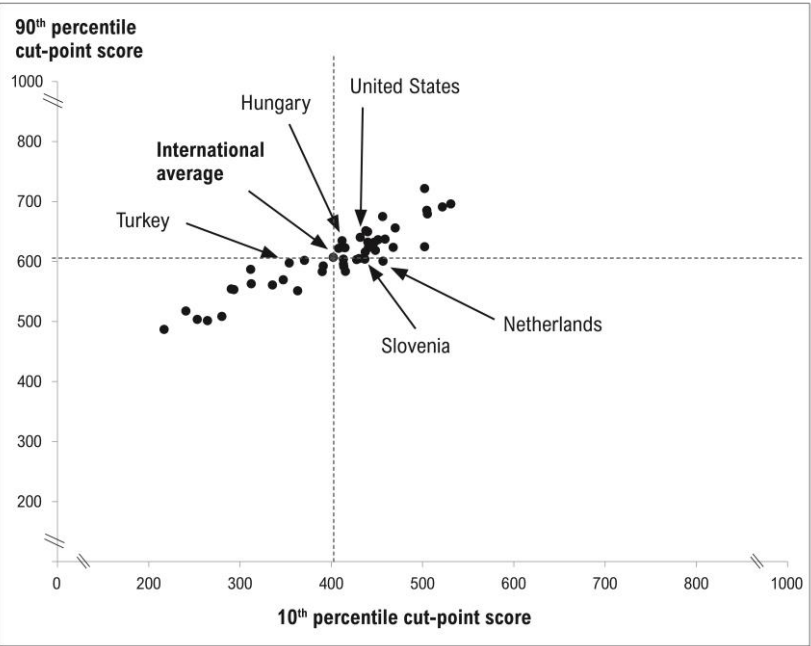
Table 1c. Differences in average mathematics scores and cut-point scores at the 10th, 25th, 75th, and 90th percentiles in Turkey and the Netherlands at grade 4 and Turkey and Chile at grade 8: 2015

	Educa- tion system	Cut- point score at 10 th percen- tile	Cut- point score at 25 th percen- tile	Aver- age score	Cut- point score at 75 th percen- tile	Cut- point score at 90 th percen- tile
Grade 4	Turkey	354	424	483	551	598
	Nether- lands	▲457	▲492	▲530	▲569	601
Grade 8	Turkey	324	385	▲458	▲531	▲599
	Chile	323	372	427	482	531

▲ Score is higher than the pairwise comparison score at $p < .05$.
Source: International Association for the Evaluation of Educational Achievement (IEA), Trends in International Mathematics and Science Study (TIMSS) 2015.

Figures 1a and 1b (for grades 4 and 8, respectively) serve as another way of representing the results discussed in examining tables 1a through 1c. Consistent with what the tables show, the United States and Hungary appear in the top right quadrant; the Netherlands, Norway, and Slovenia lie within or very close to the bottom right quadrant; Turkey lies within or very close to the upper left quadrant; and Chile appears in the bottom left quadrant of figure 1b. Turkey and the Netherlands serve as a particularly striking example when looking at figure 1a as well as table 1c. At grade 4, the lowest-performing students scored more than 100 points higher in the Netherlands compared to Turkey, while these countries’ highest-performing students performed similarly.

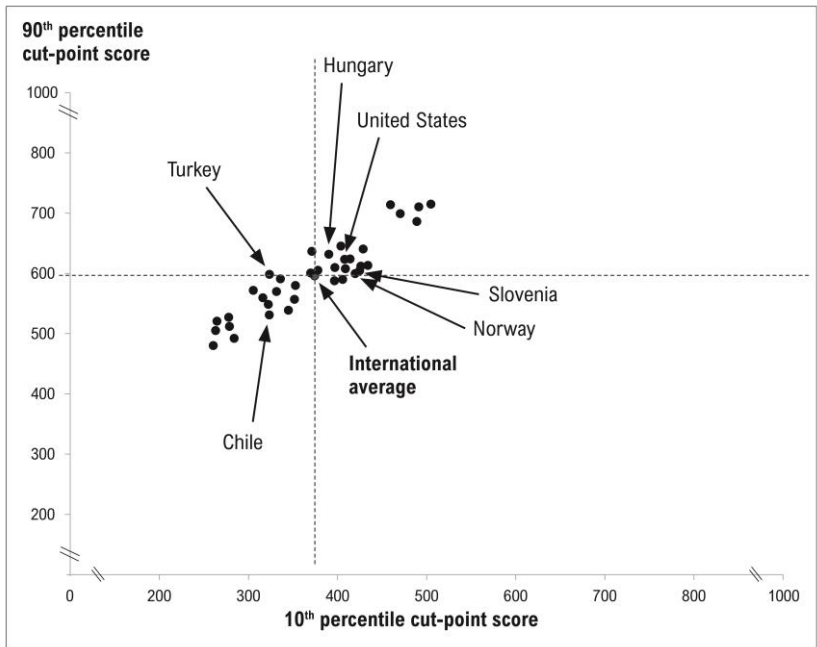
Figure 1a. Cut-point scores of fourth-grade students in mathematics at the 10th and 90th percentiles, by education system: 2015



Note: Each of the 48 education systems included in this analysis appears as a dot, and several education systems highlighted in the text are labeled here for illustrative purposes. The international average is the average of these 48 education systems, with each one weighted equally. The dotted lines are vertical and horizontal lines intersecting at the international average.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 2015.

Figure 1b. Cut-point scores of eighth-grade students in mathematics at the 10th and 90th percentiles, by education system: 2015



Note: Each of the 37 education systems included in this analysis appears as a dot, and several education systems highlighted in the text are labeled here for illustrative purposes. The international average is the average of these 37 education systems, with each one weighted equally. The dotted lines are vertical and horizontal lines intersecting at the international average.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 2015.

Research Question 2: *What is the extent of the variation seen across education systems in the size of students' within-country achievement gaps in mathematics and how is the size of these achievement gaps related to education systems' average performance?*

In addressing the second set of research questions, we define an achievement gap as the distance between the 10th and 90th percentile cut-point scores. For each grade, we ordered the education systems in a figure by the size of the achievement gap, from smallest (at the top) to largest (at the bottom). Due to space constraints, the figures shown here (2a and 2b) do not show all education systems, but rather, just the top seven, the middle seven (including the international average), and the bottom seven.

In 2015, the size of the mathematics achievement gaps varied substantially across education systems. For example, the Netherlands and Belgium (Flemish) at grade 4 (144 and 156 points, respectively) and Canada, Slovenia, and Norway at grade 8 (180 points in all three) had a more equitable distribution of student performance, while Oman, Qatar, and the United Arab Emirates had relatively large performance gaps at both grades (250 points or more in all three at both grades) (figures 2a and 2b). In the United States, the gap was 209 points at grade 4 and 216 points at grade 8.

The gaps tended to be smaller at grade 4 than at grade 8 (for the education systems on average, 205 compared to 222 points, respectively), though the difference between the largest gap and the smallest gap (i.e., Jordan and the Netherlands at grade 4 and Turkey and Canada at grade 8) was larger at grade 4 than at grade 8 (133 compared to 96 points, respectively).

Figure 2a. Average mathematics scores and achievement gaps of fourth-grade students at the 10th, 25th, 75th, and 90th percentiles, by education system: 2015

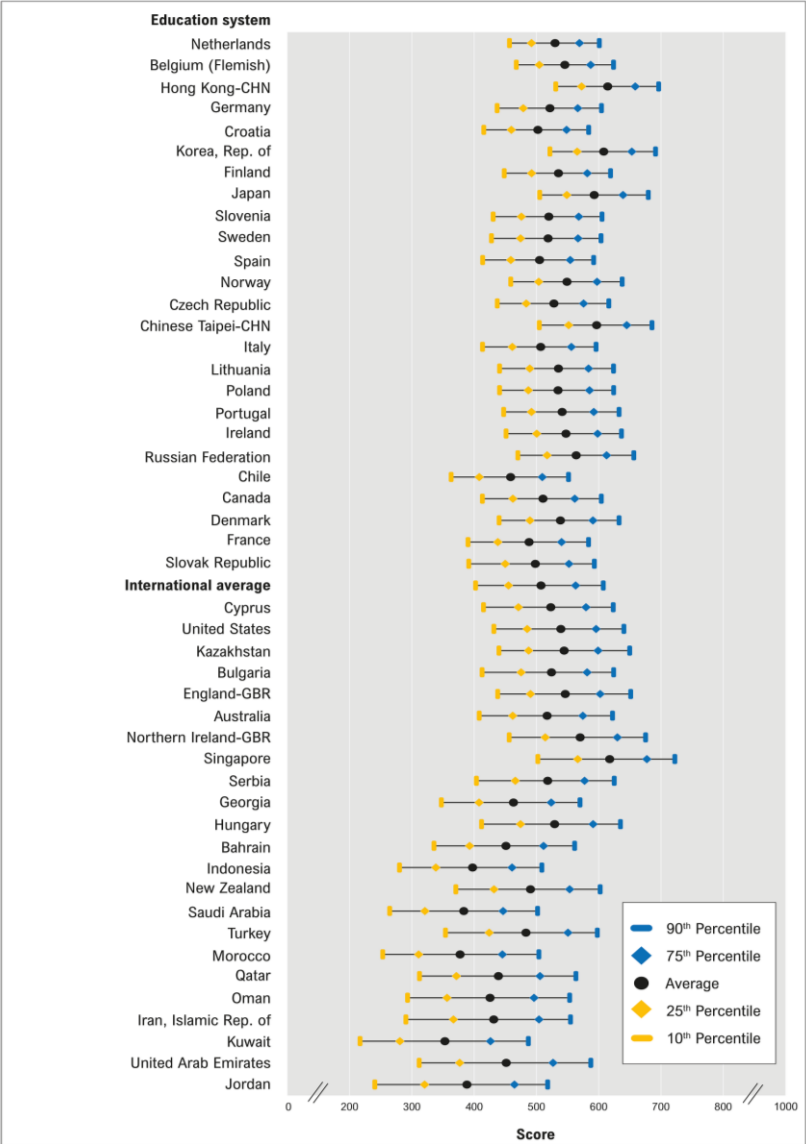
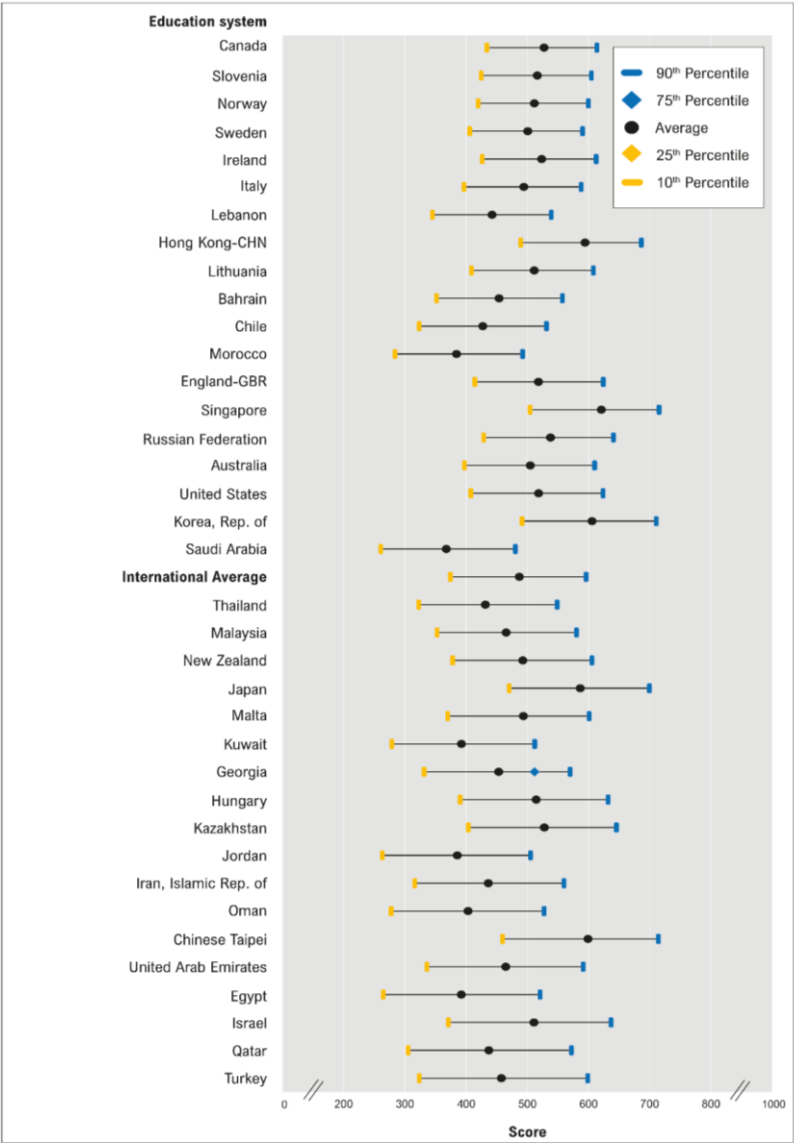


Figure 2b. Average mathematics scores and achievement gaps of eight-grade students at the 10th, 25th, 75th, and 90th percentiles, by education system: 2015



Using this graphical representation, it is also possible to investigate whether the variation in the size of these within-country achievement gaps is related to the variation in education systems' overall average mathematics performance. For example, do education systems that have low mathematics scores, on average, also tend to have small achievement gaps between low- and high-performing students?

As shown in figures 2a and 2b, education systems that scored very low on average (e.g., Jordan and Kuwait at grade 4 and Egypt and Oman at grade 8) also tended to have some of the largest achievement gaps, and education systems that scored very high on average also tended to have some of the smallest achievement gaps – though this latter finding was primarily found at grade 4. When we computed the correlation coefficient between education systems' average mathematics scores and the size of their achievement gaps between low- and high-performing students, we found a fairly strong negative correlation at grade 4 ($r = -.705, p < .001$) and a negative correlation that was much weaker and not statistically significant at grade 8 ($r = -.267, p = .110$). That is, at the country level, smaller achievement gaps tended to be associated with higher average scores at grade 4, but not grade 8.

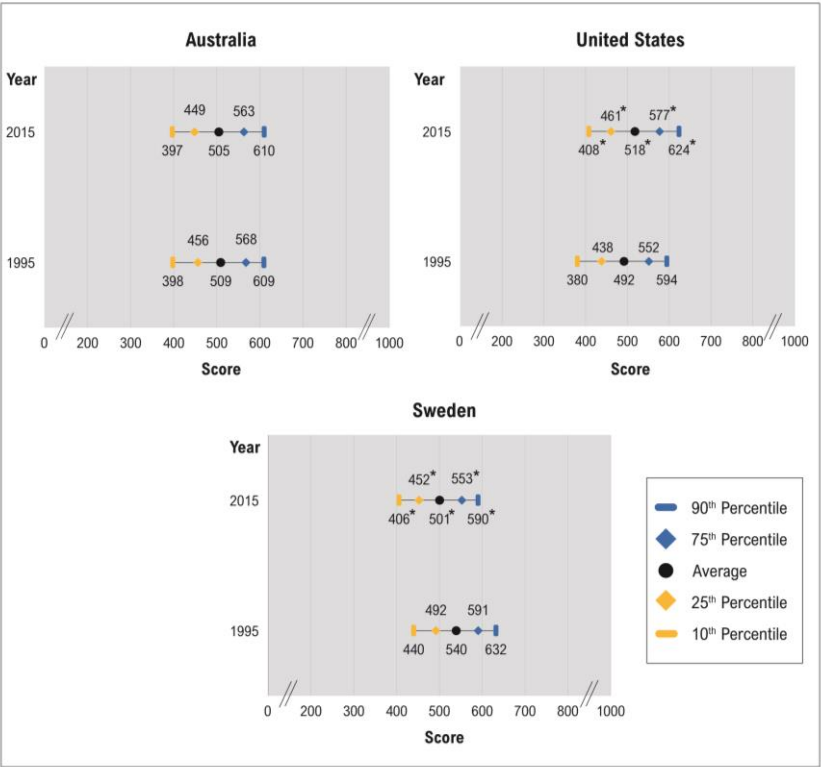
Research Question 3: Across education systems, has the mathematics achievement of low- and high-performing students changed over time, and how does this compare with the change in education systems' average performance?

The third research question was analyzed using TIMSS fourth- and eighth-grade data from 1995 and 2015 to examine changes in average mathematics scores and scores at the 10th, 25th, 75th, and 90th percentiles. The figures in this section are similar to those used to address the second research question; however, each figure focuses on selected education systems and provides data for two time points: 1995 and 2015. Collectively, the set of figures includes examples of education systems that represent different patterns of change or lack thereof from 1995 to 2015 in average mathematics scores and scores at the 10th, 25th, 75th, and 90th percentiles.

To begin, Australia at grade 8 is an example of a country where the data points are similar in 1995 and 2015 and the achievement gaps are almost the same size (figure 3a). Any observed differences between the two years are not statistically significant. In the United States, the performance of eighth-graders increased on average and across the points of the achievement distribution as shown from 1995 to 2015. In Sweden, however, the opposite pattern is seen: The performance of eighth-graders decreased on average and across the points of the achievement distribution as shown from 1995 to 2015.

In Iran at both grades, student performance increased on average from 1995 to 2015, but this improvement did not extend to the low or lowest performing students (figure 3b).

Figure 3a. Differences in average mathematics scores and achievement gaps of eighth-grade students at the 10th, 25th, 75th, and 90th percentiles in Australia, the United States, and Sweden: 1995 and 2015

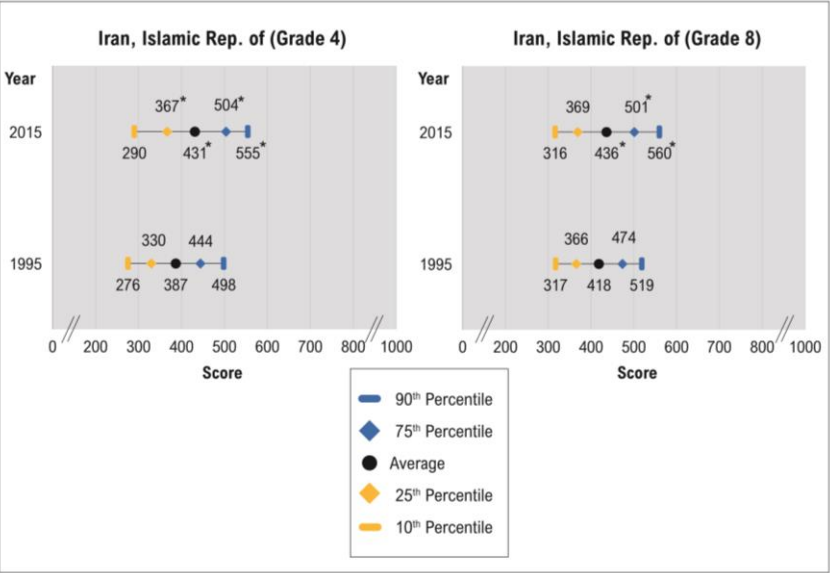


*Average score or cut-point score in 2015 is statistically different than the average score or cut-point score in 1995 at $p < .05$.

Note: The achievement gaps represented here show the distance between the 10th and 90th percentile cut-point scores, with the 25th and 75th percentiles and average scores also shown. For Australia, there are no measurable differences for the average scores and cut-point scores at the 10th, 25th, 75th, and 90th percentiles in 2015 compared to 1995.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 1995 and 2015.

Figure 3b. Differences in average mathematics scores and achievement gaps of fourth- and eighth-grade students at the 10th, 25th, 75th, and 90th percentiles in the Islamic Republic of Iran: 1995 and 2015



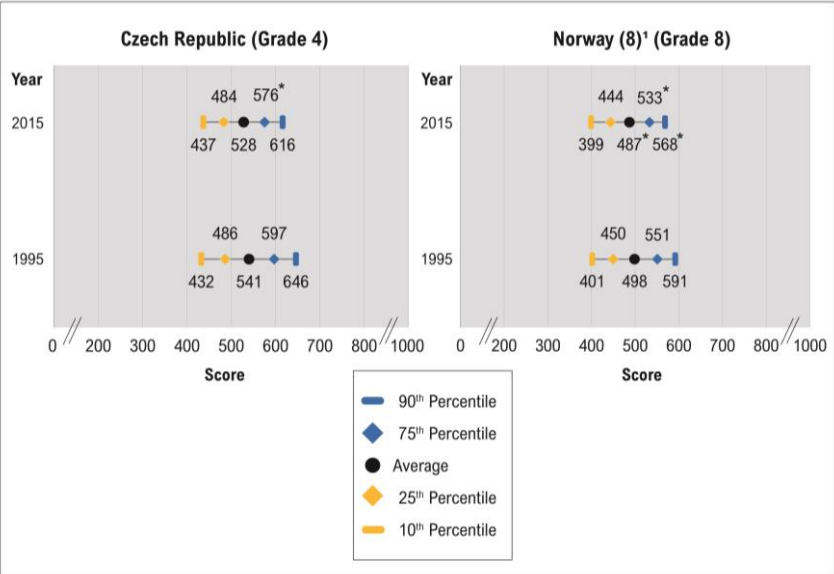
*Average score or cut-point score in 2015 is statistically different than the average score or cut-point score in 1995 at $p < .05$.

Note: The achievement gaps represented here show the distance between the 10th and 90th percentile cut-point scores, with the 25th and 75th percentiles and average scores also shown.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 1995 and 2015.

In the Czech Republic at grade 4 and Norway at grade 8, student performance declined on average from 1995 to 2015 (figure 3c). However, when looking across the achievement distribution, it can be seen that this decrease in performance was limited to the high side of the distribution.

Figure 3c. Differences in average mathematics scores and achievement gaps of fourth-grade students in the Czech Republic and eighth-grade students in Norway at the 10th, 25th, 75th, and 90th percentiles: 1995 and 2015



*Average score or cut-point score in 2015 is statistically different than the average score or cut-point score in 1995 at $p < .05$.

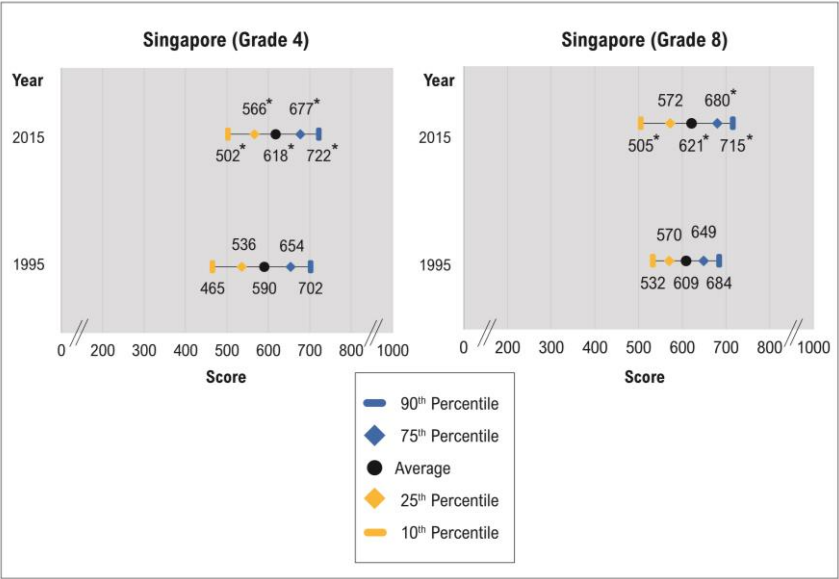
¹ For TIMSS 2015, Norway revised its 8th-grade assessed population to consist of students in their 9th year of schooling to obtain better comparisons with Sweden and Finland. In previous TIMSS cycles, Norway assessed students in their 8th year of schooling, which was defined as 8th grade, but has been redefined as 7th grade because the first year of schooling in Norway is now considered the equivalent of kindergarten. To maintain trend with previous TIMSS cycles, in 2015 Norway also collected data from students in their 8th year of schooling, which is used in trend analyses and shown here as Norway (8).

Note: The achievement gaps represented here show the distance between the 10th and 90th percentile cut-point scores, with the 25th and 75th percentiles and average scores also shown.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 1995 and 2015.

In Singapore, the performance of fourth-graders increased on average and across the points of the achievement distribution as shown from 1995 to 2015 (figure 3d). At grade 8, however, scores increased on average and at the high side of the achievement distribution while scores actually declined for the lowest performing students. Educators and policymakers only looking at average country performance over time would miss these kinds of differences in the achievement of low- and high-performing students.

Figure 3d. Differences in average mathematics scores and achievement gaps of fourth- and eighth-grade students at the 10th, 25th, 75th, and 90th percentiles in Singapore: 1995 and 2015



*Average score or cut-point score in 2015 is statistically different than the average score or cut-point score in 1995 at $p < .05$.

Note: The achievement gaps represented here show the distance between the 10th and 90th percentile cut-point scores, with the 25th and 75th percentiles and average scores also shown.

Source: International Association for the Evaluation of Educational Achievement, Trends in International Mathematics and Science Study (TIMSS), 1995 and 2015.

Research Question 4: *Across education systems, has the size of the mathematics achievement gaps changed over time?*

To answer the fourth research question, the achievement gap at each grade and year for each education system is calculated by subtracting the cut-point score at the 10th percentile from the cut-point score at the 90th percentile; the change in gaps is calculated by subtracting the 1995 gap from the 2015 gap.

From 1995 to 2015, the mathematics achievement gaps tended to narrow at grade 4 but to widen at grade 8. In 11 of the 17 education systems at grade 4, the achievements gaps narrowed, ranging from 12 points in Norway to 45 points in Portugal; only in Iran did the gaps widen (by 41 points) (figure 4a). On average across the 17 education systems, the gaps narrowed by 16 points. At grade 8, however, the gaps widened by an average of 8 points across the 16 education systems (figure 4b). The gaps widened in Hungary (40 points), Iran (42 points), Japan (31 points), and Singapore (59 points), and narrowed in only one country – Norway (21 points). In the United States, there was no statistically significant change in the size of the gap at either grade. Figures 3b and 3c show what produced the changes in the score gaps in Iran and Norway at grade 8. For example, the smaller achievement gap in Norway in 2015 compared to 1995 is the result of the cut-point score for students at the 90th percentile having declined over time.

Figure 4a. Change over time in mathematics achievement gaps of fourth-grade students at the 10th and 90th percentiles, by education system: 1995 and 2015

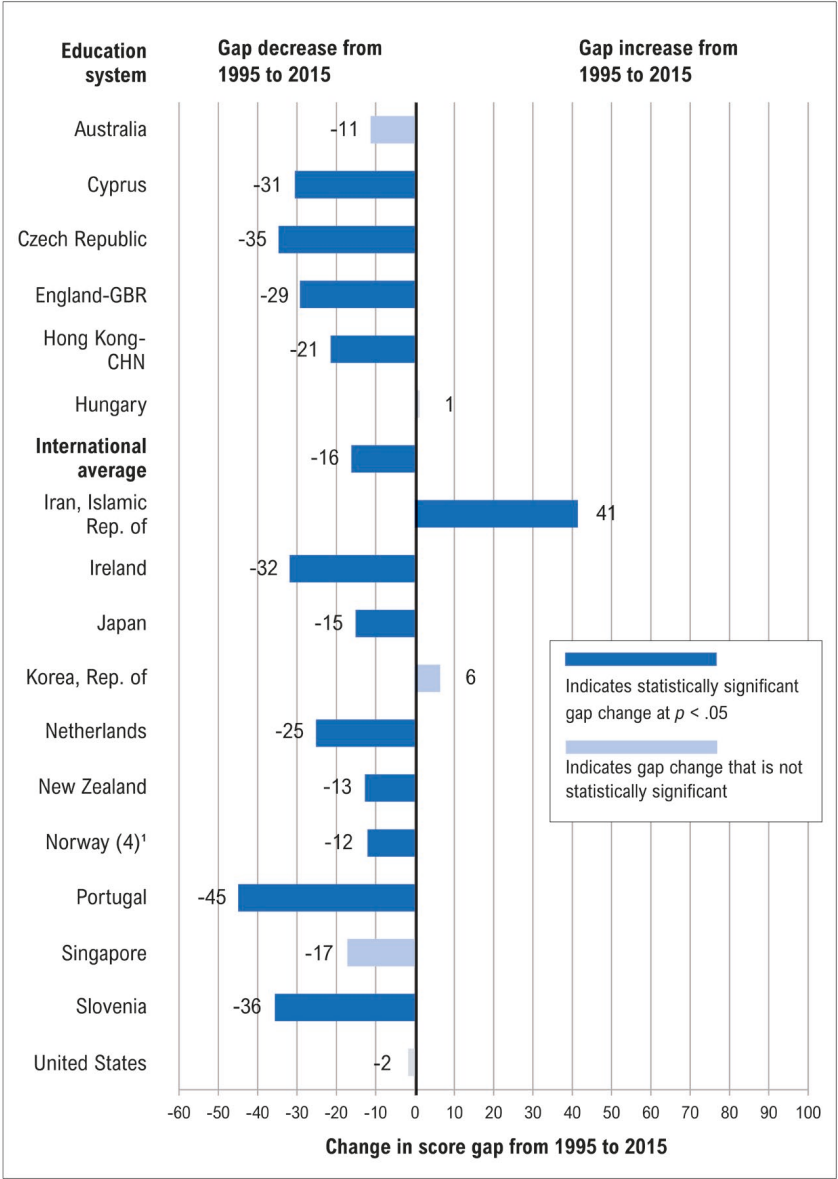
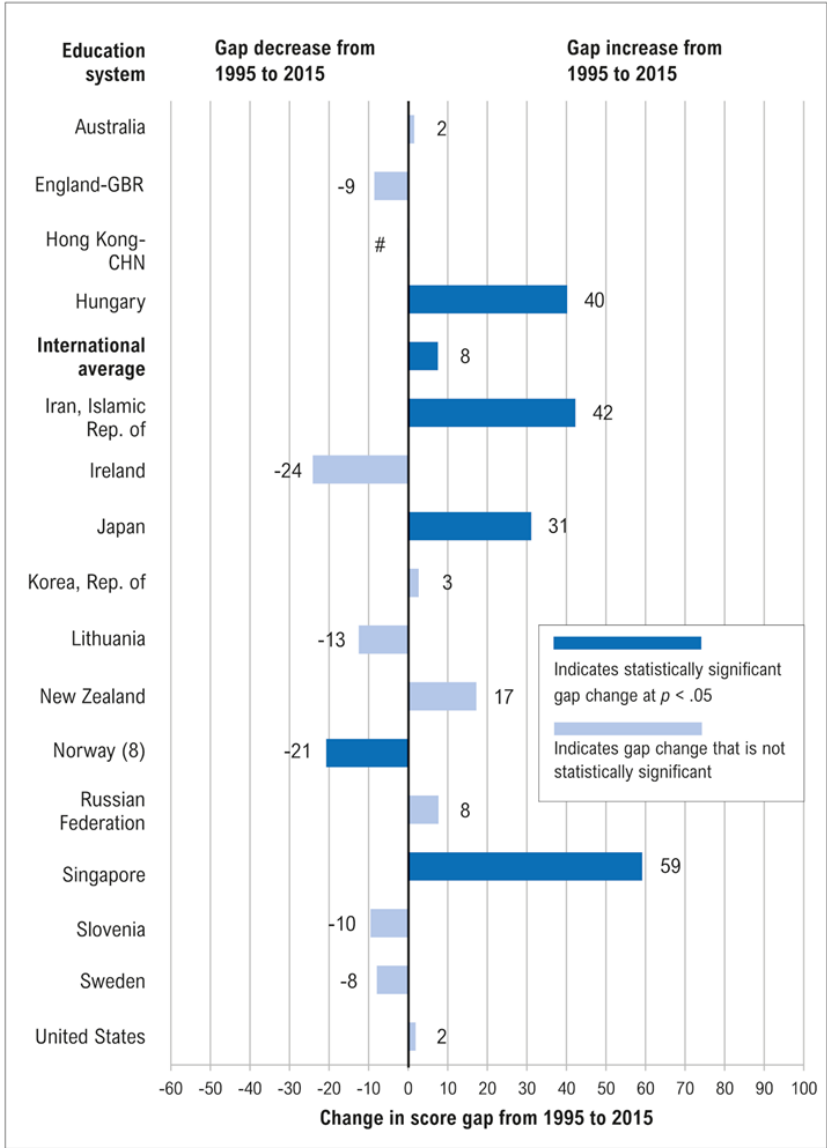


Figure 4b. Change over time in mathematics achievement gaps of eighth-grade students at the 10th and 90th percentiles, by education system: 1995 and 2015



Research Question 5: *Using country-level data, what is the relationship between income inequality and mathematics achievement gaps?*

Using the Gini coefficient and 2015 TIMSS data, the mathematics achievement gaps were related to country-level income inequality, but differently by grade and for OECD compared to non-OECD countries. At grade 4, the relationship was positive for OECD countries ($r = .431, p = .025, N = 27$) and there was no relationship among less developed (non-OECD) education systems: $r = .023, p = .933, N = 16$), which is consistent with what was hypothesized. At grade 8, the relationship was positive and almost statistically significant for OECD countries ($r = .483, p = .058, N = 16$), and for non-OECD education systems a negative relationship was found ($r = -.600, p = .018, N = 15$).

4. Discussion

There are several conclusions that can be drawn from this study. First, there were considerable cross-national differences in the mathematics achievement of low- and high-performing students. Second, the size of mathematics achievement gaps varied substantially across education systems, with some having a more equitable distribution of student performance and others having large performance gaps. Third, examining an education systems' average achievement over time can mask significant change that may be occurring with low- and/or high-performing students, and several examples were provided that illustrate this.

Soon after the release of TIMSS 2015 results, an article appeared in *The Straits Times* titled, "Singapore students top maths, science rankings," and with the following subtitle: "Key global study also shows improvements in reasoning, applied learning and progress made by weaker students" (Teng 2016). The article cited "progress made by weaker students" based on the finding that in 2015 only 1 percent of Singaporean fourth-graders scored below the lowest international TIMSS benchmark in mathematics (i.e., scoring below 400), and this percentage was much lower than for the international average. However, this article fails to mention, as shown in this analysis using percentiles of achievement, the lack of progress made by weaker students at grade 8, where mathematics scores declined for the lowest performing Singaporean students from 1995 to 2015. Although not presented in this paper, the mathematics scores of the lowest performing eighth-graders in Singapore were higher in 1995 than all of the other subsequent years of TIMSS data collection (i.e., 1999, 2003, 2007, and 2011), and there was no statistically significant change in these scores from 2011 to 2015.

Fourth, the study brought to light some troubling findings at eighth grade compared to fourth grade: We found that achievement gaps tended to be larger at eighth grade; smaller achievement gaps tended to be associated with higher average scores at grade 4, but not at grade 8; and the achievement gaps tended to narrow at grade 4 but widen at grade 8 from 1995 to 2015. These grade differences imply that education systems need to be concerned about struggling students falling further behind as they progress from earlier to later years of schooling. Furthermore, these grade differences suggest that efforts by education systems to be both high performing overall and to have a more equitable distribution of student performance become increasingly challenging as students progress through school. Finally, future research might examine if ability tracking, which tends to occur more in the later grades, is contributing to these grade differences. For example, in some countries, such as Austria and Germany, students are tracked by ability into schools enabling to take up university studies afterwards and schools with lower examination degrees as early as age 10, while other countries (e.g., France and the United States) start tracking into different school types or branches much later (Guyon/Maurin/McNally 2012).

It is desirable to have a negative correlation between countries' average mathematics scores and the size of their achievement gaps – that is, smaller achievement gaps between low- and high-performing students associated with higher average scores. Also, given the importance of equity in education, it is desirable for countries to have small achievement gaps between low- and high-performing students and to reduce the size of these gaps over time. However, as the study shows, the narrowing of an achievement gap does not always occur because of increased performance.

Future research should explore these findings at different grades and ages, at different time points, across subject areas, or by breaking down the results by various student characteristics. For example, patterns of low and high performance may differ for males and females and across subject areas. Future research could use additional data from TIMSS and other international large-scale assessments (e.g., PISA).

In thinking about policy implications, we would argue that education systems that are committed to fostering equity and opportunity and looking to attain technological and economic competitiveness should be concerned about maximizing the learning potential of both their low- and high-performing students. In a 2016 article that appeared in *Educational Leadership*, Celine Coggins argues that you will get a policymaker's attention if you do work that addresses the three main pressures that policymakers face. We think research like this has implications in two of the three areas that she points out: "promoting equity" and "allocating scarce resources" (the third issue that she cites is "addressing accountability issues").

Finally, this study suggests that the effort among industrialized countries to reduce the disparity between low- and high-performing students may also help to reduce income inequality, and vice versa. Future research could further explore the relationship between mathematics achievement gaps and country-level income inequality, including the unexpected negative relationship at grade 8 for non-OECD education systems. Future research could incorporate additional student variables as well as school variables from the TIMSS database, along with additional country-level contextual variables from outside sources.

References

- Anderson, Jenny/Shendruk, Amanda (2019): The best students in the world, charted. Quartz. <https://qz.com/1759506/pisa-2018-results-the-best-and-worst-students-in-the-world/>
- Badescu, Mircea/D'Hombres, Beatrice/Villalba, Ernesto (2011): Returns to education in European countries: Evidence from the European Community Statistics on Income and Living Conditions (EU-SILC). Joint Research Centre, European Commission. Luxembourg: Publications Office of the European Union.
- Barone, Carlo/van de Werfhorst, Hermann G. (2011): Education, cognitive skills and earnings in comparative perspective. In: *International Sociology*, 26, 4, pp. 483-502.
- Coggins, Celine (2016): What it takes to get a policymaker's attention. In: *Educational Leadership*, 73, pp. 28-32.
- Coleman, James S./Campbell, Ernest Q./Hobson, Carol J./McPartland, James/Mood, Alexander M./Weinfeld, Frederick D./York, Robert L. (1966): *Equality of educational opportunity*. Washington, DC. Department of Health, Education, and Welfare.
- Coughlan, Sean (2016): Singapore tops global education rankings. BBC News. <http://www.bbc.com/news/education-38131731>
- Gurney-Read, Josie (2016): Revealed: World pupil rankings in science and maths – TIMSS results in full. The Telegraph. <http://www.telegraph.co.uk/education/2016/11/29/revealed-world-pupil-rankings-science-maths-timss-results/>
- Guyon, Nina/Maurin, Eric/McNally, Sandra (2012): The effect of tracking students by ability into different schools. In: *Journal of Human Resources*, 47, 3, pp. 684-721. <https://doi.org/10.3368/jhr.47.3.684>
- Heyneman, Stephen P./Loxley, William A. (1983): The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. In: *American Journal of Sociology*, 88, 6, pp. 1162-1194.
- LaRoche, Sylvie/Joncas, Marc/Foy, Pierre (2016): Sample Design in TIMSS 2015. In: Martin, Michael O./Mullis, Ina/Hooper, Martin (eds.): *Methods and Procedures*

- in TIMSS 2015, pp. 3.1-3.37. <http://timss.bc.edu/publications/timss/2015-methods/chapter-3.html>
- Lofgren, Emma (2016): Swedish pupils gain ground in global education rankings. The Local.se. <http://www.thelocal.se/20161130/swedish-pupils-gain-ground-in-education-rankings-timss>
- Omigazy, Dana (2016): Kazakh students rank higher than U.S., Germany in Math and Science. The Astana Times. <http://astanatimes.com/2016/12/kazakh-students-rank-higher-than-u-s-germany-in-math-and-science/>
- Peaker, Gilbert Fawcett (1971): The Plowden children four years later. London: National Foundation for Educational Research in England and Wales.
- Plowden, Bridget (1967): Children and their primary schools: A report of the Control Advisory Council for Education (England), Vol. 1. London: Her Majesty's Stationery Office.
- Scheerens, Jaap/Hendriks, Maria A. (2004): Benchmarking the quality of education. In: European Educational Research Journal, 3, 1, pp. 101-399.
- Teng, Amelia (2016): Singapore students top maths, science rankings. Singapore Strait Times. <https://www.straitstimes.com/singapore/education/singapore-students-top-maths-science-rankings>
- UNESCO (2012): International Standard Classification of Education ISCED 2011. Montreal: UNESCO Institute of Statistics. <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf>
- Wright, Helen (ed.) (2019): Estonia tops tables in PISA international education rankings. ERR News. <https://news.err.ee/1009832/estonia-tops-tables-in-pisa-international-education-rankings>

IV. The Management and Use of Digital Data in Education

Section Editors:

Sieglinde Jornitz, DIPF | Leibniz Institute for Research and Information in Education

Laura Engel, George Washington University

The Management and Use of Data in Education and Education Policy: Introductory Remarks

Sieglinde Jornitz¹ and Laura C. Engel²

1. Datafication – trends in education and education policy: an introduction

Education systems around the world share a fundamental aim to improve the overall development, growth, and learning outcomes of young people. Over time, the belief has grown that evidence is key to achieving that essential objective. This rests on a basic idea that evidence will bring optimal results by increasing empirical knowledge about successes and areas for improvement. In some sense, the focus on using scientific knowledge to address educational problems is not new (Noah/Eckstein 1969; National Research Council 2012). What does appear to be novel in recent decades, however, is the growing and at times dogmatic commitment to the role that data play in optimizing the quality of education system performance, and the narrow application of student achievement outcomes as the leading measure of both “quality” and “effectiveness” in many systems.

Along with the mounting datafication of education systems have emerged new data infrastructures, constructed and powered by an increasing reliance on evidence from assessments, as well as a commitment to holding governments, political leaders, and teachers accountable (see, e.g. Williams/Engel 2012). For example, since the 1990s, and linked with new global trends in accountability, more systems around the world are participating in international, regional, and national learning assessments (Kamens/Benavot 2011). At a global level, international large-scale assessments (ILSAs) and the resulting rankings of countries in key academic subjects have grown in significance, often steering “international processes of education policy formation” (Edwards 2012). The largest international education survey to date, the Programme for International Student Assessment (PISA), has provided education policy actors with access to cross-national evidence of educational achievement, helping to fuel the use of these international data in

-
- 1 Sieglinde Jornitz is Senior Researcher at DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main. Email: jornitz@dipf.de
 - 2 Laura Engel is Associate Professor of International Education and International Affairs at the George Washington University (GW), Director of the International Education Program, and co-chair of the GW UNESCO Chair in International Education for Development. Email: lce@gwu.edu

national and sub-national reform efforts (see, e.g., Takayama 2007; Grek 2009; Sellar/Lingard 2014; Engel 2015; Engel/ Frizzell 2015). These assessments not only feed the proliferation of policy advice, but also lead to new data infrastructures and platforms individuals can access for secondary data analysis.

In addition, over the past decade, there is a proliferation of new data tools and applications used to monitor and improve teaching and learning. These new tools are now available to systems, individual schools and other stakeholders including families who are now able to track student learning in real time (Katz 2012).

At a first glance, research on the utility of data in education indicates a universal and international orientation as many digital systems and instruments are distributed world-wide. It might be assumed, for example, that functionalities and work-arounds are the same for every user – regardless of where people live and work. Moreover, the programmed learning systems or systems for education policy and administration are mostly conceptualized for an international market, and only the surfaces of the systems and their respective language differ from country to country. However, the extent to which data are utilized in different systems and by different actors within those systems to improve practice are understudied. There are also key differences in governance and management of data infrastructures across systems.

2. Datafication of education: promises and potential hazards

International discourse on data utilization in education spans many perspectives, ranging from pleas for new and better use of digital education to critiques about its pitfalls. For example, results from the International Computer and Information Literacy Study (ICILS) have been used for appeals for increased investment in infrastructure and teacher training and professional development related to the use of data technologies in education (Watkins/Engel/Hastedt 2016; Eickelmann/ Labusch 2019). At the same time, critiques and cautions continue to mount, for example, regarding data abuse (Lankau 2015; 2019) or the addictive potential of digital tools (Bleckmann/Eckert/Jukschat 2012; Bleckmann/ Jukschat 2015), which is particularly concerning as some critics have fundamentally emphasized the necessity of social interaction for adolescents (Rittelmeyer 2018). Two of the main primary promises and potential hazards are exemplarily outlined in the following.

2.1 The hope for individualized learning

The usage of digital data in educational administration and educational practice is most of all linked to a hope for more individualized actions, because technical systems are better able to process more data at a higher speed. Williamson characterizes such development as “real-time governance of the individual” (Williamson 2016b: 134). At the core of this desire we find the idea of a personalized learning environment. Selwyn points out that “digital technologies are seen to enhance student’s control over the nature and form what they do, as well as where, when and how they do it” (Selwyn 2011: 16). Education technology supports students in self-organizing their learning paths. The instruments and tools shall contribute to a reduction of workload for teachers; they support the monitoring of learners’ progress, the management of teaching materials or other administrative tasks (Selwyn 2011: 17). These tools might be able to overcome the dilemma that there is ultimately never a chance to serve students appropriately, owing to a shortage of time and resources and too little knowledge about an individual student. At school, a teacher often has to divide attention across all students in class. In a classroom setting, students are usually expected to learn according to the set standards, and a teacher’s lessons will be oriented toward the average. Educational administration usually works with statistical data, which assess the individual students’ situation via a mean score without being able to meet the real needs of the children, their parents or their social environment.

In each case, digital data promise a solution to the dilemma of not serving the individual student enough. Digital technology can collect and algorithmically analyze an incredibly large amount of data (Stalder 2017). For educational administration, this does not only concern resource planning but most of all the forecast of risk that can duly be identified. Measures might then be taken even before a child is in real danger of failure. For instruction, it is likewise possible to forecast future scenarios. However, the potential of digital infrastructures is even higher regarding teaching. For example, teaching of an entire classroom might be replaced with individual instruction. The learning software might present each student with the task he or she needs for a particular level, to reach the next one. Such an adaptive type of learning would take personalization to the extreme. A teacher’s role would then mean assisting the student in managing the task.

Personalized learning does not only depend on respective digital systems, but the systems also require a high number of personalized data. It is likely that many such data would have to enter the system in addition to achievement-related data, and correctly and incorrectly solved tasks. To deliver exact profiles, socio-economic data would also be needed, such as background information like number of siblings, parents’ age and profession, and health. Such data can be collected “from *in utero* through to the school

years” (Lupton/Williamson 2017). Hence, this would not only lead to personalized learning by datafication, but to a system of dataveillance (Lupton/Williamson 2017; see for the extension of an educational data science: Williamson 2016a). All available data would be visible and accessible by the state and its related institutions, if there were no legal framework that minimizes or defines clearly the connection to data. In this regard, the US and Europe have chosen to follow different pathways of data privacy protection.

The positive momentum of support is thus changed into a negative moment of surveillance, corresponding not only to a violation but rather a relinquishment of a democratic state which protects its citizens (and in the case of schools: under-age citizens). The implementation of such learning systems also reveals another aspect: An existence of learning theories and development scenarios for school subjects and thematic areas is assumed that can be modelled by computer programs. Yet, this is not the case (for Germany: Jornitz 2018). In most cases, education technologies model and process school curriculum topics in rather simple ways. Students are instead trained by multiple-choice formats rather than being educated in critical reflection and understanding of reasoning (Rittelmeyer 2018). In this regard “learning takes the form of pre-packed curriculum content and teachers are positioned as providers and supporters of students’ navigation through set activities” (Selwyn 2011: 107f.; Selwyn 2016). Moreover, digital instruments seldom serve to personalize and individualize in an authentic manner. Students are grouped according to their level of knowledge (low, medium, high) and provided with respective tasks. It is an open question whether this procedure lets students remain at a certain level of knowledge or if such a tool assists them in reaching a higher level. Studies indicate a limitation of data usage and the increase of competition and incessant insecurities in classrooms (Neumann 2019). Hence, personalized systems would need to strike a balance that enables support, self-organization, and help while avoiding surveillance and submission.

2.2 The hope for greater equity and equality in learning opportunities

The analysis of personalized data is moreover linked to the hope of overcoming inequality and thus leading to a fairer learning environment in school. From this perspective, digital infrastructures are adapted to the individual and deficits are thus recognized and outbalanced. Digital technology consistently processes data in the same manner, while human actions are deficient. In this regard, digital instruments might be better than teachers in providing students with materials and resources that match their skills; “learners can enjoy access via the internet to a more diverse range of

learning opportunities” (Selwyn 2012: 15). The hope is focused on a decreasing school drop-out-rate and an increasing percentage of adolescents entering employment. Although schools might fulfil their remit to enable all children a strong start regardless of their parental status, digital data would reveal inequalities sooner, and pave the way for timely actions, thus acting as a predictive analytics technique (Williamson 2016b).

Such scenarios are challenged by research demonstrating that inequality is not removed but even generated. Ben Williamson analyzed a wide range of studies and comes to the conclusion that “instruments are bearers of values and interpretations of the social world that are materialized and operationalized by particular and concrete techniques and tools, and that as a result have the capacity to partly structure policies, determine how actors behave and privilege certain representation of problems to be addressed” (Williamson 2016b: 125).

Data allocate students according to categories that are not as indisputable as they might appear at a first glance. In many cases, such data are used as an equivalent for a certain quality, without being that quality themselves (Mau 2018). A standard example taken from the PISA background questionnaires relates to a question that asks for the number of books in each household in order to assess the socio-economic background, respectively the academic status of students’ parents, assuming that they have only a vague idea regarding their parents’ salary or their academic status. The number of books is taken as an approximate value which is principally considered a good, robust indicator, but can occasionally lead to an incorrect assessment of cases.

Such wrong appraisals cannot be avoided even if data infrastructures are expanded. Inequality will remain because individual cases are not recognized by a system. Such a system can be characterized as “[p]redictive and prescriptive” (Williamson 2016b: 136). Given that these data are only an approximation, they are always also tied to limitations and reductions that need to be recognized (Selwyn 2016: 64). Only the perception of the systems’ limitations will enable a realistic assessment of their scope, because “this reductionism [is] also apparent in what was *not* being captured and recorded in the schools’ digital data” (Selwyn 2016: 64; emphasize in original). It is therefore necessary to train professionals who handle the systems once they have been implemented. If there is no intention to abstain from using such systems and platforms, it seems essential to argue for an evidence-informed rather than an allegedly evidence-based action fed by data, and a rationally substantiated decision against the systems should always be possible. After all, “[...] technologies are subjected continually to a series of complex interactions and negotiations with the social, economic, political and cultural contexts into which they emerge” (Selwyn 2011: 41).

3. Approaching the research

Digital instruments are not only an object of research but challenge scientific methods and methodologies, including qualitative and quantitative aspects. Researchers who are working with qualitative methods need a basic technical understanding to approach digital education. It is possible to conduct interviews with software engineers who design education technological products – likewise, interviews can be run and analyzed with users of such software in educational administrative and education practical contexts. By doing so, we gain insights into the handling and routines in operating particular data systems that are invisible in daily lives. Such qualitative assessments can reveal which decisions have to be made for dealing with the systems. They raise doubts regarding the alleged neutrality or objectivity of the data (Selwyn 2016; Williamson 2016b). Such research is oriented toward operating practices. It concentrates on understanding the handling of such data-generating systems to elicit their logic and structure, thus revealing how a human agent turns to the data and interprets them. The research can shed light on the blind spots in process and decision chains within educational contexts – blind spots that occur in the interaction of humans with machines.

This orientation of research initially does not affect technical operations but enables an appraisal of the agents' routines which are, in a certain respect, limited by processing and output logic of the software. This technological logic is first of all meant to be analyzed and disclosed from a social scientific perspective. Not only is a technological understanding required but an interdisciplinary collaboration of social scientists with technical experts seems necessary in the field (Decuyper 2019). Other than collaborations with computer scientists in general, however, this does not concern programming for social scientist purposes. Computer programmers together with social scientists rather develop methods to visualize software operation processes.

An example from school administration may serve to illustrate this. Across the world, school administrators are increasingly using data sources not only to calculate their budgets, equipment and human resource needs but also to identify students who are at risk of failing school and dropping out of the education system. Many diverse data sources can be processed in such data systems. Besides finance and human resource data, general assessments of student achievement (national and international) are entered into the systems, as well as socio-economic data on the region, individual details concerning students and many more. Ultimately, there is no limit or programming boundary to the quantity and type of processed datasets. One might even fear that data sets which are available to an administrative body

and are not limited by data protection regulations, might have a tendency to induce a desire for entering such data into processing (cf. Williamson 2017).

Working with the systems requires that analyzed data are visualized via a so-called dashboard. Without that kind of reduced and visualized presentation, users would be challenged to run their own calculations. The task is thus referred to the algorithm underlying the data mining systems (Williamson 2016b). Therefore, it is not the user of the software system who determines the algorithm by which data are processed and calculated. Educational administrators can ultimately only work via the dashboard presentations. In most cases, attention is guided via color schemes ranging from green to yellow to red whose meanings are known from other areas of social life. Decision-making processes are thus pre-shaped at a base level.

Scientists need to learn how to disclose and understand these algorithmic structures and data mining processes if they wish to take a critical distance to the software. So far, scientists have attempted to approach the structures via their usage as such, and via an analysis of paths and screen presentations (Decuyper 2019); however, the process is arduous and limited. Actually, this means that science itself would need a license for a respective software product, feed data into it and then put it to the test. In a further step, one might even be able to read a technical code in a technical analysis. Teams of social scientists who are interested in such critical studies need to develop new methods of disclosing which datasets interact and how.

Digital instruments also reveal new opportunities for quantitative science. Currently, a focus is on an increased generation of data from practice, for example via an analysis of log file data on click rates (Naumann 2008) to better understand the underlying learning process. Students who are working on their computers thus become technically transparent in their course of action. In this case, existing methods are also expanded and processing methods which require a technical understanding to analyze the data need to be developed. Scientists who work with quantitative methods are expanding their set of methods by concentrating on the data that are always automatically generated by the system, and they try to make these data usable in the interest of educational research.

On this basis, scientists working with quantitative methods collaborate with computer scientists to develop instruments for instructional processes. Digital structures do not only allow for an increased generation of data but also for faster data processing. It is thus possible to almost deliver results in real time. At this point, research is being conducted in programming software that enables teachers to assess the classes' respectively students' individual achievement and thus adapt didactic decisions more quickly.

In a broad sense, such research can be allocated to learning analytics. The focus lies on assisting teachers and students by usage of data in real time. The Society for Learning Analytics Research (SoLAR) defines learning analytics

as “[...] the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Ferguson 2012).

An optimization of learning processes is thus targeted, as well as an influence of instruction via the data, by means of immediate feedback to answers that are entered into the system. In this sense, a machine is superior to a teacher who can only direct his or her attention to selected students in the classroom. A machine can deliver an immediate response to each student and inform the student about the correctness or incorrectness of an entry. For this purpose, many datasets are used in Learning Analytics platforms. First of all, this concerns student log data but possibly also further data on their learning progress. Such data are used to forecast learning outcomes or simply calculate the next task. Learning Analytics can be viewed as a support system for teachers, assisting them in the improved and sometimes quicker identification of learning difficulties and offering suggestions for further didactic procedures.

4. Cross-national trends in data utilization in education

Research communities in different countries also reflect on how various data are regarded and utilized in and across education systems worldwide to inform policy and practice. In the US, for example, a strong focus on the uses of data in education policy can be found. The American Educational Research Association (AERA), the leading national research society with more than 25,000 members, hosts a special interest group on “Data-driven decision making in education” (AERA-SIG 179). Multiple reports have focused on the role of data in national policy-making trends (National Research Council 2012; Singer/Braun/Chudowsky 2018). In contrast, there is no equivalent national research network in the German Educational Research Association (*Deutsche Gesellschaft für Erziehungswissenschaft – DGfE*).

The most influential publications on the importance and challenges in society and education through software and digital instruments have been written by Australian, British and US-based researchers. Germany and many other European countries are benefitting from this work and have started to contribute to the discourse. This difference may signal and suggest that there is varied awareness of, and priority assigned to, the utilization of data in education policy-making. In part this is because data-driven policy-making in education has followed divergent pathways (see comparison of standards-based reforms in Wallner et al. 2020).

The case of Germany is interesting because of the much slower technological development with regard to the usage of digital data in education. In

Germany, the discourse on digital data in education started very late in comparison to the US, UK or Australia and is still dominated by a narrative of lagging behind or of losing connection to an international process and development. Most German schools lack the digital or technological infrastructure needed to use and teach with learning software tools. At the same time, Germany has a teacher education system that centers on the subject taught and teachers have a high level of confidence in being responsible for the teaching process in the classroom. Therefore, teachers have to be convinced that teaching with digital resources or software instruments offers an asset.

Compared to education practice, education governance – as outlined before – established a digital data infrastructure for its purpose. The German education monitoring strategy marks the starting point of regularly delivering datasets of school education for education policy. In Germany, data have been regularly used for education governance since the late 1990s, which became evident for everybody since the so-called “PISA shock” in 2000. For many years, Germany did not participate in large-scale assessment studies of the International Association for the Evaluation of Educational Achievement (IEA). In this regard, the mathematics and science study TIMSS for students in 8th grade of the IEA marked a new beginning. This participation led to an agreement among the education ministers of the 16 Länder or states – called *Konstanzer Beschluss* (“Resolution of Constance”).³ This document was fundamental for the establishment of an educational monitoring infrastructure that installed a system of gathering, analyzing and distribution of data for the entire education system in Germany. In this line, the education ministers decided to regularly conduct achievement tests for students at a certain age (called “VERA” – *Vergleichsarbeiten* (comparative tests)), that are compulsory for grade 3 and 8 and focus on the school subjects of Mathematics, German and the first foreign language (English or French). Yet, it took another approximately ten years until the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic (in German: *Kultusministerkonferenz* – KMK) published a strategy paper that merged several instruments into one monitoring education system.⁴ This strategy consists of a bundle of instruments that produce data for the policy context and is centered on the political target of measuring and ensuring quality in education. For the German national education report (in

3 See: Kultusministerkonferenz (1997): Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland – Konstanzer Beschluss – (Beschluss der Kultusministerkonferenz vom 24.10.1997). Available at https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1997/1997_10_24-Konstanzer-Beschluss.pdf

4 KMK (2016): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Berlin: KMK. Available at https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf

German: *Bildungsbericht*), data are taken from instruments like international large-scale assessments (see chapter 3 in this volume), examinations of learning results and competency levels in certain subjects and selected school years as well as statistical data. This report is published every two years with different topics in focus, like cultural and aesthetic education in 2012, inclusive education in 2014, and most recently on results and effects of education in 2018.⁵

Meanwhile, Germany and its citizens have thus become used to the publication and discussion of educational data. In comparison to other countries and especially in comparison to the US, this is a fairly new development that started twenty years ago. Still, there is resistance from politicians, practitioners and parents. The data are therefore used for information purposes only rather than being a source for decision-making in terms of accountability. The agreed societal consensus takes education data as an objective source, but the decisions that are linked to them are discussed and well-justified.

In this line, the German development has a counterpart at the European level that uses education data to constitute a European Education Area. Because the European Union has no political mandate for regulating or governing the education systems of the member states, the Union uses data as a political instrument. For coordination purposes, the member states have agreed on certain targets and benchmarks in the policy sector of education. Such benchmarks encompass the increase of participation in lifelong learning, student achievement results and tertiary education attainment as well as the decrease in student drop-out rates.

Given the fact that the European Commission as the executive authority of the European Union lacks political instruments to control or interfere, the Commission has focused on stimulating action by visualizing data in education. Since 2011, the European Commission has annually collected educational data from its member states and published the *Education and Training Monitor*.⁶ The latest volume was published in 2019 and it shows impressively what data can do.

The annual Monitor Report (European Commission 2019) uses data on two levels. The first level establishes a European Education Area by visualizing a data-set for all European member states (cf. Figure 1 and 2).

5 See for education reports: <https://www.kmk.org/themen/bildungsberichterstattung.html> or short versions in English: https://www.bildungsbericht.de/en/the-national-report-on-education/education-in-germany?set_language=en

6 Website of the European Commission: https://ec.europa.eu/education/policy/strategic-framework/et-monitor_en

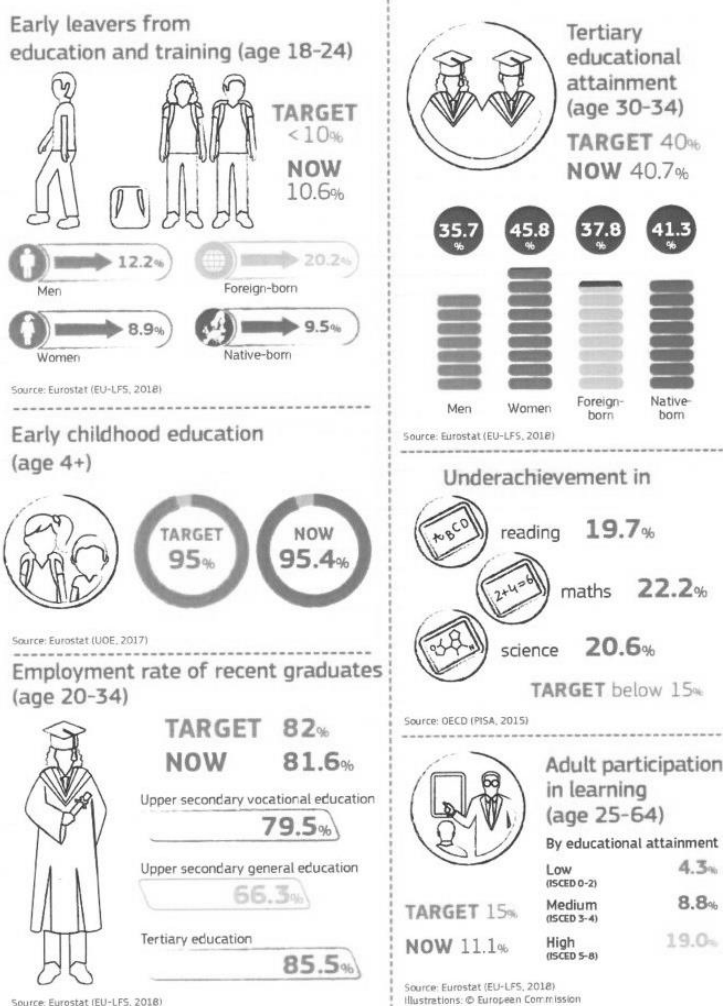
Figure 1. EC: Education and Training Monitor 2019

EU targets for 2020 in education and training

		Current	Target
Headline target	1		
	Early leavers from education and training The share of 18 to 24 year-olds having attained ISCED level 0-2 and not receiving any formal or non-formal education or training in the four weeks preceding the survey.	10.6%	Below 10%
Other targets	2		
	Tertiary educational attainment The share of 30 to 34 year-olds having successfully completed ISCED level 5-8.	40.7%	At least 40%
	3		
	Early childhood education The share of children aged 4 to the age of compulsory primary education who are participating in education.	95.4%	95%
	4		
	Underachievement in reading, maths and science The share of 15 year-olds failing to reach level 2 in the OECD's PISA for reading, mathematics and science.	Reading: 19.7% Maths: 22.2% Science: 20.6%	15%
	5		
	Employment rate of recent graduates The share of employed 20 to 34 year-olds having successfully completed ISCED 3-8 one to three years preceding the survey and who are no longer in education or training.	81.6%	82%
	6		
	Adult participation in learning The share of 25 to 64 year-olds who received formal or non-formal education or training in the four weeks preceding the survey.	11.1%	15%

Source: Eurostat (EU-LFS 2018 for 1, 2, 5 and 6; UOE 2017 for 3) & OECD (PISA 2015 for 4). Note: ISCED 0 = early childhood; ISCED 1 = primary education; 2 = lower secondary education; 3 = upper secondary education; 4 = post-secondary non-tertiary education; 5 = short-cycle tertiary education; 6 = bachelor's or equivalent level; 7 = master's or equivalent level; 8 = doctoral or equivalent level.

Figure 2. EU targets for 2020 in Education



Source: EC: Education and Training Monitor 2019: 4

Europe appears as a whole, as one region where each member state has the same target for its education system. A European Education Area is created by offering these data as a summation of each member state. But at the same time, these data do not inform about how they were calculated and in which

appropriate rate they were positioned. The aggregated data cannot show how differently member states deal with the targets and from which level they start. By visualizing one type of data for the European Union, all differences are made invisible. For instance, one country might still be far from reaching the goals, while another is doing better.

Therefore, a second level is necessary. The Monitor Report of the European Commission also presents data on several education topics for each member state, for example statistical data on teachers and teaching, participation in early childhood education or digitally equipped schools (EC 2019). By looking closer at the figures, two kinds of presentation can be identified: one that lists the member states and its measures in alphabetical order and another one that ranks them by increasing measures. These two kinds of data visualization put the member states in direct competition to each other. Even when they are listed in alphabetical – i.e. the most neutral – order, they are placed in comparison to their adjacent partners.

Even though the European Commission has no mandate for the member states' political decision-making in education, data are being used to push each of the member states in a similar direction towards the development of their education systems. By gathering the data at hand and compiling them in an appealing way, the European Education and Training Monitor becomes an instrument and a source for a data-sensitive education policy. The European Union takes its chance to become a powerful player in the policy field of education through the use of data (Grek 2009; Lawn/Grek 2012).

In the US, the past three decades of educational reform seemingly have been dominated by a commitment to data-driven decision-making. The rationale for this commitment is aptly stated by Datnow, Park, and Wohlstetter (2007): "Using data to improve decision making is a promising systemic reform strategy" (p. 10). There are a range of uses for data, but overall proponents of using data to guide decision-making see it as essential to allowing "school systems to learn more about their school, pinpoint successes and challenges, identify areas of improvement, and help evaluate the effectiveness of programs and practices" (Datnow/Park/Wohlstetter 2007: 10).

Data have been regularly used in US educational governance since the 1980s, though more on state-level through national assessments rather than international ones. Internationally, the US played a key role in the establishment of different international large-scale assessments, and has remained a frequent participant in international assessments since their inception. However, it has generally drawn on data from these international assessments less frequently than its own longstanding national assessments, like the National Assessment of Educational Progress. Take PISA for example. While PISA has had a more influential role in systems like Germany, it has generally received less policy attention and influence in the

US (Niemann/Hartong/Martens 2018). While all states do participate in PISA as part of the national sample, a small number of states have elected to pay to participate in PISA to receive state results (Engel/Frizzell 2015; Engel/Rutkowski 2018). In addition, individual schools opt to participate in PISA for Schools (Rutkowski 2015).

The datafication of US education policy appears more strongly linked to the intensification of discourse around an educational crisis and need for reform, initiated by the 1983 report “A Nation at Risk”. This reform underscored the need for common standards across states and the use of tests to hold schools, districts, and states accountable (Koretz 2008; Engel/Olden 2012). With the passage of the “No Child Left Behind” (NCLB) legislation in 2002, standards-based assessments were made high stakes. The federal legislation effectively mandated all states to use a standards-referenced assessment in grades 3-8 and one secondary grade (Koretz 2008). NCLB developed an annual yearly progress system, whereby schools and states accepting federal funds were required to report results in relation to performance standards and show that all students were performing at a proficient level.

The state and federally-driven reform agenda has continued to underscore the use of standards-based assessments as mechanisms for school improvement, whereby school effectiveness is strongly linked to indicators of student performance. For instance, the 2009 announced Race to the Top’s \$360 million Assessment Program, which linked access to federal funding with individual state’s use of standards-based assessments, encouraged and paved the way for additional assessments aligned with the Common Core State Standard Initiative. States choose which assessments they wish to use, including Smarter Balanced Assessment Consortium, the Partnership for Assessment of Readiness for College and Careers (PARCC), the Scholastic Assessment Test (SAT) or the American College Testing (ACT), and alternatives that states designed and/or purchased on their own (Wallner et al. 2020).

These developments mark the now well-established trend to link results of student achievement data to (1) forms of accountability of schools, administrators, and/or teachers; (2) improve teaching and learning at school and classroom levels; and (3) overall system reform (Williams/Engel 2012). Across the US, local districts and states also link standardized testing to teacher evaluation, along with a range of data collected from classroom observations, student surveys, and other measures. For example, the District of Columbia Public Schools have more robust evidence-based approaches to teacher evaluation, drawing on multiple measures, including classroom observations, student achievement data, student survey data on classroom culture, and teacher contributions to school culture (DCPS n.d.).

Additionally, there has been a rapid growth in the range of new educational technologies and digital tools, like PowerSchool, Seesaw, and Google Classroom, now regularly employed to improve classroom instruction, create student portfolios, and track student learning against learning objectives. Some of these different digital tools and applications are designed to allow families to monitor student progress and track learning in real time (Katz 2012). While increasingly prevalent, access to and use of these technologies both by schools and families varies greatly (Katz 2012). Although these and other developments are beyond the scope of this introduction, schools in the US produce these data by experimenting in which ways they might contribute to the teaching and learning process. Once developed, they are ready to be used and connected to various educational platforms and infrastructures. In many ways, it is an open and in some cases unregulated space for education policy and practice, driven by a range of non-governmental stakeholders.

5. Advancing understandings of data uses in education: comparative perspectives

Given the divergent trends related to data utilization in educational governance and practice across systems, an international exchange about the national contexts of using data for the education processes seems needed. It bears in mind the opportunity to broaden national views and to see how differently data are analyzed and interpreted. The usefulness of digital data varies from skeptical to euphoric appraisals, while countries (or supranational institutions) often act differently in supporting digital infrastructures for and in schools.

In this volume, contributors discuss the possibilities that digital data, data infrastructures and data flows offer for an improvement of education settings, but also pay attention to problematic aspects around the growing importance and also the increasing amounts of data to be handled, organized and used within the education system. By giving room to transatlantic perspectives, the contributions not only broaden the view on the worldwide trend towards digitalizing education, but also raise questions that are important for further – comparative and cooperative – research.

The contributions focus on (1) different aspects related to the datafication of educational governance, including the key agencies and the dynamics involved in the production and utilization of big data; and (2) the possibilities of data use in school administrative and teaching practice. In the first contribution, *Sigrid Hartong* presents some insights into a comparative

project carried out in Germany and the US with regard to different uses of data in school administration agencies. She outlines how monitoring is being made by the technical systems and what kind of sense-making is formed by these systems.

In the second contribution, *Steven Lewis* looks at the OECD's instrument "PISA for Schools". He demonstrates how the OECD as *the* leading agency of international education invented a tool that will serve individual schools for gathering data. Closely linked to the successful large-scale assessment study on students' knowledge, the schools receive an instrument that is able to align local education success (or failure) to an international acknowledged standard. By inspecting the reports and recommendations "PISA for Schools" offer, *Steven Lewis* unfolds the schematized results.

The next two contributions indicate different possibilities of data use in school administrative and teaching practice. *Bernard Veldkamp, Kim Schildkamp, Merel Keijsers, Adrie Visscher and Ton de Jong* carried out a study in the Netherlands that broadens the view for data usage in practice. The insights are not only useful for the country the study originates from, but refer to an international audience. The authors explored the potential of big data in formal education by interviewing Dutch stakeholders, revealing purposes, usage and challenges linked to big data.

In the fourth contribution, *Elmar Souvignier, Birgit Schütze, Karin Hebbecke and Natalie Förster* present a study that is deeply rooted in the US-American development of instruments for monitoring learning progress. The research group focuses on the web-based system for measuring student progress called "quop". Developed for German schools, the authors give an example how digital instruments can fulfill requirements of technical adequacy and simplicity. With such regularly short term testings they offer teachers a tool that is suitable for classroom practice. They share their insights with regard to research on its technical adequacy and teaching effects.

Together, these contributions raise important issues about the different spheres of data usage in education and help broaden the scope for an ongoing debate about the utility of data in research and practice, as well as demonstrating the power and value of international exchange in better understanding the ways in which data are used in different systems.

References

- Bleckmann, Paula/Eckert, Judith/Jukschat, Nadine (2012): Futile search for a better life? Two biographical case studies on women with depression and video game dependency. In: *Advances in Dual Diagnosis*, 5, 3, pp. 137-146.

- Bleckmann, Paula/Jukschat, Nadine (2015): The integrated model of (dys)functionality: Reconstructing patterns of gaming as self-medication in biographical interviews with video game addicts. In: FQS – Forum qualitative Sozialforschung, 16, 3, Art. 8.
- Datnow, Amanda/Park, Vicki/Wohlstetter, Priscilla (2007): Achieving with Data: How high-performing school systems use data to improve instruction for elementary students. Report. <http://people.uncw.edu/kozloffm/AchievingWithData.pdf>
- Decuypere, Mathias (2019): Researching educational apps: ecologies, technologies, subjectivities and learning regimes. In: Learning, Media and Technology, 44, 4, pp. 414-429. doi: 10.1080/17439884.2019.1667824
- District of Columbia Public Schools (n.d.): IMPACT: The DCPS Evaluation and Feedback System for School-Based Personnel. <https://dcps.dc.gov/page/impact-dcps-evaluation-and-feedback-system-school-based-personnel>
- Edwards, Brent D. (2012): Researching international processes of education policy formation: Conceptual and methodological considerations. In: Research in Comparative and International Education, 7, 2, pp. 127-145.
- Eickelmann, Birgit/Labusch, Amelie (2019): ICILS 2018 #Deutschland auf einen Blick. International Computer and Information Literacy Study. Münster: Waxmann.
- Engel, Laura C. (2015): Steering the national: Exploring the education policy uses of PISA in Spain. In: European Education, 47, 2, pp. 100-116.
- Engel, Laura C./Frizzell, Matthew O. (2015): Competitive comparison and PISA bragging rights: Sub-national uses of the OECD's PISA in Canada and the US. In: Discourse: Studies in the Cultural Politics of Education, 36, 5, pp. 665-682.
- Engel, Laura C./Olden, Kate (2012): One size fits all: Globalization and standardization of US education. In: Shaklee, Beverly/Bailey, Supriya (eds.): Internationalizing teacher education. Lanham, MD: Rowman and Littlefield, pp. 77-92.
- Engel, Laura C./Rutkowski, David (2018): Pay to play: What does PISA participation cost in the U.S.? In: Discourse: Studies in the Cultural Politics of Education.
- European Commission (2019): Education and Training Monitor 2019. Luxemburg: European Union. https://ec.europa.eu/education/resources-and-tools/document-library/education-and-training-monitor-eu-analysis-volume-1-2019_en
- Ferguson, Rebecca (2012): Learning analytics: drivers, developments and challenges. In: International Journal of Technology Enhanced Learning, 4, 5/6, pp. 304-317. <http://dx.doi.org/doi: 10.1504/IJTEL.2012.051816>
- Fraillon, Julian/Ainley, John/Schulz, Wolfram/Friedman, Tim/Duckworth, Daniel (2019): Preparing for Life in a Digital World. IEA International Computer and Information Literacy Study 2018. International Report. Amsterdam: IEA.
- Grek, Sotiria (2009): Governing by numbers: The PISA 'effect' in Europe. In: Journal of Education Policy, 24, 1, pp. 23-37.
- Kamens, David H./Benavot, Aaron (2011): National, regional and international learning assessments: Trends among developing countries, 1960-2009. In: Globalisation, Societies and Education, 2, pp. 285-300.
- Katz, Cindi (2012): Just managing: American middle-class parenthood in insecure times. In: Heiman, Rachel/Freeman, Carla/Liechty, Mark (eds). The global

- middle classes: Theorizing through ethnography. Santa Fe, New Mexico: School of Advanced Research Press, pp. 169-188.
- Koretz, Daniel (2008): *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Kultusministerkonferenz (2016): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Berlin.
https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf
- Lankau, Ralf (2015): Das Lernen verlernen? Digitale Medien und Unterricht. In: *Pädagogische Korrespondenz*, 52, pp. 42-58. <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-148606>
- Lankau, Ralf (2019): Vom Unterrichten zum Bildungscontrolling. Über die Blindheit gegenüber den Zielen der 'Digitalisierung' von Schule und Unterricht. In: Bleckmann, Paula/Lankau, Ralf (eds.): *Digitale Medien und Unterricht. Eine Kontroverse*. Weinheim/Basel: Beltz, pp. 54-66.
- Lawn, Martin/Grek, Sotiria (2012): *Europeanizing education: governing a new policy space*. Oxford: Symposium Books.
- Lupton, Deborah/Williamson, Ben (2017): The datafied child. The dataveillance of children and implications for their rights. In: *New Media & Society*, 19, 5, pp. 780-794.
- Mau, Steffen (2018): *Das metrische Wir. Über die Quantifizierung des Sozialen*. Berlin: Suhrkamp.
- Manolev, Jamie/Sullivan, Anna/Slee, Roger (2018): The datafication of discipline. *ClassDojo, surveillance and a performative classroom culture*. In: *Learning, Media and Technology*, 44, 1, pp. 36-51. doi: 10.1080/17439884.2018.1558237
- National Research Council (2012): *Using Science as Evidence in Public Policy*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/13460>
- Naumann, Johannes (2008): Log file analysis in hypertext research: An overview, a meta-analysis, and some suggestions for future research. In: Cañas, José J. (ed.): *Workshop on Cognition and the Web: Information Processing, Comprehension and Learning*. Granada: University of Granada, pp. 53-57.
- Neumann, Eszter (2019): Setting by numbers: datafication processes and ability grouping in an English secondary school. In: *Journal of Education Policy*, pp.1-23.
- Niemann, Dennis/Hartong, Sigrid/Martens, Kerstin (2018): Observing local dynamics of ILSA projections in federal systems: a comparison between Germany and the United States. In: *Globalisation, Societies and Education*, 16, 5, pp. 596-608. doi: 10.1080/14767724.2018.1531237
- Noah, Harold J./Eckstein, Max A. (1969): *Toward a science of comparative education*. New York: The Macmillan Company.
- Rittelmeyer, Christian (2018): *Digitale Bildung. Ein Widerspruch*. Oberhausen: Athena.
- Rutkowski, David (2015): The OECD and the local: PISA-based test for schools in the USA. In: *Discourse: Studies in the Cultural Politics of Education*, 36, 5, pp. 683-699. doi: 10.1080/01596306.2014.943157
- Sellar, Sam/Lingard, Bob (2014): The OECD and the expansion of PISA: New global modes of governance in education. In: *British Educational Research Journal*, 40, 6, pp. 917-936.

- Selwyn, Neil (2011): *Schools and Schooling in the Digital Age. A critical analysis.* London/New York: Routledge.
- Selwyn, Neil (2016a): *Is Technology Good for Education?* Cambridge: Polity Press.
- Selwyn, Neil (2016b): 'There's so much data': Exploring the realities of data-based school governance. In: *European Educational Research Journal*, 15, 1, pp. 54-68. doi: 10.1177/1474904115602909
- Singer, Judith, D./Braun, Henry I./Chudowsky, Naomi (2018): *International education assessments: Cautions, conundrums, and common sense.* Washington, DC: National Academy of Education.
- Spring, Joel (2018): *The American School. From the Puritans to the Trump Era.* 10th edition. New York/London: Routledge.
- Stalder, Felix (2017): *Kultur der Digitalität.* 3rd Labuschedition. Berlin: Suhrkamp.
- Takayama, Keita (2007): A Nation at Risk crosses the Pacific: Transnational borrowing of the U.S. crisis discourse in the debate on education reform in Japan. In: *Comparative Education Review*, 51, 4, pp. 423-446.
- Urban, Wayne J./Wagoner, Jr., Jennings L./Gaither, Milton (2019): *American Education. A History.* 6th edition. London/New York: Routledge.
- Wallner, Jennifer/Savage, Glenn/Hartong, Sigrid/Engel, Laura C. (2020): Laboratories, co-producers and venues: Illuminating the varying roles of subnational governments in the formation of standards-based assessments in federations. In: *Comparative Education Review*, 64, 4.
- Watkins, Ryan/Engel, Laura C./Hastedt, Dirk (2015): Is the "net generation" ready for digital citizenship? Perspectives from the IEA International Computer and Information Literacy Study 2013. IEA's Policy Brief Series, No. 6. Amsterdam: IEA.
- Williams, James/Engel, Laura C. (2012): How do other countries evaluate teachers? In: *Phi Delta Kappan*, 94, 4, pp. 53-57.
- Williamson, Ben (2016a): Coding the biodigital child: the biopolitics and pedagogic strategies of educational data science. In: *Pedagogy, Culture & Society*, 24, 3, pp. 401-416. doi: 10.1080/14681366.2016.1175499
- Williamson, Ben (2016b): Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. In: *Journal of Education Policy*, 31, 2, pp. 123-141. doi: 10.1080/02680939.2015.1035758
- Williamson, Ben (2016c): Political computational thinking: policy networks, digital governance and 'learning to code'. In: *Critical Policy Studies*, 10, 1, pp. 39-58. doi: 10.1080/19460171.2015.1052003
- Williamson, Ben (2017): *Big Data in Education. The digital future of learning, policy and practice.* London: Sage.

Digital Education Governance and the Productive Relationalities of School Monitoring Infrastructures

Sigrid Hartong¹

1. Introduction: the rise of digital education governance and the transformation of school monitoring infrastructures

Even though the use of data for education governance itself is nothing new, recent years have marked a new era as the increasingly digital and automated formation, recoding, storage, manipulation and distribution of data has become an essential feature of government (Houben/Priegl 2018; Selwyn 2014: 1). As Williamson (2017: 4) documented, “[s]oftware and digital data are becoming integral to the ways in which educational institutions are managed, how educators’ practices are performed, how educational policies are made, how teaching and learning are experienced, and how educational research is conducted.”

In fact, to date, much research on the datafication and digitalization of education has focused on how to effectively produce, implement and use data or software, while often considering digital technologies as neutral or simply technical. In recent years, however, a growing body of scholars has responded to this purely instrumental approach by calling for more *critical data studies* (for overview see Iliadis/Russo 2016) which, instead of understanding data as given, focus on their capabilities and power (e.g. West 2017). Such research explicitly raises questions “[...] about the [normative, *added S.H.*] nature of data, how they are being produced, organized, analyzed and employed, and how best to make sense of them and the work they do” (Kitchin/Lauriault 2014: 1). Contributing to this approach while referring more specifically to the governance of education, different scholars have introduced the term *digital education governance* to capture this growing insertion of “[...] digital technologies, software packages and their underlying standards, code and algorithmic procedures” (Williamson 2015: 1) into the political, administrative and practical spheres of education and the resultant impact on the conduct of multiple actors (see also Hartong 2016, 2018; Landri 2018).

The research project *Data Infrastructures and the Digitalization of Education Policy – A Comparison between Germany and the United States*

1 Sigrid Hartong is Professor for Sociology at the Faculty of Humanities and Social Sciences at the Helmut-Schmidt-University in Hamburg. Email: hartongs@hsu-hh.de

(funded by the *German Research Foundation/DFG* 2017-2020 and situated at the Helmut-Schmidt-University in Hamburg, Germany) contributes to this important field of digital education governance studies by focusing on the ongoing transformation of school monitoring infrastructures, particularly in state-level school administration.² It is thus argued that monitoring infrastructures have always represented a central dimension of education governance, because they produce powerful representations of schools, teachers and students which administrators then use for *real acting* to guide and legitimize governmental decision-making (including high stakes decisions such as the opening and closure of schools and resource distribution) (West 2017). Over the past decade, school monitoring has become deeply affected by increasing digitalization and automatization, which has not only significantly expanded the amount of (available or aspirational) data, but also continuously accelerated (automated) data production and processing. Simultaneously, digital data generated through the application of personalized classroom technologies increasingly feed state agencies' monitoring tools, fostering a direct link between teaching/learning activities and governmental action (Hartong 2019; Williamson 2017).

While there is a steadily increasing body of research on the (demanded) transformation of monitoring systems in state education agencies across different countries (see for example González-Sancho/Vincent-Lancrin 2016; Dederling 2015; Conaway et al. 2015), a critical understanding of the capacities and powers of such digital monitoring infrastructures has remained surprisingly underdeveloped. Responding to this pressing need, the *Data Infrastructures* project represents an earnest attempt to make visible as least part of the actual data infrastructures and flows which practically enact the *doing* of monitoring in state³-level education agencies.⁴

2 The empirical insights presented in this chapter partly build on interviews which have been conducted by project member Annina Förschler.

3 We use the term 'state' here to describe subnational units of educational authority, as commonly used in the US. It seems important to note that such subnational authorities are usually named *Länder* in Germany. Yet, we use the US term here for the purpose of alignment between the cases.

4 Methodologically, the analysis builds on materials such as organization charts, policy papers, documentation on the development and usage of data instruments, online data dashboards, as well as semi-structured interviews with state agency experts (including school supervising agencies, institutes of quality assurance as well as IT-institutions responsible for processing monitoring data) responsible for different data tasks (in four different states, two in the US and two in Germany).

2. Disentangling monitoring infrastructures using a relational approach

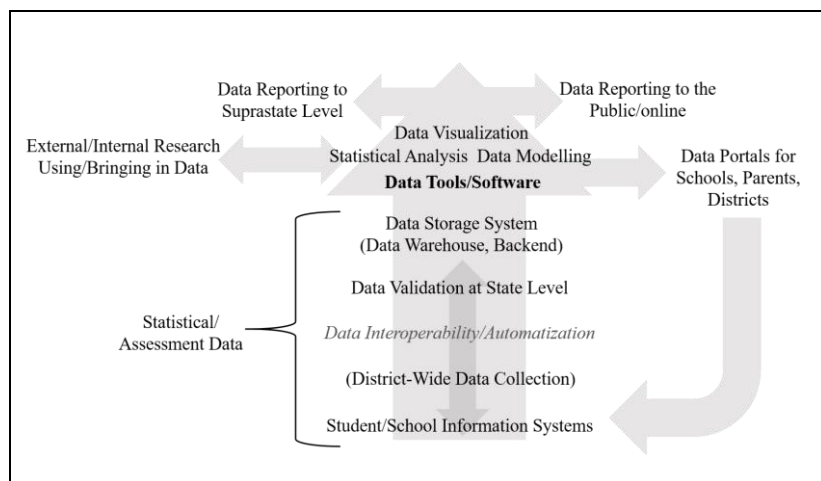
The research project seeks to disentangle monitoring infrastructures in state-level education agencies by asking *which* particular representations of schools, teachers and students become fabricated as data, and *how*. We thus adopt a relational approach as is increasingly used in critical data studies both within and beyond the field of education (e.g. Kitchin/Lauriault 2014; Landri 2018; Williamson 2017; Sellar 2017). In general, a relational approach understands phenomena such as digital education governance as complex, constantly moving, techno-social entanglements or infrastructures of objects and subjects that become “[...] assembled around [...] data and around its socio-technical de- and recontextualization practices” (Hartong 2018: 135). Consequently, a key element of understanding monitoring from a critical perspective lies in tracing and, ultimately, disentangling these configurations of subjects and objects (see also Landri 2018). Thus, relations fabricated through “[...] practices of sorting, naming, numbering, comparing, listing, and calculating” (Lury et al. 2012: 3) play a central role, whether performed by humans or (automated/algorithmized) technologies, because they not only relate things *as* data to other things *as* data in particular ways, but build up particular spaces of comparison and visibility (Savage 2019: 9; Thompson/Cook 2015: 734). In other words, while the relation of data introduces “[...] new continuities into a discontinuous world” (Lury et al. 2012: 3) (e.g. by creating a numerical space of assessment that relates students to each other by their assessment results), it simultaneously creates new discontinuities and differences (e.g. between different performance groups or assessment domains).

As emphasized in critical data studies literature, all of this has important political implications, particularly when applied to systems of (high or low stakes) accountability, as in the case of school monitoring. In other words, and building on Kitchin and Lauriault (2014: 4-5), monitoring infrastructures are always “[...] expressions of knowledge/power, shaping what questions can be asked, how they are asked, how they are answered, how the answers are deployed, and who can ask them” (see also Ruppert et al. 2017). I will discuss such political implications further when illustrating specific examples of *doing* monitoring in state education agencies in the next section.

3. (Re-)Building relations for monitoring: some empirical insights

In all state-level agencies under study (see footnote 2), the technical infrastructure and processes of school monitoring are, at least theoretically, relatively clearly designed, with data undergoing a journey from data collection, via validation, processing and modelling, to reporting (Figure 1):

Figure 1. A typical technical infrastructure of data-based school monitoring in German and US state-level education agencies.



Source: Hartong/Förschler 2019: 4.

Behind this typical technical infrastructure, however, lies enormous complexity, which in practice includes various interdependencies, scopes of action, and data flowing back and forth multiple times. Consequently, the data experts we interviewed clearly contrasted their work around data with linear procedures or loop circle models, instead describing it as highly experimental and messy, requiring them to manoeuvre between very different logics, stakeholders, and problems. Against this backdrop, tracing how school monitoring data actually enter or are selected into a particular form, related to each other, and – affected by these forms and relations – become information or governing knowledge (Sellar 2017; Thompson/Sellar 2018: 4-5), is far from easy. Still, however, we identified somewhat typical moments, contexts and challenges in relation-making, which many of our interviewees in fact described as very controversial (for a closer analysis of these contexts see

Hartong/Förschler 2019), and which were shown to have a significant impact on the capabilities and powers of the monitoring infrastructure. Next, I will use examples to illuminate such contexts, using both an epistemological perspective and a perspective which Drucker (2010) describes as “graphesis”.

3.1 Doing monitoring as epistemological relation-making

Every stage of collecting, validating, processing, modelling or reporting data in state education agencies represents the many ways “[...] a governing entity can define what variables are important [...] and, by extension, what’s not important” (Mattern 2015), in order to fabricate a valid representation of schools, teachers and students. Thus, for example, a key challenge of building monitoring infrastructures in state education agencies is selecting which data is collected via student or school information systems and *defining* that data. This moment of definition appears all the more important given that all the state education agencies we studied invested heavily in reducing data duplication and data alternatives for measuring the same phenomenon. They did so by implementing centralized school information systems (which, in the US, are named *State Longitudinal Data Systems*, SLDS), data standards (e.g. what format particular data – for example age, gender or socio-economic status – can have, whether it is measured using numbers or letters etc.), data business rules (e.g. defining the terminological framework of data collection, including how particular data relate to other data) and interoperability frameworks (e.g. standardizing the technical interfaces of data collection). Such standardization procedures define which data can (or cannot) be fed into the system and how data “[...] is captured because it conforms to the rules and hypothesis” (Drucker 2010: 7).

Beyond the complexity of data collection, epistemological relation-making then plays a key role in the analysis of collected data, which particularly includes practices of indicator-based modelling. Thus we identified what Kitchin et al. (2015: 8-9) have similarly documented in their study on city dashboards: that there are several types of indicators, spanning from single and composite indicators, to descriptive and contextual indicators, diagnostic, performance and target indicators, to predictive and conditional indicators – each type of indicator again inscribed with particular ideas about ways of presenting data and governing schools (e.g. ideas of *good schooling* influencing the numerical targets schools are expected to meet). Furthermore, modelling with multiple indicators always includes weighting procedures, and thus decisions about which indicators should count less or more than others, and how calculation is performed (e.g. using regression analysis). Our interviewees made very clear that relative weighting was shown to have a profound effect on the resulting performance scores (see also

Kitchin et al. 2015: 22) and, thus, the expectations and consequences schools might face. For example, the *College and Career Ready Performance Index* (CCRPI), which the state department in Georgia, US, uses as the key performance measurement to hold schools accountable, alone builds on nine different models (measuring e.g. content mastery, progress, closing gaps and graduation rates for different student subgroups in each school), which each again consist of multiple applications and weighting procedures (see <http://www.gadoe.org/CCRPI>).

Given that a key role of state-level education agencies is to support schools, teachers or students in need, another productive moment of data-based relation-making lies in the definition and modelling of *neediness*. Measuring neediness is not only based on failure to reach performance goals, but also depends on making visible and recognizing the circumstances that underprivileged schools or students might be facing, ultimately seeking to increase the fairness of monitoring and governance. As an example, the German state Hamburg uses a social index (*Hamburger Sozialindex*) as part of its monitoring system to classify the socio-economic status of schools, which is then used not only to determine state-provided resources (which increase as a school's index score lowers), but also to statistically calculate 'school peers' for evaluating test performance (deemed to be *fair comparison*). Another example is the *Early Warning Indicator System* (EWIS) used in the US state Massachusetts, which flags individual students deemed unlikely to pass particular educational milestones based on particular student characteristics (<http://www.doe.mass.edu/ccr/ewi/>). The model thus relationally creates particular groups of students that are perceived as being "at risk" (see also Ratner 2019), which then also legitimizes particular digital forms of "targeted surveillance" (Hansen 2015: 213). In fact, the social index in Hamburg or EWIS in Massachusetts are just two examples of how explicit or implicit assumptions about neediness and risk underlie *every* indicator used for school monitoring. In each case, such assumptions are linked to particular norms, values and political expectations, thus paving the way for what other digital governance studies have already described as "predictive regulation" (e.g. Williamson 2017; Mattern 2015). Such predictive regulations are further empowered by attempts to link school-level data to other data sources from both early childhood and post-graduation, already clearly visible in so-called P-20⁵ data systems used in the US (in Germany, similar data linkages have not yet been implemented).

While such an epistemological perspective puts its emphasis on the definition, relation and thus representation of educational phenomena *as* data, it is at the same time closely linked to the visualization of these data or, to build on Drucker (2010), their graphical expression.

5 P-20 stands for the integration of data from preschool (sometimes even pre-kindergarten) to high school, college, and the workforce.

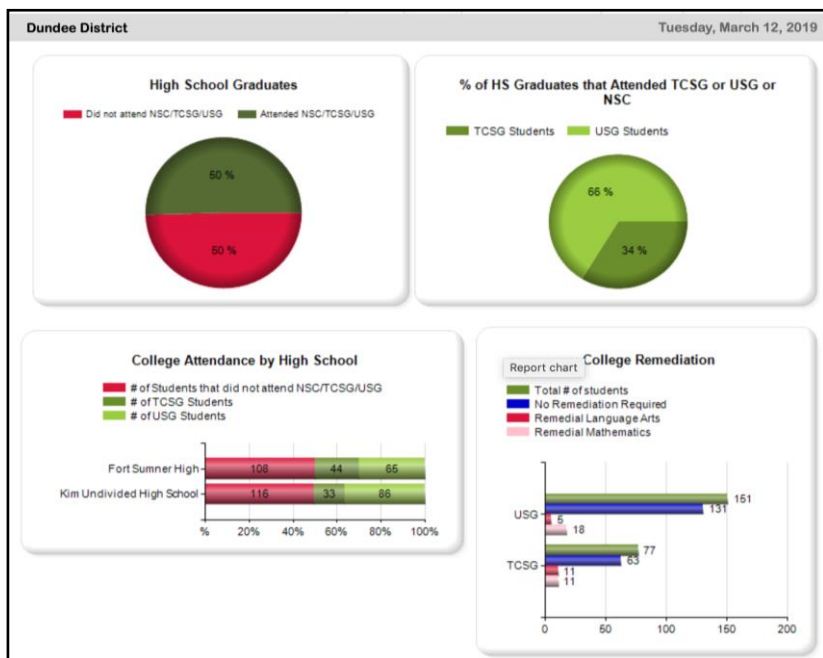
3.2 Doing monitoring as graphical relation-making

In 2010, Drucker argued for the need to establish a more critical understanding of visual knowledge production – which she terms *graphesis* – and its growing relevance for today’s digital governance technologies. Quite similar arguments have been made in critical data studies, for example by Williamson (2015: 2) who stated that “[...] graphical forms of display invite particular forms of social action from its audiences”, or by Kitchin and colleagues (2015: 6) who illustrated how the powerful realistic epistemology of dashboards is closely linked to their visual presentation.

In line with this argumentation, it seems important to also look at school monitoring infrastructures from a perspective of visualization as productive relation-making, in other words, the creation of “[...] a form that is already an argument” (Drucker 2010: 17). This not only includes the visual formation of reporting tools (e.g. data dashboards used in monitoring portals, see below), but in fact refers to all stages of the data journey. For example, when we ask which data is collected from school information systems using what kind of business rules, this should also include the graphical structures these rules imply. The easiest illustration of this is a table, such as a timetable (e.g. for guiding and standardizing data collection), which functions “[...] by putting discrete cells of information into a meaningful syntactic relation with each other” (Drucker 2010: 18).

However, when observing monitoring infrastructures, the most salient area of graphesis is the *data-out* process, which includes the fabrication of data platforms, dashboards and portals for different audiences (schools, teachers, parents and other governing agencies). In all our studied monitoring infrastructures, we found a highly complex surface of various graphical (often interactive) dashboards, including diagrams, maps, flow charts and histograms. While each of these dashboards not only visually defines particularly related entities, they are also inscribed with valuation, as the following dashboard from Georgia’s *State Longitudinal Data System* (SLDS) illustrates (see Figure 2):

Figure 2. The high school feedback tool as part of the Georgia State Longitudinal Data System



Source: SLDS Demo version <https://sldstrn.gadoe.org/SLDSDemoWeb/helpdesk/LDSHelpDesk.aspx?Name=Power%20School>.

The dashboard, which is called *Highschool Reporting*, aims at providing information to districts (among others) about the postgraduation careers of their former students, using not only different forms of graphical expression (e.g. pie and bar charts), but also specific pictorial, mostly traffic-light, colors. While the producers of monitoring usually argue that such visual strategies facilitate readability and meaningfulness for different audiences, we also see how visual expression or colors powerfully shape what questions can be asked, how they are answered, and how the answers are deployed and to whom. In fact, we found that such a strategic design of graphical expressions is becoming increasingly important for state education agencies that seek to compress and communicate as much data as possible to various audiences, but simultaneously fear the dangerous pitfalls of visual misinterpretation from data non-experts – particularly when (unintended) causations could be assumed. Consequently, the agencies we studied increasingly invest in dashboard user training and support, which, again, alongside the visualizations

themselves, should be investigated as a powerful part of the enactment of monitoring infrastructures.

4. Outlook

As these selected examples illustrate, there are different perspectives which may offer a fruitful way to look at school monitoring infrastructures from a critical data studies perspective. Such a perspective explicitly challenges the idea of data-based monitoring as neutral, evidence-guided and de-politicized, and instead emphasizes the various, yet often hidden, moments of re-politization (Hansen 2015: 204), which lie in the powerful *infrastructuring* of visibilities and invisibilities (see also West 2017: 1). These visibilities and invisibilities shape the numerical and graphical representation of schools, teachers and students, which administrators, but also dashboard users, then use to act upon, and consequently change, the educational world. Additionally, because new data continuously produces the need for more and better data (Thompson/Sellar 2018), in state education agencies there is also a constant increase in data management and business ruling, ultimately shifting more and more attention towards what is captured on screens and dashboards. It is important to mention, however, that behind the fabrication of data representations lie various moments of relation-making where “change is immanent in conduct” (Ruppert 2012: 129), which is to say that monitoring infrastructures always enact multiple things at the same time. In other words, digital education governance, at least in our cases under study, does not appear to produce single centers of calculation and data power, but instead multiple infrastructures and often messy practices that together perform calculation, commensuration and representation work. Consequently, a key task for digital education governance studies lies in what Gray and colleagues (2018: 1) recently described as promoting *data infrastructure literacy*, which is “[...] the ability to account for, intervene around and participate in the wider socio-technical infrastructures through which data is created, stored and analyzed”. In that regard, the presented project on the ongoing transformation of school monitoring only provides some initial ideas and findings, seeking to pave the way for further studies.

References

- Conaway, Carrie/Keesler, Vanessa/Schwartz, Nathaniel (2015): What research do state education agencies really need? The promise and limitations of state longitudinal data systems. In: *Educational Evaluation and Policy Analysis*, 37, 1, pp. 16-28.
- Dederling, Kathrin (2015): Entscheidungsfindung in Bildungspolitik und Bildungsverwaltung. In: Altrichter, Herbert/Maag Merki, Katharina (eds.): *Handbuch neue Steuerung im Schulwesen*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 53-73.
- Drucker, Johanna (2010): Graphesis. In: *poetess archive journal*, 2, 1. <https://journals.tdl.org/paj/index.php/paj/article/view/4>
- González-Sancho, Carlos/Vincent-Lancrin, Stéphan (2016): Transforming education by using a new generation of information systems. In: *Policy Futures in Education*, 14, 6, pp. 741-758.
- Hansen, Hans K. (2015): Numerical operations, transparency illusions and the datafication of governance. In: *European Journal of Social Theory*, 18, 2, pp. 203-220.
- Hartong, Sigrid (2016): Between Assessments, Digital Technologies, and Big Data: The Growing Influence of 'Hidden' Data Mediators in Education. In: *European Educational Research Journal*, 15, 5, pp. 523-536.
- Hartong, Sigrid (2018): The transformation of state monitoring systems in Germany and the US: relating the datafication and digitalization of education to the Global Education Industry. In: Parreira do Amaral, Marcelo/Steiner-Khamsi Gita/Thompson, Christiane (eds.): *Researching the Global Education Industry – Commodification, the Market and Business Involvement*. Cham, Switzerland: Palgrave Macmillan, pp. 157-180.
- Hartong, Sigrid (2019): "Bildung 4.0? Kritische Überlegungen zur Digitalisierung von Bildung als erziehungswissenschaftliches Forschungsfeld". In: *Zeitschrift für Pädagogik*, 3, pp. 424-444.
- Hartong, Sigrid/Förschler, Annina (2019): Opening the black box of data-based school monitoring: data infrastructures, flows and practices in state education agencies. In: *Big Data & Society*, 6, 1, pp. 1-12.
- Houben, Daniel/Priest, Bianca (eds.) (2018): *Datengesellschaft. Einsichten in die Datafizierung des Sozialen*. Bielefeld: Transcript.
- Iliadis, Andrew/Russo, Federica (2016): Critical data studies: An introduction. In: *Big Data & Society*, 3, 2, pp. 1-7.
- Kitchin, Rob/Lauriault, Tracey P. (2014): Towards critical data studies: Charting and unpacking data assemblages and their work. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Kitchin, Rob/Lauriault, Tracey P./McArdle, Gavin (2015): Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. In: *Regional Studies, Regional Science*, 2, 1, pp. 6-28.
- Landri, Paul (2018): *Digital Governance of Education: Technology, Standards and Europeanization of Education*. London: Bloomsbury.
- Lury, Celia/Parisi, Luciana/Terranova, Tiziana (2012): Introduction: The becoming topological of culture. In: *Theory, Culture & Society*, 29, 4-5, pp. 3-35.

- Mattern, Shannon (March 2015): Mission Control: A history of the Urban Dashboard. In: Placesjournal.org. <https://placesjournal.org/article/mission-control-a-history-of-the-urban-dashboard/>. Accessed 26th March 2019
- Ratner, Helene (2019): Describing children at risk: experiments with context. In: Discourse: Studies in the Cultural Politics of Education, 40, 1, pp. 16-28.
- Ruppert, Evelyn (2012): The governmental topologies of database devices. In: Theory, Culture & Society, 29, 4-5, pp. 116-136.
- Ruppert Evelyn/Isin, Engin/Bigo, Didier (2017): Data politics. In: Big Data and Society, July-September, pp. 1-7.
- Savage, Glenn (2019): What is policy assemblage? In: Territory, Politics, Governance. <https://www.tandfonline.com/doi/full/10.1080/21622671.2018.1559760>
- Sellar, Sam (2017): Making network markets in education: the development of data infrastructure in Australian schooling. In: Globalisation, Societies and Education, 15, 3, pp. 341-351.
- Selwyn, Neil (2014): Data entry: towards the critical study of digital data and education. In: Learning, Media and Technology, 40, 1, pp. 1-18.
- Thompson, Greg/Cook, Ian (2015): Becoming-topologies of education: deformations, networks and the database effect. In: Discourse: Studies in the Cultural Politics of Education, 36, 5, pp. 732-748
- Thompson, Greg/Sellar, Sam (2018): Datafication, testing events and the outside of thought. In: Learning, Media and Technology, 43, 2, pp. 139-151.
- West, John (2017): Data, democracy and school accountability: Controversy over school evaluation in the case of DeVasco High School. In: Big Data & Society, June, pp. 1-16.
- Williamson, Ben (2015): Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. In: Journal of Education Policy, 31, 2, pp. 123-141.
- Williamson, Ben (2017): Big Data in Education: The digital future of learning, policy and practice. Newbury Park: SAGE.

Data, Diagnosis and Prescription: Governing Schooling through the OECD's PISA for Schools

Steven Lewis¹

1. Introduction

This chapter explores *PISA for Schools*, an instrument developed by the Organisation for Economic Cooperation and Development (OECD), in collaboration with a diverse array of (largely US-based) partner organizations, including philanthropic foundations, not-for-profit agencies and commercial edu-businesses. PISA for Schools, a school-based variant of the OECD's influential Programme for International Student Assessment (PISA) test, not only assesses school performance in reading, mathematics and science against international schooling systems, but also promotes examples of what the OECD presents as best practices from notionally world-class schooling systems (i.e., as measured by PISA), as well as the policy expertise of the OECD itself. This arguably reflects the expanding scope, scale and explanatory power of the OECD's education policy work (Sellar and Lingard 2014), which helps extend the relevance of PISA beyond national policymakers and political leaders into decidedly more *local* schooling spaces (i.e., schools and schooling districts). Specifically, my focus here is how PISA for Schools helps to constitute new spaces and relations of global education policymaking, and how these emergent relational or *topological*, spatialities enable the OECD to influence how schooling is locally thought and practiced.

The emergence of global governance in education has been documented during the previous two decades (Lewis/Lingard 2015; Meyer/Benavot 2013), with such global processes, discourses and relations recognized as exerting considerable influence over how schooling is enacted in national and, increasingly, subnational (e.g., state/province, schooling district, school) spaces. While the nature and effects of these developments have often been examined at the level of national (and subnational/state) schooling systems, there has been less consideration given to how such global policy ensembles seek to influence, and actually do influence, local schooling spaces. I wish to emphasize here the relational and productive capacities of space to examine how the OECD can now exercise educational governance by, topologically

1 Steven Lewis is an ARC DECRA Fellow at the Education Governance and Policy (EGP) group within the REDI (Research for Educational Impact) Centre of Deakin University. Email: steven.lewis@deakin.edu.au

speaking, “reaching into” (Allen/Cochrane 2010: 1075) more practice-focused schooling spaces, rather than remaining at the policy level of the global and nation-state *vis-à-vis* the main PISA test. Given the significant, and frequently documented, normative influences exerted by main PISA and the OECD on national schooling systems (Fischman et al. 2019; Rautalin/Alasuutari/Vento 2019), it seems logical that PISA for Schools should warrant a similar level of critical scrutiny, particularly for its potential to respatialize relations of educational governance and position schools within what is now a global space of measurement and comparison.

In what follows, I first briefly describe the PISA for Schools test. Then, I introduce my theoretical framework, which draws together diverse thinking around commensuration, the increasing role of data, and processes of *datafication* (Hartong/Piattoeva 2019; Jarke/Breiter 2019; Lewis/Holloway 2019; Lycett 2013) in contemporary schooling governance and practice. In particular, I employ Simons’ (2015) notion of governing by examples to understand how the inclusion of best practices – alongside quantitative performance data – within the PISA for Schools report constitutes a unique form of evidence, facilitating new modalities of global education governance within decidedly local schooling spaces; that is, *governing by best practice* (Lewis 2017). Best practice in this way can thus be considered an integral form of *soft* qualitative evidence – such as PISA-informed policies and practices – that works alongside *hard* quantitative performance data. My analyses suggest that PISA for Schools exerts a governing influence through both numbers *and* examples, which allows the OECD to discursively and normatively constrain how world-class schools and systems, and their policies and practices, are defined.

2. PISA for Schools: the test and report

PISA for Schools – known in the USA as the *OECD Test for Schools* (based on PISA) – is similar in design and appearance to the main PISA survey, comprising a two-hour written test that assesses the ability of 15-year-old students to apply their acquired classroom knowledge in reading, mathematics and science to notionally real-world situations. Like the main PISA exam taken by schooling systems, PISA for Schools is not aligned to any particular national curriculum. Unlike the main PISA test, however, PISA for Schools assesses (and compares) a school’s local performance in reading, mathematics and science against that of schooling systems. In addition to assessing student performance, the test contains student and principal questionnaires that generate contextual information about particular in-school (e.g., class disciplinary climate) and out-of-school influences (e.g., student attitudes

towards reading) on student learning. These contextual questionnaires ask students questions about the learning environment and student engagement with their teachers and school classes, while principals respond to questions concerning school resourcing, governance and the socio-economic makeup of the school community. Such contextual information allows subject performance data (in reading, mathematics and science) to be reported against relative socio-economic advantage, as well as student attitudes towards the teaching and learning of these respective subjects.

Development of the program began in 2010, with English-speaking US, UK and Canadian schools invited by the OECD in late 2011 to participate in a pilot study. This was designed to equate the new school-based test with main PISA, so that direct comparisons could be made between school (PISA for Schools) and schooling system (main PISA) performance. PISA for Schools test items were developed according to the relevant PISA assessment frameworks for reading, mathematics and science, and equated to the existing PISA scales (Level 1 to Level 6) by simultaneously anchoring them with main PISA “link items” against a common PISA metric. This process enabled PISA for Schools scores for reading, mathematics and science to be reported against the established PISA proficiency scales, and against the performance of schooling systems as measured by main PISA. Following a successful field trial of 127 schools, PISA for Schools was officially launched in the USA in April 2013, and made available to all eligible schools and districts throughout the country. Since this time, PISA for Schools has experienced a significant expansion in terms of its availability and administration². As of 2020, PISA for Schools is available in twelve languages across fourteen countries, and it has been cumulatively administered in more than 2,200 schools globally (OECD 2019a).³

Another key feature of PISA for Schools is the school-level report provided by the national accredited provider (OECD 2017). All schools participating in PISA for Schools receive a report that analyzes their students’ performance and contextual data, as well as providing examples of best practices from high performing international schooling systems (e.g.,

-
- 2 Janison Education Group (‘Janison’), an Australian for-profit education technology company, was announced in 2019 as the global provider of the software platform on which the online version of PISA for Schools is delivered. Since then, it has signed agreements with the National Service Providers (NSPs) of Brazil (June 2019) and the Russian Federation (September 2019). In October 2019, Janison announced that it was also accredited to be the sole NSP for all U.S. schools. At the time of publication, with Janison as the accredited NSP for the U.S., schools pay US\$5,000 to participate in the online version PISA for Schools.
 - 3 PISA for Schools is now available in the following 14 jurisdictions: Andorra, Brazil, Brunei Darussalam, China (PRC), Colombia, Japan, Pakistan, Portugal, Russia, Spain, Thailand, the United Arab Emirates, the UK and the USA. It is also deliverable in the following 12 languages: Arabic, Basque, Catalan, English, Galician, Japanese, Mandarin (Chinese), Portuguese, Russian, Spanish (Castilian), Thai and Welsh.

Shanghai-China, Finland, Singapore) and excerpts from the OECD's broader educational policy research. However, and besides the graphs and tables representing a school's specific data around student performance or local contextual factors, the report is otherwise entirely identical for all participating schools within the same national jurisdiction (e.g., the US). For instance, the examples of best practice within the report, as well as the excerpts from other OECD research publications, are *identical* for all US schools, and there are no modifications to the report contents to acknowledge a school's specific context (e.g., whether a school is deemed high/low performing on PISA for Schools). This arguably promotes the logic that *all* schools both equally require and can benefit from the same OECD policy lessons, even if such assumptions problematically downplay the role of local context and non-educational effects to performance on standardized assessments like PISA (Feniger/Lefstein 2014; Meyer/Schiller 2013; Tan/Yang 2019).

3. Commensuration, datafication and governing by examples

Commensuration, or the “transformation of different qualities into a common metric” (Espeland/Stevens 1998: 314), is by no means a recent phenomenon. Much attention has previously been paid to the role of numbers and statistics in the historical constitution of the nation-state as a knowable, and governable, political space (see Desrosières 1998; Hacking 1990; Porter 1995). Indeed, these data help inscribe the very spaces they purport to represent, achieving what has been described as “the mutual construction of statistics and society” (Sætnan/Lomell/Hammer 2011: 1), and numbers have played a central role in helping to constitute a commensurate global education policy field (Lingard and Rawolle 2011). However, while the productive capacities of numbers and data are largely beyond question, it is worth problematizing precisely *what* is produced in these processes of commensuration, and particularly how these common spaces of measurement can “render some aspects of life invisible or irrelevant” (Espeland/Stevens 1998: 314). As Ball (2003: 217) argues in his examination of performativity upon the soul of the teacher, such data-driven commensuration helps translate “complex social processes and events into simple figures or categories of judgement”, which often has considerable consequences for how teachers and teaching itself are constituted (Holloway/Brass 2018; Lewis/Holloway 2019). Moreover, abstracting complex qualities into simple and reductive quantities through data-driven processes of commensuration “will unavoidably channel users

towards some kinds of inferences and/or actions more readily than others” (Lycett 2013: 384; emphasis added). It is these dual effects of commensuration, simultaneously both reductive *and* productive, that help to illuminate how internationally comparative measures of schooling performance and PISA for Schools in particular, help to enable the governance of education.

Building on such processes of commensuration is the increasing focus on data, and especially digital data. To this end, I consider the *datafication* of education as enabling (and even encouraging) every aspect of schooling, students and teachers to be constituted *as data* – to be collected, analyzed, surveilled and controlled (Bradbury 2019; Selwyn/Henderson/Chao 2015; Williamson 2017). This inclination to datafication has been followed, in turn, by the emergence of new *digital technologies* (e.g., data dashboards, learning platform observation apps, etc.), *services* (e.g., data analysis) and even *professionals* (e.g., data stewards, technology coaches), subjecting schools and schooling systems to unforeseen levels of surveillance and control. It is important to note, however, that constructions of schooling accountability, practices and leadership are never purely technical procedures, but are instead a complex entanglement of very different (technical and social) logics, practices and problems (see, for instance, Hartong/Förschler 2019; Hartong/Piattoeva 2019; Lewis/Hardy 2017). Far from somehow being neutral or objective, such data-centric processes – of collecting, recoding, storing, analyzing, distributing and comparing data – have now become integral features of contemporary modes of digital educational governance (Hartong 2016; Thompson/Sellar 2018; Williamson 2016).

These putatively objective data have also been used to legitimate prospective policy decisions in what has been described as *evidence-informed* policymaking (Lingard 2013). Similar to the production of data being informed by contingent socio-technical factors, the use of such evidence is never purely objective, but is instead always mediated by political judgments, prioritization and values. Even so, the centrality of *hard* data to educational governance and policymaking should not lead us to overlook newer modalities that incorporate other *soft(er)* forms of qualitative evidence, including examples of what works. Such evidence-informed policymaking can be considered, in this instance, to have progressed from merely addressing, on the basis of performance measures and comparisons, “*Is reform necessary?*” Indeed, perhaps the more pressing question these forms of evidence force us to now ask is “*What type of reform is necessary?*” Simons (2015: 715) usefully describes this evolution of governing through data as “governing by examples”:

[G]overning through evidence is not only about governing by numbers but also includes a mode of *governing by examples*. To a large extent, the examples of good practice are examples of good performance and are being decided upon available

numerical performance data. In that sense, governing by examples is to be regarded as complementary to governing by numbers. (emphasis added, SL)

Here, *qualitative* forms of evidence – such as narrative accounts, examples of successful practices and even educators’ own professional experiences – are used to provide additional richness to enumerations of performance, but these qualitative accounts are still framed in terms of their ability to improve *quantitative* performance. That is, for best practices to “work”, they must demonstrate the ability to improve performance in a way that can then be captured *quantitatively* (e.g., via standardized tests, such as PISA for Schools). This has arguably led to a disproportionate focus by researchers, policymakers and educators seeking to determine the policies and practices of top-performing schooling systems (Auld/Morris 2016; Lewis 2017).

Herein is the central premise of most (if not all) large-scale international assessments, where culturally different and geographically distant schooling systems and schools are rendered relationally – or *topologically* – proximate through reference to common measures and metrics (see Lewis/Sellar/Lingard 2016). This creates a situation whereby school performance is not only able to be compared but, in fact, *should be* compared, and where such comparisons are seen as a valid way of informing local schooling policies through a looking around at, and learning from, the global. Taking this rationale of policy borrowing from successful schooling systems to its ultimate (if not necessarily logical) conclusion, I would argue, in agreement with Kamens (2013: 124), that “[i]f one can compare school systems [or schools] in terms of their characteristics and outcomes, the idea of borrowing features from the ‘best’ systems is a natural corollary”. As we shall see, however, the rationales underpinning the search for decontextualized, data-driven best practices can lead to a significant “oversimplification of more complex contexts and issues” (Wiseman 2010: 4). This can, in turn, produce problematic consequences for the local teachers and school leaders who might attempt to uncritically borrow examples of what works.

4. Defining “what works” through data

A central aspect of the OECD’s educational governance is arguably the creation of a commensurate space of PISA measurement, within which participating schools and schooling systems are rendered knowable and comparable through reference to PISA data and assessment frameworks. This putative commonality then enables PISA for Schools performance, and especially any perceived *difference* in performance data between schools and high-performing schooling systems, to be used to justify school-level reform

measures (see also Lewis 2018). However, the question of *which* reforms should be implemented, and *how* such reforms might be undertaken, remain stubbornly unanswered on the basis of performance data alone.

It is here that the inclusion of global best practices in the PISA for Schools reports helps the OECD to further steer local processes of schooling reform, with these qualitative examples of successful policies and practices accompanying the quantitative data that compares local (school) and international (schooling system) PISA performance. Besides simply measuring a school's relative performance, a key governing modality of PISA for Schools is the promotion of certain strategies, policies and practices from high performing schooling systems to participating schools. To this end, the OECD has mandated the inclusion of prominent breakout boxes in the PISA for Schools report that highlight the policies and practices of celebrated PISA poster children, including Shanghai-China, Singapore, Finland and Japan. Significantly, these schooling systems have been determined (on the basis of their performance on main PISA) to be "the world's top performing school systems" (OECD 2019b), with the implication being that schools now have a ready prescription of *how* they should act in order to be among other notionally top-performing systems. Such practices also help to validate and strengthen the policy credentials of the OECD, as the inclusion of what works from PISA-validated schooling systems suggests that these policy solutions are already *tried and tested*. By establishing this pedigree of successful implementation in other high performing systems, the OECD is clearly encouraging local educators to have confidence in the efficacy of the proffered policy reforms – namely, that what works actually *works*.

Best practice is thereby understood entirely by reference to schooling system performance on main PISA, while other potential considerations of best practice are excluded. We can see then the productive power of such discourses, and how it is the OECD (and not teachers, schools or districts) that ultimately controls *who* is high performing and, in turn, *which* are the best practices responsible for such performance. Even the concept of best practice itself is presented through PISA for Schools in a largely unproblematized and self-evident manner, as though participating teachers and schools should no more question the notion of best practice than they should the OECD's representation of these very practices. As noted in the PISA for Schools *Technical Report*,

[...] the PFS [PISA for Schools] provides important peer-to-peer learning opportunities for educators – locally, nationally and internationally – as well as the opportunity to share good practices to help identify "what works" to improve learning and build better skills for better lives. (OECD 2015: 9)

Moreover, the OECD (2013: 5) even suggests the "sharing of effective practices" between international schooling systems and local schools via the PISA for Schools report is a "logical next step" when school leaders look to

implement schooling reform processes. Teachers are thus presented with a deceptively linear relationship between i) measuring schooling performance, ii) determining what works within other putatively successful schooling systems and then iii) adopting these self-same practices in order to improve learning outcomes at local schooling sites.

The inclusion of best practice in the PISA for Schools reports specifies an ensemble of qualitative evidence from school systems with quantitative success on main PISA, providing the necessary complementarity between quantitative and qualitative forms of evidence, and governance by numbers *and* examples (see Simons 2015). Poor local performance on PISA for Schools, especially when compared with that of high-performing schooling systems, arguably encourages participating schools to adopt the OECD's proffered examples of best practice, where the hard evidence of numerical data authoritatively validates the soft examples of best practice. Further reflecting this complementarity of numbers and examples, schools are seemingly encouraged to look to Shanghai-China, a normative "looking east" (Sellar/Lingard 2013a) that is presumably based on the municipality's world-leading performance on PISA. The OECD's logic here is, in turn, inescapable: successful performance is attributable to successful practices, and such practices can be readily transferred between settings and contexts.

The supposed link between success on PISA and the implementation of successful policies thus presents such examples of best practice in, arguably, a causal light, as though the adoption of certain schooling policies is directly responsible for (measurable) improvements in student performance data. However, this largely ignores the numerous non-policy factors that can (and frequently do) influence student learning and PISA performance outcomes (Feniger/Lefstein 2014; Meyer/Schiller 2013). Instead, policy is positioned as the overwhelming influence on school performance while culture is understood as something *external* to schooling, rather than culture being central to how education is locally understood and given meaning. As such, there is little overt consideration given to how participating schools and notionally high performing systems might also be substantively *different* in terms of socio-economic, cultural, historical or geographic factors. This decoupling of best practice from its original context demonstrates the largely epistemological nature of the OECD's global educational governance and influence, which depends on "stressing the importance of policy factors over the effects of cultural and social context" (Sellar/Lingard 2013b: 723).

It is this PISA-mediated linking of performance and best practice that enables the OECD to normatively define both what schools should strive towards (i.e., PISA-world class status) *and* how they should notionally attain such goals (i.e., adopt global best practices), with the processes of data-driven diagnosis and prescription being inseparably intertwined and, importantly, the OECD positioned as *the* global expert on matters of education policy. This

sense that the OECD “knows best” is clearly evident, providing education policy advice that seemingly elides contextual considerations within, and between, local schools and national schooling systems, reducing the potential for schools to individualize their policy responses in ways that address and acknowledge local contexts. As Grek (2013: 707) rather tellingly notes, this supposedly universal advice reflects the OECD’s imbrication of knowledge and policy so that knowledge *is* policy, in which “expertise and the selling of undisputed, universal policy solutions drift into one single entity and function”.

I should emphasize here there is nothing innately wrong with local educators accessing the work of the OECD, or any other policy authority for that matter, to help inform their teaching practice and reform measures. However, it *is* arguably problematic when PISA data becomes the dominant (or only) contribution to this process, with the danger being that the OECD becomes the overwhelming authority on schooling, rather than just one voice amongst many. I would also stress here how the increasing reliance on data as the means to understand and evaluate schooling, and the subsequent necessity of external data experts (e.g., statisticians, data technicians) to analyze and interpret these data, risks displacing other more professionally-oriented forms of expertise and knowledge, such as that possessed by the teaching profession (see, for instance, Lewis/Holloway 2019). This shifts not only where expertise is located, but also how such expertise is determined – what becomes most valued is the ability to understand and respond to data in a way that will, in turn, produce favorable improvements to data. In this way, the OECD may well be able to authorize what counts as valued evidence for the schools and districts that choose to participate in PISA for Schools, thereby limiting the possible ways in which schooling might be alternatively understood and practiced. We can thus see how the ready-made nature of the OECD’s proffered best practices facilitates their local uptake by schools and districts, but without first ensuring that these practices are understood in the context of the countries and systems from which they are being borrowed, or how they might align with the context.

5. Conclusion: data-driven diagnoses and PISA for Schools

I have argued here that PISA for Schools facilitates international school-to-system (and school-to-school) comparisons, situating participating schools and schooling systems within a common global education policy field (Lingard/Rawolle 2011). Importantly, this also allows their local performance data to be evaluated against notionally high-performing or fast-improving schooling systems, as determined by the results of main PISA (e.g., Shanghai-

China, Singapore, Finland). While certainly not the first time that transnational data have helped to produce commensurate global or regional education policy spaces, the inclusion of *individual schools* marks what is, arguably, a significant development. In this sense, the OECD presents PISA for Schools as a logical *next step* for local policymakers and educators, being an effective means to obtain knowledge on school performance in the same way that main PISA purportedly evaluates national systems. Participating schools can thus receive the *imprimatur* of the OECD, demonstrating to local, national and international stakeholders that they are an OECD-approved world-class institution that adequately prepares its students for educational success in the global economy. The ability of PISA for Schools to produce legitimate and internationally recognized proof of a school's performance may thus make such evidence a valued commodity for local communities, and especially so for schools that are doing well in relation to national under-performance on main PISA (e.g., the decreasing national performance of the USA).

In effect, PISA for Schools serves a dual role, providing a data-driven diagnosis of local performance *and* a prescription of the policies that should be implemented to improve performance. Consequently, the dominant rationale around best practice in the PISA for Schools report might best be described as solutions looking for a problem, with the OECD ostensibly determining which set of global best practices is most appropriate for local implementation by all schools in all circumstances. Arguably, this makes sitting the test, and the data that are generated, somewhat redundant beyond providing schools with the impetus to act upon the OECD's policy recommendations. In this, we can perhaps see evidence of what Jessop (2008) describes as "policy Darwinism", whereby certain policies – in this instance, those of the OECD – come to discursively and materially dominate, and possibly even exclude, other articulations and futures of schooling.

References

- Allen, John/Cochrane, Allan (2010): Assemblages of state power: Topological shifts in the organization of government and politics. In: *Antipode*, 42, 5, pp. 1071-1089.
- Auld, Euan/Morris, Paul (2016): PISA, policy and persuasion: Translating complex conditions into education 'best practice'. In: *Comparative Education*, 52, 2, pp. 202-229.
- Ball, Stephen (2003): The teacher's soul and the terrors of performativity. In: *Journal of Education Policy*, 18, 2, pp. 215-228.

- Bradbury, Alice (2019): Datafied at four: The role of data in the 'schoolification' of early childhood education in England. In: *Learning, Media and Technology*, 44, 1, pp. 7-21.
- Desrosières, Alain (1998): *The politics of large numbers*. Translated by Camille Nash. Cambridge: Harvard University Press.
- Espeland, Wendy Nelson/Stevens, Mitchell L. (1998): Commensuration as a social process. In: *Annual Review of Sociology*, 24, pp. 313-343.
- Feniger, Yariv/Lefstein, Adam (2014): How not to reason with PISA data: An ironic investigation. In: *Journal of Education Policy*, 29, 6, pp. 845-855.
- Fischman, Gustavo E./Marcetti Topper, Amelia/Silova, Iveta/Goebel, Janna/Holloway, Jessica (2019): Examining the influence of international large-scale assessments on national education policies. In: *Journal of Education Policy*, 34, 4, pp. 470-499.
- Grek, Sotiria (2013): Expert moves: International comparative testing and the rise of expertocracy. In: *Journal of Education Policy* 28, 5, pp. 695-709.
- Hacking, Ian (1990): *The taming of chance*. Cambridge: Cambridge University Press.
- Hartong, Sigrid (2016): Between assessments, digital technologies and big data: The growing influence of 'hidden data mediators in education. In: *European Educational Research Journal*, 15, 5, pp. 523-536.
- Hartong, Sigrid/Förschler, Annina (2019): Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies. In: *Big Data & Society*, 6, 1, pp. 1-12.
- Hartong, Sigrid/Piattoeva, Nelli (2019): Contextualising the datafication of schooling: A comparative discussion of Germany and Russia. In: *Critical Studies in Education*, pp. 1-16.
- Holloway, Jessica/Brass, Jory (2018): Making accountable teachers: The terrors and pleasures of performativity. In: *Journal of Education Policy*, 33, 3, pp. 361-382.
- Jarke, Juliane/Breiter, Andreas (2019): Editorial: The datafication of education. In: *Learning, Media and Technology*, 44, 1, pp. 1-6.
- Jessop, Bob (2008): *State power*. Cambridge: Polity Press.
- Kamens, David H. (2013): Globalization and the emergence of an audit culture: PISA and the search for 'best practices' and magic bullets. In: Meyer, Heinz-Dieter/Benavot, Aaron (eds.): *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books, pp. 117-139.
- Lewis, Steven (2017): Governing schooling through 'what works': The OECD's PISA for Schools. In: *Journal of Education Policy*, 32, 3, pp. 281-302.
- Lewis, Steven (2018): PISA 'Yet To Come': Governing schooling through time, difference and potential. In: *British Journal of Sociology of Education*, 39, 5, pp. 683-697.
- Lewis, Steven/Hardy, Ian (2017): Tracking the topological: The effects of standardized data upon teachers' practice. In: *British Journal of Educational Studies*, 65, 2, pp. 219-238.
- Lewis, Steven/Holloway, Jessica (2019): Datafying the teaching 'profession': Remaking the professional teacher in the image of data. In: *Cambridge Journal of Education*, 49, 1, pp. 35-51.
- Lewis, Steven/Lingard, Bob (2015): The multiple effects of international large-scale assessment on education policy and research. In: *Discourse: Studies in the Cultural Politics of Education* 36, 5, pp. 621-637.

- Lewis, Steven/Sellar, Sam/Lingard, Bob (2016): 'PISA for Schools': Topological rationality and new spaces of the OECD's global educational governance. In: *Comparative Education Review*, 60, 1, pp. 27-57.
- Lingard, Bob (2013): The impact of research on education policy in an era of evidence-based policy. In: *Critical Studies in Education* 54, 2, pp. 113-131.
- Lingard, Bob/Rawolle, Shaun (2011): New scalar politics: Implications for education policy. In: *Comparative Education*, 47, 4, pp. 489-502.
- Lycett, Mark (2013): 'Datafication': Making sense of (big) data in a complex world. In: *European Journal of Information Systems*, 22, 4, pp. 381-386.
- Meyer, Heinz-Dieter/Benavot, Aaron (eds.) (2013): *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books.
- Meyer, Heinz-Dieter/Schiller, Kathryn (2013): Gauging the role of non-educational effects in large-scale assessments: Socio-economics, culture and PISA outcomes. In: Meyer, Heinz-Dieter/Benavot, Aaron (eds.): *The emergence of global educational governance*. Oxford: Symposium Books, pp. 207-224.
- OECD (2013): *The OECD Test for Schools (based on PISA): Questions and answers (U.S. version)*. Paris: OECD Publishing.
<http://www.oecd.org/pisa/aboutpisa/pisa-basedtestforschools.htm>
- OECD (2015): *PISA-based Test for Schools technical report*. Paris: OECD Publishing.
http://www.oecd.org/pisa/aboutpisa/PfS_TechReport_CRC_final.pdf
- OECD (2017): *How your school compares internationally: OECD Test for Schools (based on PISA)*. Paris: OECD Publishing.
- OECD (2019a): *PISA for Schools*. <http://www.oecd.org/pisa/pisa-for-schools/>
- OECD (2019b): *PISA for Schools: FAQs*. <http://www.oecd.org/pisa/aboutpisa/pisa-based-test-for-schools-faq.htm>
- Porter, Theodore (1995): *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Rautalin, Marjaana/Alasuutari, Pertti/Vento, Eetu (2019): Globalization of education policies: Does PISA have an effect? In: *Journal of Education Policy*, 34, 4, pp. 500-522.
- Sætnan, Ann Rudinow/Lomell, Heidi Mork/Hammer, Svein (2011): Introduction: By the very act of counting: The mutual construction of statistics and society. In: Sætnan, Ann Rudinow/Lomell, Heidi Mork/Hammer, Svein (eds.): *The mutual construction of statistics and society*. Oxon: Routledge, pp. 1-17.
- Sellar, Sam/Lingard, Bob (2013): Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. In: *Comparative Education* 49, 4, pp. 464-485.
- Sellar, Sam/Lingard, Bob (2013b): The OECD and global governance in education. In: *Journal of Education Policy* 28, 5, pp. 710-725.
- Sellar, Sam/Lingard, Bob (2014): The OECD and the expansion of PISA: New global modes of governance in education. In: *British Educational Research Journal*, 40, 6, pp. 917-936.
- Selwyn, Neil/Henderson, Michael/Chao, Shu-Hua (2015): Exploring the role of digital data in contemporary schools and schooling: '200,000 lines in an Excel spreadsheet'. In: *British Educational Research Journal*, 41, 5, pp. 767-781.

- Simons, Maarten (2015): Governing education without reform: The power of the example. In: *Discourse: Studies in the Cultural Politics of Education*, 36, 5, pp. 712-731.
- Tan, Charlene/Yang, Jeong A. (2019): Path-dependency or path-shaping? An analysis of the policy to target exam-orientation in South Korea. In: *Critical Studies in Education*, pp. 1-16.
- Thompson, Greg/Sellar, Sam (2018): Datafication, testing events and the outside of thought. In: *Learning, Media and Technology*, 43, 2, pp. 139-151.
- Williamson, Ben (2016): Digital education governance: An introduction. In: *European Educational Research Journal*, 15, 1, pp. 3-13.
- Williamson, Ben (2017): *Big data in education: The digital future of learning, policy and practice*. Los Angeles: Sage.
- Wiseman, Alexander (2010): The uses of evidence for educational policymaking: Global contexts and international trends. In: *Review of Research in Education*, 34, 1, pp. 1-24.

Big Data Analytics in Education: Big Challenges and Big Opportunities

*Bernard Veldkamp¹, Kim Schildkamp², Merel Keijsers³, Adrie Visscher⁴
and Ton de Jong⁵*

1. Introduction

In many sectors (medicine, transport, chain stores, etc.), large amounts of data are being collected and stored for further analysis. As stated in a review by Piety (2019), much has been invested in the use of data, and policy makers are continuously looking at the field of (big) data for solutions. Within the field of education, data can be obtained from students or teachers for specific purposes, they can be stored by third parties in administrative systems, and data can be recorded from the interaction of participants with online systems. The increase in the amount of data, together with an increased availability and accessibility of data in electronic form, and the linking of previously separated data files, is labeled 'big data'. Big data can be used to gain more insight into specific processes, to predict for example achievement, and to develop measures for improving education. Big data have the following characteristics (Laney 2001):

- Volume: It involves large quantities of data
- Variety: Data sources and the data itself differ
- Velocity: Data are added and updated continuously.

These three Vs denote that big data expand along various dimensions. The volume of the data is not the only dimension along which big data evolves. The variety of the data also increases. New technology and applications are introduced into classrooms, each of them generating new data types and sometimes even new data formats. For example, most children have smart-phones nowadays. This enables the use of all kinds of online learning tools,

-
- 1 Bernard Veldkamp is Professor of Research Methodology and Data Analytics at the University of Twente. Email: b.p.veldkamp@utwente.nl
 - 2 Kim Schildkamp is Associate Professor at the Faculty of Behavioural, Management, and Social Sciences at the University of Twente. Email: k.schildkamp@utwente.nl
 - 3 Merel Keijsers is PhD Student at the HITlab UC in Christchurch. Email: merel.keijsers@pg.canterbury.ac.nz
 - 4 Adrie Visscher is Professor at the University of Twente at the Faculty of Behavioural, Management and Social Sciences. Email: a.j.visscher@utwente.nl
 - 5 Ton de Jong is Department Head of Instructional Technology and Professor on Instructional Technology at the University of Twente. Email: a.j.m.dejong@utwente.nl

but also the real-time collection of data that provides teachers with rapid feedback (Bijlsma et al. 2019). Finally, data are streaming to computer servers at a different pace, which has to be synchronized.

Big data in education can come from different source types. It can be collected from participants, such as students or teachers directly (for example, assessment data collected with student monitoring systems), from administrative systems (e.g., national databases), or from online learning systems (interaction data). A small inventory of available data in the Netherlands (Veldkamp et al. 2017) identified ten different types of data sources, namely data from:

1. Student monitoring systems
2. National assessments at the end of primary and secondary education
3. National educational surveys
4. International surveys, such as PISA, TIMSS, and PIRLS
5. The Dutch Inspectorate of Education
6. Student evaluations
7. Online learning environments and MOOCs
8. Teacher meeting minutes
9. Formal administrations of, for example, sick-leave, or the absence of both teachers and students
10. All kinds of unstructured sources, such as teachers' notes, Wi-Fi tracking, and online search histories.

In education, big data involves a variety of data types about various levels of the educational systems, on complex and social interactions, stored at different places and in multiple systems, which need to be connected in order to be able to analyze processes taking place in education, and to improve education. The potential of big data for education has been increasingly recognized and knowledge of patterns in data can be used for improving education (Bongers/Jager/Te Velde 2015). However, a number of issues related to big data require attention, such as privacy and ethical issues, responsibility, availability and the quality of the data.

This chapter is based on a big data study conducted between November 2016 and February 2017 in the Netherlands (Veldkamp et al. 2017). In this study, the potential of big data in primary education, secondary education, vocational education, and higher education in the Netherlands was explored based on the views of various Dutch stakeholders. Insights from the literature were combined with these stakeholder opinions. We wanted to uncover (1) the purposes big data are being used for, (2) the challenges that the field of education faces when dealing with big data and (3) the opportunities that big data could offer.

2. Method⁶

Dutch schools have much autonomy, and many decisions are made at the level of the school (OECD 2010). The Dutch Government is only responsible for the general education policy, financial structures, admission requirements, and for the structure and objectives of the educational system (EP-Nuffic 2015). There is no central curriculum for primary or secondary education. Learning objectives are broadly formulated for the different stages and different tracks of the educational system. Only one national assessment at the end of primary education, and one at the end of secondary education exist (OECD 2008). This means that, within the boundaries that the national examinations set, schools can decide on their teaching and learning methods and curriculum design, including the subjects to be taught and the content of these subjects (Béguin/Ehren 2011; OECD 2008). As a consequence, the topic of big data in schools involves many stakeholders.

For this study, we contacted 31 institutions with different types of expertise in the field of big data. All of them were willing to participate. Each of the institutions selected the people that would be interviewed. In order to explore the opinions of Dutch experts and stakeholders on our topic, thirty-three interviews were conducted. The interviewees included individuals working at organizations that generate and manage data, scientists who conduct research on (big) data, policy makers, school staff, the business community, lawyers, experts on ethical issues, and experts on the technical storage and retrieval of data.

Based on issues identified in the literature and the opinions of experts we consulted, three different kinds of semi-structured interviews (focusing on the purposes of big data, the challenges, and opportunities) were developed that included sets of questions that matched with the role and the expertise of the respondents. We discussed the following set of topics:

- What is your vision on big data in education?
- What kinds of data do you manage?
- Is educational science ready for working with big data?
- What do you think about privacy and big data analytics?
- Would it be a good idea to organize a national comprehensive database for educational data that can be used for big data analytics?

With respect to each of these topics a set of sub-questions was predefined to obtain more detailed information. These sub-questions dealt with issues like expectations regarding the value of big data analytics, how data is managed, why respondents held certain opinions, etc. Other, more focused interviews

6 The methods of this study are summarized here; for more details: see Veldkamp et al. 2017.

were prepared for legal experts. We interviewed lawyers from three different universities, an intellectual property lawyer, and the Dutch Data Protection Authority. We asked these legal experts the following kind of questions. What is your vision on big data in education? Is current legislation adequate? What is your view on the ownership of educational data? What are your expectations regarding future developments? Finally, with ethicists from three Dutch universities, we discussed the following questions. Which ethical concerns do educational researchers have to take into account? Which common practices do you observe when it comes to educational data? What issues are researchers in the field of big data confronted with from an ethical point of view?

The interviews were recorded, transcribed and annotated. We counted how often respondents mentioned each topic. A complete overview of these counts can be found in Veldkamp et al. (2017). Next, we grouped these data into three main topics: Purposes, challenges, and opportunities.

3. Results

We will first present the results with respect to the purposes of big data in education according to the different stakeholders identified in our study. Next, various challenges of big data in education are summarized, and finally, various opportunities that were found in the literature and/or were mentioned by the respondents are presented. The results of this study are summarized in Figure 1, and will be further explained thereafter. Organizational, ethical, and social implications were placed at the top of the figure, as these will influence the use of technology, as well as the purposes for which big data can and cannot be used. Technological challenges and opportunities in turn, will also impact the use of big data for certain purposes. Finally, organizational and human capacity have been placed at the left side of the Figure, as this will influence legal and ethical challenges, the technological challenges and opportunities, as well as the actual use of big data. Now we will turn to the details of Figure 1.

Figure 1. Summary of results

Organizational and human capacity	Laws, ethics and social implications
<ul style="list-style-type: none"> • Lack of organizational structures, such as data infrastructure, leadership at different levels of the systems, collaboration in the use of big data, for example in professional learning communities ↔ investments in organizational infrastructure • Lack of expertise, experts and data literacy ↔ Working together in multidisciplinary teams, partnerships between stakeholders • Risk of bias and subjectiveness ↔ Investment in research and tools 	<ul style="list-style-type: none"> • Data protection laws • Privacy and anonymity in the age of big data? • Combining data and the risk of data leaks • Data trails and who has access to which data? • Future consequences and the ‘right to be forgotten’, the role of coincidence in the age of prediction • Data collection is not value free • Problems such as false negatives, untraceable decisions, profiling, labeling, stigmatization, discrimination, self-fulfilling prophecies, and true identity being replaced by digital identity • Does the importance of improving the quality of education for many outweigh the interests of individuals? • Possible increase of inequality in society
	Technological challenges and opportunities
	<ul style="list-style-type: none"> • Data availability: Do we have the right data available to answer our questions? Less data available on concepts difficult to measure. Not everything can be captured in data ↔ More and more data available, growing exponentially. More tools available for data analysis and use • Risk of goal displacement ↔ Use of data to provide new insights needed to adjust education to the needs of the students
	Purposes
	<ul style="list-style-type: none"> • Improving the quality of decision making • Goal setting • Monitoring • Planning and scheduling • Selection • Curriculum design • Identifying and solving problems: improving the quality of programs, leading, teaching and learning processes • Research purposes

4. Purposes

Within schools, different actors with different roles can be distinguished. Each of them might use big data for different purposes, but the overall purpose can be described as the improvement of the quality of decision making. Leaders at all levels of the system can use big data for benchmarking and profiling (Veldkamp et al. 2017). It can demonstrate how well (or poorly) an organization is performing compared to other similar organizations. This can help leaders in setting the goals for their organization, for example, with respect to educational quality indicators, set by themselves or the government. They can use big data for setting goals at the different levels of the system (Romero/Ventura 2010). What are important goals at the level of the school, what are important classroom level goals, and what are student level goals? These goals can pertain to cognitive goals, such as (aggregated) student achievement results, but also to non-cognitive goals, such as well-being and the socio-emotional development of students.

School leaders can also use big data for monitoring purposes, for example, to monitor to which extent the goals set are being accomplished. For this purpose, dashboards are becoming increasingly popular in educational institutions. User friendly software tools, such as Power BI or Tableau, are available and facilitate the development and support of personalized dashboards that bring together information from various sources and administrative systems. They present the data graphically in one or a few overviews and provide a toolkit for basic analyses. This way, leaders can monitor budget, personnel, sick leave, presence and performance on a continuous basis (Veldkamp et al. 2017).

Based on continuous monitoring, leaders also can identify and solve problems. Big data can assist in the analysis of problems and the causes of these problems (Manyika et al. 2011). The problem can be defined in a measurable manner (e.g., these schools or students are underperforming, their average score is x, and our goal is y), and big data can be used to identify the root causes of underperformance, which can help in solving these problems. In this sense, big data can also assist managers in decision making to improve the quality of an educational organization (Manyika et al. 2011). This may also pertain to decisions with regard to how to use (human) resources and materials (Liñán/Pérez 2015; Romero/Ventura 2010). For example, big data can help in identifying professional development opportunities needed in the organization.

Teachers can also make use of big data for several purposes. Similar to leaders, they can use big data for goal setting at the classroom level and the individual student level. Big data can also help teachers to monitor the quality of teaching and learning in their classrooms. Based on big data, they can

evaluate the quality of their own instruction, monitor student performance, map out the progress and learning of individual students, and check to which extent goals are being accomplished (Liñán/Pérez 2015; Romero/Ventura 2010). Big data enables teachers to identify learning problems at an earlier stage and to develop appropriate follow-up actions accordingly. Teachers can take both proactive and remediation measures (Dede 2016; Romero/Ventura 2010). For example, based on achievement data, classroom observation data, and motivation data, teachers may decide that they need to differentiate more in their classroom.

Similarly to leaders and teachers, purposes of big data for *students* also relate to goal setting, identifying (learning) problems, and solving these problems. Based on different kinds of data available about students' performance (coming from formal tests, teacher observations, logfiles, papers, or their own reflection on learning), big data analytics can be applied to provide the students with personalized feedback about where they are in their learning processes (compared their goals and current performance level). The students can analyze their own learning, compare their learning to the learning of other students, and based on the analyses, identify possible problems, and take decisions on next steps (Liñán/Pérez 2015; Romero/Ventura 2010). Big data can even be used to recommend activities, resources, books, assignments, and/or courses that may be helpful for (improving) the learning processes of students (Liñán/Pérez 2015; Romero/Ventura 2010). In the interviews it was also mentioned that big data offers the possibility to give more personalized lessons. Weaker students can receive extra support. Also, individual students can have more autonomy and more ownership over their own learning process. Finally, big data analytics can provide students with feedback on the most suited learning track after primary education (e.g., a pre-vocational track or a pre-university track in secondary education) (Veldkamp et al. 2017).

In the interviews, *researchers* stressed the importance of having access to data and they drew attention to the possibilities of creating a national database of educational data and the opportunities this would offer for them. The time saved by researchers when they no longer have to collect the data themselves was seen as a huge advantage of such a database. In terms of its content, the possibilities for research into early diagnosis and for establishing links with data from other sources, such as data on health, developments in the locality/region or data from the parents were mentioned (Veldkamp et al. 2017).

Finally, big data can be used by *course providers, training institutes, schools, colleges, and universities* to make better decisions (Romero/Ventura 2010). Using big data, recommendations can be made for specific courses for specific (groups of) students. Big data can also be used to predict what is needed to improve student learning. It can be used to reduce the number of students dropping out of school, in a cost-effective way. Big data can be used for planning and scheduling and for selection, both at the intake and the

progression of students into different directions within a study program. Finally, big data are important for quality improvement, for example, by evaluating and improving educational programs and teacher performance (Kane/Rockoff/Staiger 2008; Romero/Ventura 2010).

In the current, exploratory, study the focus was on actors who are active in schools. Different stakeholders, such as publishers, the Ministry of Education or the Dutch Inspectorate of Education, could also use big data for other purposes related to education. Some of the purposes might strengthen each other. Both teachers and students are focused on optimizing learning results. Purposes such as the evaluation of students' learning progress by teachers, and students obtaining insight into their own learning progress, both benefit from assessment of student performance. Other purposes might conflict with each other. For example, the purpose of the efficient planning of materials by leaders, and the purpose of developing new materials by publishers might conflict with each other. Moreover, answering the various questions requires different types of data and different analyses, each of which comes with its own questions about, for example, the availability and quality of data (Veldkamp et al. 2017). Finally, even though the actors might intend to use big data for all of these purposes, this does not mean that this usage can be realized yet.

5. Challenges for big data

The field of big data analytics needs to address several challenges, before it can fulfill its potential of educational improvement, as reflected in the following statement by Valerie Strauss made in 2016:

Data from PISA, for example, suggests that the 'highest performing education systems are those that combine quality with equity'. What we need to keep in mind is that this statement expresses that student achievement (quality) and equity (strength of the relationship between student achievement and family background) of these outcomes in education systems happens at the same time. It doesn't mean, however, that one variable would cause the other. Correlation is a valuable part of evidence in education policy-making but it must be proven to be real and then all possible causative relationships must be carefully explored (Valerie Strauss, 'Big data' was supposed to fix education. It didn't. It's time for 'small data.' From: Washington Post; May 9, 2016)

In this section, we distinguish between legal challenges, ethical and social challenges, technological challenges, and human capacity challenges.

With regard to *legal challenges*, our society is facing much needed but complex data protection laws (Boyd/Crawford 2012; Enyon 2013; Ferguson 2012; Liñán/Pérez 2015; Manyika et al. 2011; Piety 2013; Veldkamp et al.

2017), such as the Family Educational Rights and Privacy Act in the USA, and the General Data Protection Regulation in the EU. Questions that need to be answered are questions such as: Which data may and may not be combined? Who owns the data? Who can access and use the data? For which goals may which data be used? When is what type of informed consent needed? How do we deal with privacy regulations? For example, concerning the question which data may and may not be combined, it is important to ensure anonymity of individuals. However, connecting different data sets makes it possible to identify individuals. Therefore, the question is, even with all the data protection laws currently in place: Does anonymity exist in the age of big data? Some of the interviewed lawyers mentioned that educational institutes often do not know how important data protection and privacy are, and how important it is that they make appropriate arrangements. On the one hand, educational institutions want to protect the privacy of their students; on the other hand, there still is a great deal of confusion about what is allowed and what is not allowed, and about privacy legislation in schools.

Next to legal challenges, *several ethical and social challenges* exist. For example, it is almost impossible not to leave a data trail, and these data trails include personal information (Franzke 2016). Also, it is very difficult to envision future consequences (ibid.), for example, of social media and its data distribution. Important is “the right to be forgotten” (Weber 2011). For example, low grades in high school should not be used years later. Another question to ask here is: What is the role of coincidence in the age of prediction (Enyon 2013)? It is important to realize that collecting and analyzing data is never value free. Even in the stage of developing the measures to collect data, decisions that are not value free have to be taken already (ibid.).

Big data also brings about certain risks and problems, such as the problem of false negatives (O’Neil 2016); the risk of profiling, labeling, stigmatization, discrimination, and self-fulfilling prophecy (Veldkamp et al. 2017). For example, it becomes much easier to identify good and weak systems/schools/teachers/students (Enyon 2013). This comes with the risk of excluding weaker students that negatively impact reaching a certain benchmark (Piety 2013). Moreover, a lack of transparency exists, it is not always clear what decisions are based on (e.g., algorithms used are not easy to understand or shared). As stated by Wang (2020) sometimes even those who have developed the algorithm do not fully know how a decision has been made, especially since the better the algorithm is, the more difficult it often is to understand (Courtland 2018).

When the topic of ethical and social challenges was discussed with university students, several respondents raised another issue. Even though big data analytics might provide many opportunities education can benefit from, it also carries the risk that teachers may focus too much on data. Students fear

that teachers might solely rely on data rather than on their own observations, especially in university courses with hundreds of participants, or in the case of online learning, and that their true identity could be replaced by a digital identity based on their data trail. They indicated that this feels alienating and undesirable (Veldkamp et al. 2017).

On the other hand, several respondents mentioned that ethical risks could be overstated. Especially with respect to scientific research, the importance of improving the quality of education in general might outweigh the interest of individuals, according to some of the respondents. An important question to ask here is: Are we identifying weak teachers to protect students, or should we protect teachers from unfair evaluations (Piety 2013)? Whereas respondents were more critical concerning commercial companies, for them, strict guidelines and policies were thought to be much more important.

Finally, big data also comes with the risk of increased inequality in society. Data will be more available about some people (e.g., people with better Wi-Fi access, students with access to expensive online practice programs) than about others, and some people will have better access to (use) these data and are more likely to benefit from it than others (Enyon 2013; Boyd/Crawford 2012). A power gap can arise between those who generate the data (students and teachers) and the people who analyze the data (Veldkamp et al. 2017).

The *technological challenges* we have to deal with before we can unlock the full potential of big data are numerous. A first set of challenges deals with the availability of data, the questions we want to ask, and the question whether available data are the right data for answering our questions. Currently, school improvement processes often start with data instead of with clear questions and goals. Vast amounts of data have been collected in schools for many years, rather than to answer specific questions. It is crucial to formulate clear questions and goals, and then collect the necessary data accordingly, especially since new questions and goals are constantly arising in areas that may be assessed less frequently (e.g., well-being, citizenship, self-regulation) (Schildkamp 2019). However, currently, it often works the other way around. The data that are available influence the questions being asked (Enyon 2013). Also, there is less data on concepts difficult to measure (Kitchin 2013), and not everything can be captured in data (Enyon 2013; Kane 2008; Piety 2013). When we adapt our research questions to the data that can be worked with, the availability of the data becomes the main issue, rather than our research questions. This risk is also referred to as goal displacement (Lavertu 2014).

Who owns the data also influences the accessibility of data for different purposes. If one is not convinced of the usefulness of big data for a specific purpose then the willingness to make data available might be limited. Another important aspect of making data available is whether the parties involved

have insight into what happens with the data. Sometimes, data are only made available under strict and specific conditions. Several stakeholders indicated that they were willing to make data available and that they saw the advantages of big data research, but only under certain conditions related to anonymization, granularity, access and limited use (Veldkamp et al. 2017). Since all data use has to follow the General Data Protection Regulations of the European Union, data can only be used for goals that have been defined before the data was collected (ibid.).

A next set of technological challenges deals with the accessibility and the quality of the data. It is difficult to connect different data sets (i.e., different data silo's), sometimes it is even impossible to retrieve the data needed from the different systems, data formats are not always aligned, and pre-processing of data costs a lot of time, which makes it also an expensive process (Enyon 2013; Kane 2008; Piety 2013; Veldkamp et al. 2017). Restrictions with regard to the quality of the data available also exist especially for unstructured data. The restrictions concern bias in the data, missing data, measurement errors, and problems with the representativeness (Boyd/Crawford 2012; Gibson/Webb 2015; Piety 2013). Moreover, big data entails combining different data sets, but combining data sets with errors increases the number of errors (Bollier 2010; Boyd/Crawford 2012).

A final set of challenges that we have identified evolves around human and organizational capacity. Technology opportunities seem to grow faster than human and organizational capacity and capabilities. Organizational structures, such as infrastructure, leadership structures, and collaboration structures (e.g., in the form of professional learning communities) are sometimes also lacking (Piety 2019). Moreover, there is not only a lack of expertise among teachers, but also a lack of experts who can assist them in the use of big data. Big data literacy is needed: How can we collect, analyze, interpret, and use big data to improve decision making in education (Enyon 2013; Lavertu 2014; Liñán/Pérez 2015; Manyika et al. 2011)? Moreover, some leaders and teachers are critical when it comes to the (benefits of) the use of big data. Furthermore, in the interpretation of big data, both bias and subjectivity play a role and the context in which the data were collected should not be forgotten (Boyd/Crawford 2012; Ozga 2009).

6. Big data opportunities

Big data offers various opportunities for education. Stakeholders see many opportunities and advantages related to technology, human capacity and capability, and real-time interventions. Their opinions about opportunities are substantiated by the literature.

Rapid developments in the field of *technology* provides us with different kinds of opportunities. Firstly, more and more data become available (e.g., social media, online learning environments, MOOCs) (Piety 2013; Williamson 2016). Instruments register and collect data about their users. These data could provide insight into user preferences, and use patterns and ways of learning, which might empower personalized learning. Moreover, new technologies are becoming available for big data (use), such as online tools to track learning over time, learning analytics, real time analysis (Piety 2013; Ferguson 2012), data mining (e.g., text mining, audio mining, video mining) and data analysis (e.g., machine learning, model training and testing) (Fayyad/Piatetsky-Shapiro/Smyth 1996; Romero/Ventura/De Bra 2004). This can be used not only to assess students' outcomes (summative testing), but also to promote student learning (formative testing). Besides, at different levels of the systems investments are being made in data infrastructures (Veldkamp et al. 2017). Wizard tools for practitioners are being developed (Romero/Ventura 2010). Based on all these developments, it has become possible to obtain detailed insights into learning, and to adjust education to the needs of the students (Veldkamp et al. 2017).

Also, in terms of *human capacity and capability*, the opportunities are growing. People are working together in multidisciplinary teams (Enyon 2013), and partnerships are created between different stakeholders, such as school-university partnerships (Veldkamp et al. 2017). Data coaches and training have become available (Dede 2016), for example in the form of data teams (Schildkamp et al. 2018). The investment in organizational infrastructure and research on data science by the government is growing, so that big data can be used as a tool in the decision making process (Piety 2019; Veldkamp et al. 2017), for example. Interest in big data analytics tools also increases among teachers and school leaders. The availability of user-friendly tools and visualization software makes big data analytics feasible for statisticians, computer scientists and engineers and many others.

Finally, by *real-time* data analysis, it is possible to predict which students are at risk and direct interventions can be applied. This enables teachers to adapt their teaching to the needs of the pupils at the right time (Veldkamp et al. 2017).

7. Big data paradoxes

The overarching purpose of the use of big data in the field of education is to predict future performance and to identify problems related to learning and development (Hrabowksi/Suess/Fritz 2011). Different stakeholders obviously have their own specific purposes for big data use in education. Roughly

speaking, a distinction can be made between using big data to (1) monitor and gain more insight into certain processes, including disproving myths and assumptions, (2) predicting (learning outcomes, study success, dropout, etc.) and (3) taking measures to improve education. The added value of big data, according to the majority of respondents, seems to lie in the new insights that big data can provide.

However, there are several challenges we need to face, as is discussed in this section. Some stakeholders might feel that the added value might not outweigh the investments needed. Teachers, for example, tend to be quite critical when it comes to the use of big data. They want to understand the underlying methods and models before they are open to implement the new insights in their classrooms. Unfortunately, it is quite difficult and time consuming to obtain the knowledge they need to understand the systems. Due to high work demands and pressure, they lack time. Therefore, teachers might not fully benefit from the possibilities of big data analysis. For the various stakeholders, different, or even conflicting, considerations play a role when it comes to using big data analysis for improving education. Based on the interviews with experts and stakeholders, Veldkamp et al. (2017) identified a number of big data paradoxes, combining the challenges and opportunities:

a) *Privacy paradox: Privacy protection versus combining different data sets*

From a legal point of view, there is an increasing focus on privacy protection measures. Specialists point out that the merging of data means that individuals can be traced, even if anonymization or pseudo-anonymization is applied. This is still unknown to many users. The question remains how a database can be set up in such a way that people's privacy is sufficiently protected, but that still facilitates the linking of different datasets, so that new insights can be obtained. Several technical solutions have been suggested, like the use of Chinese walls to separate information sources, keys, and encryptions. But it is an open question if these tools provide sufficient protection and if they are trusted by the general public.

b) *Clustering paradox: Combining data versus data leaks*

When educational data are combined in a single database, data are becoming more accessible and security can be improved. However, there are some risks involved, like mistakes in linking data. A central database entails a higher risk of data leaks.

c) *Individual context paradox: Disconnect context from the data, to ensure the privacy of individuals versus context needed to interpret data*

Analyzing big data without taking into account the specific context of collecting data offers many advantages. Privacy is more protected, since it is much harder to identify the data sources, and data can be aggregated at higher levels, which increases the potential of big data analytics. On

the other hand, for a correct interpretation of the data, the context is needed to prevent biases.

- d) *Give and take paradox: Decisively asking for data versus being hesitant to provide data to others*

Many respondents are interested in data that can be used to answer their questions; however, they are reluctant to share their own data for answering other people's questions.

- e) *Technology capacity paradox: Fast growing opportunities for collecting and analyzing data versus slowly growing human capacity*

The field of big data collection and analysis is developing rapidly and leads to increasing possibilities. New platforms and tools are becoming available. On the other hand, the capacity of people seems to be lagging behind.

8. Conclusions

Our society is inundated with data, leading to several challenges and opportunities with regard to the use of big data, as described in this chapter. Legal rules and regulations are becoming more and more important. The laws created in this area, such as the Family Educational Rights and Privacy Act in the USA, and the General Data Protection Regulation in the EU, are clearly needed, but it also needs to be investigated to what extent these new laws hinder the opportunities that big data analytics can provide. Policy makers may need to devise different rules for different stakeholders. They may need to distinguish between rules with regard to the use of big data by school staff, by researchers, and by commercial organizations. Laws and policies need to be clear about who owns which data, and how the safety of data storage and privacy of individuals is ensured. Based on our study, we also recommend to policy makers to implement “the right to be forgotten” in these laws. Once a student leaves school, the data about him/her should be anonymized and/or at only available at an aggregated level.

Based on this study, we recommend to researchers that with each big data study and publication of the outcomes, the quality of the original data (accuracy, consistence, representativeness) should be reported with regard to technology. Also, the data analysis techniques used should be reported, as well as information on the context in which the data were collected. Finally, a protocol for how to store the data used in a standardized manner needs to be developed, and it might be explored whether setting up one (inter)national database is desirable and feasible. Moreover, with regard to technology several research questions exist, like which aspects of the big data use process can be taken over by technologies (such as machine learning and artificial

intelligence), and which aspects still require human decision making, and thus human capacity.

Human capacity development is crucial in the field of big data. We need to train more people in data science. This does not only pertain to the technical side of big data, but also to how to discuss big data with a lay audience, and how to train school staff in the interpretation and use of big data analyses. It would also help to develop tools, which enable big data analysis by (trained) school staff, who can then look at and use the data in their own context. As stated above, technology can take over part of the collection, and analysis of big data, but human decision-making is needed with regard to which decisions to take and implement, based on the analysis. Perhaps a good start would be to develop big data professional development programs for school boards and educational leaders.

The use of big data requires both technology (high tech) and professional development (human touch). This requires the collaboration of different stakeholders. As stated by Wang (2020) this includes those who have in-depth knowledge of the data (e.g., data providers) in their specific context (e.g., school leaders), the people who will be influenced by the decisions (e.g., teachers, parents, students, communities), and those who develop the big data analytics and algorithms. Furthermore, this means that to be able to make use of big data, it is essential to understand the context of the data, prioritize people, and focus on student interests and needs (Ibid). The key question for scientific research to answer is: How can big data analytics contribute to education? It is clear that big data can be used for a large number of purposes by different actors in the field of education. The list of possible purposes seems almost endless. The availability of tools and software provides many opportunities to realize them. Therefore, the future of big data analytics in education seems very bright. As Piety (2019: 414) states: “These new techniques will support different kinds of understandings about instructional and student processes across the praxis landscape”. It should be mentioned though that many challenges still exist. The various paradoxes mentioned exemplify that the field of educational big data analytics is still in its infancy, and that the development of human capacity is urgently needed. In our opinion this field would benefit much from research that illustrates how big data contributes to better education in such a way that it accounts for transparent, comprehensible and replicable data use.

References

- Béguin, Anton/Ehren, Melanie Catharina Margaretha (2011): Aspects of accountability and assessment in the Netherlands. In: *Zeitschrift für Erziehungswissenschaft*, 14, pp. 25-36.
- Bollier, David (2010): *The Promise and Peril of Big Data*. The Aspen Institute.
- Bongers, Frank/Jager, Cor-Jan/Te Velde, Robbin (2015): *Big Data in onderwijs en wetenschap*. Utercht: Dialogic.
- Boyd, Danah/Crawford, Kate (2012): Critical questions for big Data. Provocations for a cultural, technological, and scholarly phenomenon, information. In: *Communication & Society*, 15, pp. 662-679.
- Bijlsma, Hannah J.E./Visscher, Adrie J./Dobbelaer, Marjoleine J./Veldkamp, Bernard P. (2019): Does smartphone-assisted student feedback affect teachers' teaching quality? In: *Technology, pedagogy and education*, 28, 2, pp. 217-236.
- Courtland, Rachel (2018): Bias detectives: The researchers striving to make algorithms fair. In: *Nature*, 558, pp. 357-357.
- Dede, Christopher J. (2016): Next steps for "Big Data" in education: Utilizing data-intensive research. In: *Educational Technology LVI*, 2, pp. 37-42.
- Daniel, Ben K. (2015): Big Data and analytics in higher education: Opportunities and challenges. In: *British journal of educational technology*, 46, 5, pp. 904-920.
- EP-Nuffic (2015): *Education System the Netherlands; the Dutch Education system described*. The Hague: Nuffic.
<https://www.epnuffic.nl/en/publications/education-system-the-netherlands.pdf>
- Eynon, Rebecca (2013): The rise of big data: what does it mean for education, technology, and media research? In: *Learning, Media and Technology*, 38, pp. 237-240.
- Fayyad, Usama/Piatetsky-Shapiro Gregory/Smyth, Padhraic (1996): From Data Mining to Knowledge Discovery in Databases. In: *American Association for Artificial Intelligence*, 17, pp. 37-54.
- Ferguson, Rebecca (2012): Learning analytics: drivers, developments and challenges. In: *International Journal of Technology Enhanced Learning*, 4, pp. 304-317.
- Franzke, Aline (2016): *Big Data Ethicist – What will the role of the ethicist be in advising governments in the field of big data?* Master's thesis. Utrecht: Utrecht University.
- Gibson, David C./Webb, Mary E. (2015): Data science in educational assessment. In: *Education and Information Technologies*, 20, pp. 697-713.
- Hrabowski, Freeman A./Suess, Jack/Fritz, John (2011): Assessment and analytics in institutional transformation. In: *Educause Review*, 46, 5, pp. 14-16.
- Kane, Thomas J./Rockoff, Jonah E./Staiger, Douglas O. (2008): What does certification tell us about teacher effectiveness? Evidence from New York City. In: *Economics of Education Review*, 27, pp. 615-631.
- Kitchin, Rob (2013): Big data and human geography: Opportunities, challenges and risks. In: *Dialogues in Human Geography*, 3, 3, pp. 262-267.
- Laney, Doug (2001): *3-D Data Management: Controlling Data Volume, Variety and Velocity*. META Group Inc, File 949, pp. 1-4.

- Lavertu, Stéphane (2014): Performance Measurement: A Balanced Solution to Inequities in Election Administration? In: *Election Law Journal*, 13, 4, pp. 550-554.
- Liñán, Laura Calvet/Pérez, Ángel Alejandro Juan (2015): Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. In: *International Journal of Educational Technology in Higher Education*, 12, 3, pp. 98-112.
- O'Neill, Cathy (2016): *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing Group.
- Manyika, James/Chui, Michael/Brown, Brad/Bughin, Jacques/Dobbs, Richard/Roxburgh, Charles/Byers, Angela Hung (2011): *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- OECD (2008): *Education at a glance 2008: OECD indicators*. Paris: Organisation for Economic Co-operation and Development (OECD).
- OECD (2010): *Education at a glance 2010: OECD indicators*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Ozga, Jenny (2009): Governing education through data in England: From regulation to self-evaluation. In: *Journal of Education Policy*, 24, 2, pp. 149-162.
- Picciano, Anthony G. (2012): The evolution of big data and learning analytics in American higher education. In: *Journal of Asynchronous Learning Networks*, 16, 3, pp. 9-20.
- Piety, Philip J. (2013): *Assessing the educational data movement*. New York: Teachers College Press.
- Piety, Philip J. (2019): Components, Infrastructures, and Capacity: The Quest for the Impact of Actionable Data Use on P-20 Educator Practice. In: *Review of Research in Education*, 43, 1, pp. 394-421.
- Romero, Cristóbal/Ventura, Sebastián/De Bra, Paul (2004): Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction. In: *The Journal of Personalization Research*, 14, pp. 425-464.
- Romero, Cristóbal/Ventura, Sebastián (2010): Educational Data Mining: A Review of the State of the Art. In: *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 40, pp. 601-625.
- Schildkamp, Kim/Handelzalts, Adam/Poortman, Cindy Louise/Leusink, Hanadij/Meerdink, Marije/Smit, Maaikje/Ebbeler, J./Hubers, Mireille Desirée (2018): *The Data Team™ Procedure: A Systematic Approach to School Improvement*. Dordrecht: Springer.
- Schildkamp, Kim (2019): Data-based decision-making for school improvement: Research insights and gaps. In: *Educational Research*, 61, 3, pp. 257-273.
- Veldkamp, Bernard P./Schildkamp, Kim/Keijsers, Merel/Visscher, Adrie/de Jong, Ton (2017): *Verkenning data-gedreven onderwijsonderzoek in Nederland*. Enschede, The Netherlands: University of Twente.
- Weber, Rolf H. (2011): The right to be forgotten. More than a Pandora's Box? In: *Journal of Intellectual Property, Information Technology and E-Commerce (JIPITEC)*, 2(2), pp. 120-130.
- Williamson, Ben (2016): Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. In: *Journal of Education Policy*, 31, pp. 123-141.

Using Digital Data to Support Teaching Practice – quop: An Effective Web-Based Approach to Monitor Student Learning Progress in Reading and Mathematics in Entire Classrooms

Elmar Souvignier¹, Natalie Förster², Karin Hebbecker³ and Birgit Schütze⁴

1. Introduction

“Is my instruction beneficial for my students?” “How can I adapt my instruction to students’ individual needs?” One key way to answer these questions – and to effective instruction in general – is to obtain objective, reliable, and valid measurements about student achievement. Moreover, such achievement measurements must be made early and repeatedly over the course of the learning process if performance assessments are to be used for educational decision-making. As these considerations fall in line with the theoretical frameworks of formative assessment (e.g., Black/Wiliam 1998) and data-based decision-making (e.g., Mandinach 2012), it follows that providing teachers with assessment information about students’ levels of achievement and about their learning progress is a promising approach to improve instruction – and, thereby, students’ learning.

Several international reviews confirm the effectiveness of using student achievement assessments for instructional decision-making (e.g., Black/Wiliam 1998; Kingston/Nash 2010; Stecker/Fuchs/Fuchs 2005). These reviews, however, also reveal that the specific ways formative assessments are realized largely differ from each other, and that the effect sizes of different approaches cover a wide range. Further, most of the research has been conducted with low achieving students and within settings of individual or small-group instruction.

One concept that has been developed since the 1980s is the approach of progress monitoring: Students’ achievements are measured with short,

1 Elmar Souvignier is Professor of Diagnostics and Evaluation in Schools at the Institute for Psychology in Education and Teaching at the University of Münster. Email: elmar.souvignier@uni-muenster.de

2 Natalie Förster is Postdoctoral Researcher at the Institute for Psychology in Education and Teaching at the University of Münster. E-Mail: natalie.foerster@uni-muenster.de

3 Karin Hebbecker is Postdoctoral Researcher at the Institute for Psychology in Education and Teaching at the University of Münster. E-Mail: karin.hebbecker@uni-muenster.de

4 Birgit Schütze is Postdoctoral Researcher at the Institute for Psychology in Education and Teaching at the University of Münster. E-Mail: birgit.schuetze@uni-muenster.de

parallel forms of tests that provide feedback for teachers on the effectiveness of the current instruction (Deno 1985). Tests on basic curricular abilities such as reading or math skills are applied over short intervals (e.g., weekly) to allow for immediate adjustments in instruction. However, given that executing, scoring, and documenting such frequent assessments is time consuming, researchers have often claimed that providing digital forms of formative assessments might support the implementation of progress monitoring (Fuchs 2004; Stecker et al. 2005). Beyond facilitating the use of repeated assessments, computer-based concepts include advantages such as being able to automatically highlight students who show little progress, provide teachers with suggestions on individualized instruction, and enable progress monitoring to be applied to all students in a classroom, which would not be feasible using paper-pencil forms of assessment.

The progress-monitoring approach entails several requirements. First, technical adequacy (reliability, validity) of the tests needs to be high. Second, the tests need to be equivalent to one another and sensitive to student progress even over short time intervals to allow for conclusions on student progress. Third, tests also have to be highly practical, which means that they have to be short and that they can be applied routinely. Finally, with respect to the goal of instructional decision-making, the results of the tests should be easily interpretable. This catalogue of requirements for measuring learning progress illustrates that the concrete demands on this approach are high – and conflicting. For example, the psychometric demands usually require sophisticated test forms, but the fact that tests must be highly practical depends on short measures that are easy to interpret.

Within the context of progress monitoring, the concept that has reached especially high visibility is curriculum-based measurement (CBM; Deno 1985), for which the key to success, as stated by Jenkins and Fuchs (2012), turned out to be “the idea of simplicity” (p. 7). In this light, it seems reasonable that to facilitate steps like test administration, scoring, documentation of results, and providing help in interpreting results, computer-based approaches should be used. Such approaches would be especially useful when the goal is to implement progress monitoring for all students in general education.

In Section 1, we describe the web-based learning progress monitoring system *quop*, which has been developed in Germany to provide teachers with a practical approach for learning progress assessment (LPA) that fulfills both requirements of technical adequacy as well as simplicity. As *quop* was extensively assessed during its development, Section 2 will summarize research on its technical adequacy, slope information, effects of providing teachers with the system, and approaches to foster assessment-based differentiated instruction in general education.

2. quop – an approach for learning progress assessment (LPA)

quop is a web-based approach for learning progress assessment that provides test concepts for reading and mathematics for grades one to six. It is designed to monitor the individual learning progress of all students in a classroom in regular education. As it is web-based, it enables a feasible, economic, and automated evaluation and documentation of students' outcomes, and the program is updated on an ongoing basis. For each grade and each domain, the system provides eight parallel tests over the course of a school year. Each test is available online for a period of three weeks. The time required to complete each test is ten to fifteen minutes. All test contents are based on the German national standards for reading and mathematics.

Reading tests in first and second grade assess the efficiency of reading processes for syllables, words, sentences, and short texts. In third and fourth grade, the tests assess reading accuracy, reading fluency, as well as text-based and knowledge-based reading comprehension using fictional and non-fictional texts. Task formats vary depending on the construct assessed and include, for example, verification tasks, maze tasks, and multiple-choice questions.

Mathematics tests assess precursors (e.g., quantity discrimination and identifying a number on a number line) and curricular competencies in the domains of numbers and operations (e.g., basic arithmetic operations), geometry, and calculating with units. Multiple-choice response formats are used. Table 1 gives an overview of all test contents and competencies in reading and mathematics in the different grades.

A prerequisite for the use of quop is a web-enabled computer with a standard internet browser. Depending on the number of computers available, students usually finish a test during self-study periods or in group sessions. Before the first test, students receive instructions from their teacher and from the computer. Moreover, they complete a short tutorial to become familiar with the test format for the different tasks and the testing procedure to be able to work independently on the tests. After completing a test, students receive computer-generated feedback on their performance for the current and former tests in the form of a graph (reading) or a table (mathematics).

Table 1. Test contents and competencies for reading and mathematics in the different grades

	Reading	Mathematics
Grade 1	Efficiency of reading processes for... syllables words sentences	Basic precursors (e.g., quantity comparison) Advanced precursors (e.g., number line) Number and operations (esp., basic arithmetic operations)
Grade 2	Efficiency of reading processes for... words sentences short texts	Basic precursors (<i>e.g., numerical comparison</i>) Advanced precursors (<i>e.g., number line</i>) Number and operations (<i>esp., basic arithmetic operations</i>)
Grade 3	Reading accuracy Reading speed Text-based reading comprehension Knowledge-based reading comprehension	Number and operations (<i>esp., basic arithmetic operations</i>) Geometry (<i>e.g., rotation of forms</i>) Calculating with units
Grade 4	Reading accuracy Reading speed Text-based reading comprehension Knowledge-based reading comprehension	Number and operations (<i>e.g., basic arithmetic operations</i>) Geometry (<i>e.g., axial symmetry</i>) Calculating with units
Grade 5	Reading fluency	Number and operations (<i>esp., calculating with natural numbers</i>) Geometry (<i>e.g. geometric bodies</i>) Calculating with units
Grade 6	Reading fluency	Number and operations (<i>esp., calculating with fractions & decimal numbers</i>) Geometry (<i>e.g., angle</i>)

Teachers have access to a teacher platform via a personal log-in, where they can obtain the results both at the class and the student level immediately after the test is finished. The results for the different competencies (e.g., reading accuracy, reading speed, text-based reading comprehension, knowledge-based reading comprehension) are presented in a table. Additionally, student growth is visualized in a graph. Teachers have the opportunity to display reference values in the form of means and standard deviations based on the sample of all students who have ever worked on the same test. To help teachers identify especially high- and low-performing students, results that are more than one standard deviation above or below the average result are highlighted in the table.

3. Research on LPA with quop

In line with Lynn Fuchs' (2004) suggestions for programmatic research on curriculum-based measurement, the development of quop was evaluated regarding three stages of research: First, technical adequacy of the tests was investigated. Second, features of slope were analyzed. Third, instructional utility was evaluated. The results from the following four paragraphs have been found in a series of studies that we conducted during the past ten years. The research program was run in German elementary schools with approximately 10.000 students from grades 1-4.

3.1 Technical adequacy of the tests

Studies on the psychometric properties of the different LPA test series in quop have addressed main test criteria like reliability and validity of the tests, but they have also investigated their equivalence and sensitivity to student progress, their ease of administration and utility, as well as their fairness. Using different designs and statistical analyses, the collection of studies covers reading and mathematics in all grades of elementary school. Results reveal that the different test series use reliable tests, with high internal consistencies usually exceeding values of $\alpha > .80$. Likewise, delayed alternate-form reliabilities are sufficient and exceed $r > .60$ in reading and $r > .70$ in mathematics (Förster/Souvignier 2011; Salaschek/Souvignier 2013; 2014). Correlations between the LPA test scores and standardized achievement tests, intelligence tests, and teacher ratings of student performance proved the convergent and divergent validity of the tests.

Taking into account the conceptual problems of evaluating test equivalence by means of alternate-form reliability, test equivalence was

studied using analyses of measurement invariance (Förster/Souvignier 2014b) and by comparing the equivalence of the test information functions (TIFs) of the different tests (Förster/Kuhn/Souvignier 2015). Results showed that the measurement models of the fourth-grade reading tests were strongly invariant over time and that TIFs showed no significant differences, with the exception of reading comprehension in one test. Using repeated measures analysis of variance, latent growth curve modeling, and latent difference score modeling, tests in reading and mathematics were found to be sensitive to student progress (Förster/Souvignier 2011, 2014b; Salaschek/Souvignier 2013, 2014).

The usability and feasibility of quop was addressed in different studies using teacher and student ratings. Moreover, the studies also considered the number of omitted LPA tests within one school year as an objective criterion. Teachers rated quop to be easy to administer, worth its effort, and to be a beneficial tool that they intend to continuously use (Förster/Souvignier 2015; Salaschek/Souvignier 2013, 2014). They observed that the students have fun when completing the tests and reported that students as young as grade one were able to conduct the tests independently. These judgments are mirrored by the students' ratings, who reported that they liked the tests and were looking forward to using the tests next year (Salaschek 2013). In different studies, the number of omitted LPA tests within one school year was found to be low; more than 90% of the students completed at least six out of eight tests during the school year. Thus, information about learning progress was available to teachers for most of the students at most points of measurement (Förster/Souvignier 2015; Förster/Kawohl/Souvignier 2018).

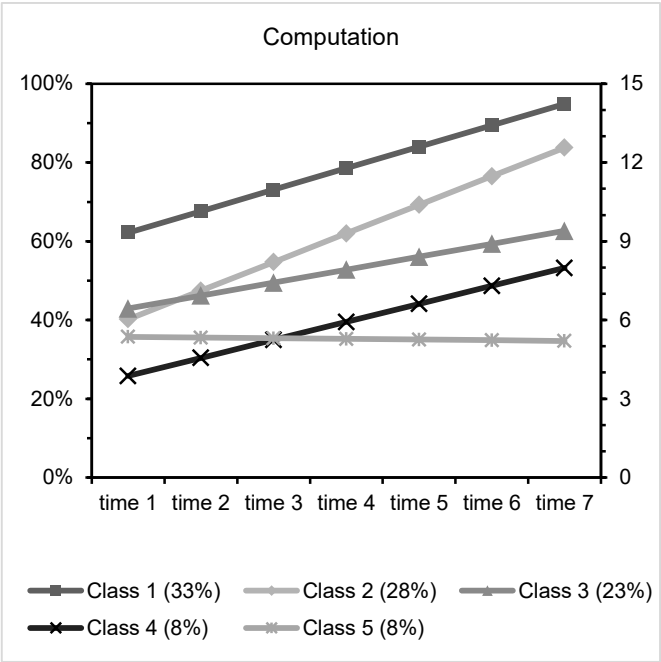
Ongoing research deals with the question of test fairness for boys and girls, for students with and without migration backgrounds, and for students with special needs using analyses of differential item functioning. Preliminary findings indicate no systematic discrimination of any student group.

3.2 Slope of students' progress

In a sample of 153 German first-grade students, Salaschek, Zeuch, and Souvignier (2014) conducted a study with quop to examine mathematics growth trajectories, focusing on the development of overall mathematics achievement and three separate mathematical competencies (basic precursors, advanced precursors, and computation). They (1) investigated whether first graders differ in mathematics growth, (2) identified classes of growth trajectories, and (3) analyzed the stability of trajectory group classifications, i.e. they examined whether students belong to similarly characterized groups across competencies. Investigations of (1) latent growth curve models revealed that for overall mathematics and for the three competencies, achievement increased during the first school year with significant variance in

students’ performance levels and slopes, indicating that students start at different levels and differ in mathematics growth. Results of (2) latent class growth analyses furthermore showed that diverse learning trajectories for overall mathematics and the three competencies exist (for an example, see Figure 1).

Figure 1. Learning growth trajectories for computation in grade one



Source: Salaschek/Zeuch/Souvignier 2014

In all competencies, most students followed cumulative growth patterns: students with higher starting performance usually showed a stronger development than students with lower starting performance, which overall led to a fan-spread pattern with persistently high-performing and persistently low-performing groups of students. For all competencies, however, some latent classes of students showed a compensatory growth, meaning that there were groups of initially low-performing students that showed steeper growth than initially high-performing children and thus caught up. Analyses of (3) class memberships revealed that in general, students in low-performing precursor

classes were less likely to be in high-performing computation classes than were students from high-performing precursor classes. The results of this study demonstrate the diversity of growth patterns in first grade mathematics and thus indicate that a single assessment at the beginning of schooling does not reliably identify students at risk of developing math difficulties. Ongoing research deals with mathematics growth trajectories in grade two (Schütze/Zeuch/Souvignier/Förster 2015) and grade three (Fleßner/Zeuch/Schütze/Souvignier 2018). These analyses have revealed patterns comparable to those in grade one.

3.3 Effects of providing teachers with LPA

The effects of LPA have been investigated in several intervention studies in general education (see Table 2 for a summary of the intervention studies presented in the following paragraphs). In one of the first intervention studies, effects of an additional teacher training were explored in addition to the implementation of LPA in 43 general education classrooms (Förster/Souvignier 2015). A quasi-experimental pretest-posttest control group design was used, with classes being assigned to either the control group or one of two LPA intervention groups (with or without training). Students' achievement was assessed at the beginning and at the end of the school year using standardized paper-pencil tests. The control-group teachers received the results from the standardized tests at the beginning of the school year, but they did not implement any kind of systematic formative assessment during the year. Classrooms in the intervention groups implemented LPA using a schedule of eight assessments during the year at intervals of three weeks. Thus, teachers in the experimental groups not only received information about their students' performance from one-time standardized tests, as did teachers of the control group, but they also formatively assessed their students' progress throughout the year to adapt instruction to individual needs. Within the intervention groups, it was further manipulated whether teachers used LPA only (LPA group) or additionally received three two-hour group training sessions on reading and interpreting LPA data and evidence-based reading fluency and reading comprehension instruction (LPA-T group). Results showed that for students in the LPA-T group, their growth in reading fluency and reading comprehension was significantly higher than for those in the control group. Likewise, compared to the control group, students in the LPA group showed higher growth in reading comprehension but not in reading fluency ($p = .059$). No differences in reading progress were found between the LPA and the LPA-T groups. By comparing the effects of LPA to a well-informed control group, these results highlight the net effect of teachers receiving learning progress information compared to receiving a one-

time standardized assessment. The amount of variance explained by affiliation to the LPA group, however, was rather small ($R^2 = .11$).

A second intervention study was conducted to evaluate whether involving students more strongly in LPA could enhance the effects of progress monitoring (Förster/Souvignier 2014a). Changing the focus from the teacher to the students was motivated by several factors. First, most research on progress monitoring and data-based decision-making has focused on teacher behavior, thus leaving a gap regarding the effects of student participation. Second, feedback on learning progress and goal achievement are key elements in self-regulated learning and might therefore enhance student achievement and positively affect student self-concept and motivation. Following a similar design as in the first intervention study, this study explored the development of reading achievement, intrinsic and extrinsic reading motivation, and individual, social, and absolute reading self-concept of 900 fourth-grade students. Classrooms were assigned to a control group, an LPA group, or an LPA group with goal-setting (LPA-G). Students in the LPA-G group specified individual goals before the LPA tests, reflected their goal achievement afterwards, and attributed their success or failure to certain causes. Results replicated findings from the first study, showing higher growth in reading achievement of LPA students compared to students in the control group. Growth in reading achievement in the LPA-G group, however, turned out to be significantly smaller compared to the LPA group and was similar to that of the control group. Moreover, unexpected negative effects of the goal-setting procedure were found on the development of intrinsic motivation and individual self-concept. While the negative motivational effects might be explained by fourth-grade students being overstrained by the goal-setting procedure, the absence of a beneficial effect of LPA combined with goal-setting on reading achievement might have arisen from the teacher paying attention to the goal-setting procedure instead of using assessment data to adapt instruction. As Stecker et al. (2005) point out in their review on the effects of curriculum-based measurement, however, the effects of progress monitoring do not occur from frequent testing alone but from teachers adapting their instruction to students' needs as indicated by the data.

3.4 Using quop for differentiated reading instruction

Given that neither the teacher training nor students' stronger involvement in LPA proved to be sufficient to enhance the effects of LPA in two intervention studies, we examined the effects of combining LPA with additional prepared teaching material, which included support for implementing feedback, differentiated reading instruction, or both.

A study by Förster et al. (2018) investigated the short- and long-term effects of combining LPA with differentiated reading instruction on reading fluency and reading comprehension. N=28 third-grade classrooms were randomly assigned to either an LPA group with differentiated instruction (LPA+DI) or a control group (CG). Teachers in the CG conducted business-as-usual-instruction, while teachers in the LPA+DI condition used quop and received prepared teaching material called the *Reading Sportsman* as well as a teacher training. The *Reading Sportsman* is designed to support differentiated reading instruction in a class-wide setting and includes two peer-based methods that were found to be effective in fostering reading fluency and reading comprehension. Students' reading fluency and comprehension were assessed at the beginning and the end of third grade and again in the middle of fourth grade. The results showed that quop and the *Reading Sportsman* can be implemented successfully. Students' growth in reading fluency was higher in the treatment conditions than in the CG ($d = .30$), and this effect remained stable. Students with lower reading skills benefited more from the treatment. No effects were found for reading comprehension. A possible explanation is that teachers applied the training method with focus on reading fluency more often, resulting in a lack of fit between students' level of achievement and the method of teaching used by the teacher. Thus, it seemed challenging for teachers to interpret progress data and use it to adapt their instruction to students' needs.

Consequently, a following study by Hebbecke and Souvignier (2018) investigated the effects of prepared teaching material designed to support teachers' interpretation of LPA data to give feedback and to provide differentiated instruction. The study also examined to what extent this approach can be implemented in regular reading lessons. In a three-group design with N=44 third-grade classrooms, an LPA group was compared to groups that additionally received prepared material and teacher training for feedback (LPA+FB) or for feedback as well as for differentiated instruction with the *Reading Sportsman* (LPA+FB+DI). Teachers in these LPA+ conditions were given further support in reading and interpreting the data by using an algorithm-based classification of students' results. Based on their quop-results, students were individually assigned to one of six profile groups. For each profile group, teachers were given a detailed description of students' strengths and weaknesses, appropriate learning goals, and training methods. Teachers could then use these descriptions to provide individual feedback (LPA+FB) and to implement individual reading instruction (LPA+FB+DI). While acceptability was high, teacher ratings of feasibility turned out to be somewhat lower. Results indicated no effects of the support on students' reading fluency and comprehension. A possible explanation is that the concept of formative assessment (LPA, feedback, differentiated instruction) was too complex and thus required changes in teachers' daily teaching

practice that were too profound to allow for effective implementation within one school year. A second possible explanation goes back to Yan and Keung Cheng (2015: 134) who summarized that “teachers probably still regard formative assessment as an added component, which needs extra time and resource, rather than an integrated part of regular instruction”.

Taken together, the studies show that the combination of LPA and the *Reading Sportsman* is fully accepted by teachers, can generally be implemented in entire classrooms (grades 2-4) in general education, and has the potential to affect teachers’ reading instruction and increase students’ learning. At the same time, teachers’ perception of feasibility is lower. Implementation of all components of formative assessment (LPA, feedback, differentiated instruction) seems to be challenging, and teachers need more support in implementing the complex concept of formative assessment. Thereby, it seems helpful to consider the implementation as a long-term and stepwise process in which teachers gain experience and develop teaching practice with one component before the next one can be implemented. Consequently, ongoing research focuses on the effects of different types of teacher support on implementation processes as well as students’ learning outcomes. In addition, prepared teaching material for differentiated instruction and feedback in reading in second grade as well as in mathematics in third grade has been developed.

Table 2. Overview of the intervention studies

	Domain, grade	Conditions	
Published research			
Förster/ Souvignier 2015	Reading 3 nd grade	CG	SAT, BAU
		LPA	SAT + LPA
		LPA+T	SAT + LPA + teacher training
Förster/ Souvignier 2014a	Reading 4 th grade	CG	SAT, BAU
		LPA	SAT + LPA
		LPA-G	SAT + LPA + students‘ goal setting
Förster, Schulte/ Souvignier 2018	Reading 3 rd - 4 th grade	CG	SAT, BAU
		LPA+DI	SAT + LPA + prepared teaching material (Reading Sportsman) + teacher training
Hebbecker/ Souvignier 2018	Reading 3 rd grade	LPA	SAT + LPA
		LPA+FB	SAT + LPA + prepared teaching material (Interpretation, Feedback) + teacher training
		LPA+FB+DI	SAT + LPA + prepared teaching material (Interpretation, Feedback, Reading Sportsman) + teacher training

Ongoing research			
Hebbecke, Meudt, Schütze/ Souvignier	Reading 2 nd - 4 th grade	LPA+FB+DI	LPA + prepared teaching material (interpretation, feedback, Reading Sportsman)
		LPA+FB+DI +T	LPA + prepared teaching material (interpretation, feedback, Reading Sportsman) + teacher training
		LPA+FB+DI +T+MP	LPA + prepared teaching material (interpretation, feedback, Reading Sportsman) + teacher training + multiplication
Schütze/ Souvignier	Mathematics 3 rd grade	CG	SAT, BAU
		LPA+T	SAT + LPA + teacher training
		LPA+FB+DI +T	SAT + LPA + prepared teaching material (feedback, differentiated instruction) + teacher training
Peters, Hebbecke/ Souvignier	Reading 2 nd grade	CG	SAT, BAU
		LPA+	SAT + LPA + teacher training (quop)
		LPA+DI	SAT + LPA + prepared teaching material (Reading Sportsman) + teacher training (quop, Reading Sportsman)

Note. CG=Control Group, LPA=Learning progress assessment, T=teacher training, G=goal setting, FB=Feedback, DI=differentiated instruction, MP=multiplication, SAT=standardized achievement tests, BAU=Business as usual instruction.

4. Challenges for learning progress assessment

Research on effects of LPA with quop shows that implementing this system results in a small but reliable improvement of students' achievement. Stecker et al. (2005), however, underline that it is not the frequent testing per se that leads to higher student achievement; in line with these findings, we also assume that successful student learning is driven by adaptive instruction by teachers who are well informed about students' levels of achievement and the progress students make. Providing teachers with additional material such as feedback sheets, classifications of students' results, and evidence-based material like the *Reading Sportsman* seem to be promising ways to support teachers in transforming assessment information into instructional decisions.

Theoretical models on data-based decision-making describe some preconditions for effective assessment-based differentiated instruction (Keuning/van Geel/Visscher 2017; Mandinach 2012; Staman/Timmermans/Visscher 2017). First, teachers need data literacy to read and interpret data. Second, from the data they must infer learning goals for their students. Third, they need to determine a strategy for goal accomplishment, and, finally, they must put these plans into action in the classroom. Each of these steps is challenging. Zeuch, Förster, and Souvignier (2017) found that teachers tend to focus on using test results to simply judge student achievement instead of inferring learning goals, as intended by formative assessment. Findings from Staman et al. (2017) and Keuning et al. (2017) point to a similar issue: They found that teachers have trouble transforming student progress data into adaptive instruction. However, this 'last step' in the process of data-based decision-making seems especially crucial.

In sum, these findings suggest that at least two major challenges arise when teachers are provided with digital data to support adaptive teaching. First, many teachers are not familiar with the theoretical concept of formative assessment. Understanding learning progress assessment as a type of feedback that supports instructional decision-making requires a (conceptual) change in well-established practice. The second challenge – adapting instruction to students' individual needs – can be addressed by providing teachers with differentiated material and teacher trainings that support evidence-based instructional approaches.

References

- Black, Paul/Wiliam, Dylan (1998): Assessment and classroom learning. In: Assessment in Education: Principles, Policy & Practice, 5, pp. 7-74.
- Deno, Stanley L. (1985): Curriculum-based measurement: The emerging alternative. In: Exceptional Children, 52, pp. 219-232.
- Fleßner, Janis/Zeuch, Nina/Schütze, Birgit/Souvignier, Elmar (2018): Learning trajectories in 3rd grade mathematics over eight measurement points: Which types are there? Paper presented at the Conference of the EARLI SIG15 – Special Educational Needs (Potsdam, September).
- Förster, Natalie/Kawohl, Elisabeth/Souvignier, Elmar (2018): Short- and Long-Term Effects of Assessment-Based Differentiated Reading Instruction in General Education on Reading Fluency and Reading Comprehension. In: Learning and Instruction, 56, pp. 98-109.
- Förster, Natalie/Kuhn, Jörg-Tobias/Souvignier, Elmar (2015): „Der Test liegt in zwei Parallelformen vor“ Untersuchung zur Äquivalenz von zwei Testverfahren am Beispiel computerbasierter Lesetests. [“There are two parallel test forms” Investigation of the equivalence of two parallel test forms using the example of computer-based reading tests] Talk at the 15th meeting of the section Educational Psychology of the German Psychological Society (Kassel, September).
- Förster, Natalie/Souvignier, Elmar (2011): Curriculum-Based Measurement: Developing a computer-based assessment instrument for monitoring student reading progress on multiple indicators. In: Learning Disabilities: A Contemporary Journal, 9, 2, pp. 65-88.
- Förster, Natalie/Souvignier, Elmar (2014a): Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. In: Learning & Instruction, 32, pp. 91-100. doi: 10.1016/j.learninstruc.2014.02.002
- Förster, Natalie/Souvignier, Elmar (2014b): Learning progress assessment in reading: How can we construct and evaluate tests to monitor student reading progress? Talk at the 7th meeting of the SIG Assessment and Evaluation of the European Association of Research on Learning and Instruction (Madrid, August).
- Förster, Natalie/Souvignier, Elmar (2015): Effects of providing teachers with information about their students' reading progress. In: School Psychology Review, 44, pp. 60-75. doi: 10.17105/SPR44-1.60-75
- Fuchs, Lynn S. (2004): The past, present, and future of curriculum-based measurement research. In: School Psychology Review, 33, pp. 188-192.
- Hebbecke, Karin/Souvignier, Elmar (2018): Formative Assessments im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. In: Zeitschrift für Erziehungswissenschaft, 21, pp. 735-765. doi: 10.1007/s11618-018-0834-y
- Jenkins, Joseph R./Fuchs, Lynn S. (2012): Curriculum-based measurement: The paradigm, history, and legacy. In: Espin, Christine A./McMaster, Kristen L./Rose, Susan/Wayman, Miya Miura (eds.): A measure of success. The influence of curriculum-based measurement on education. Minneapolis, MN: University of Minnesota Press, pp. 7-23.

- Keuning, Trynke/van Geel, Marieke/Visscher, Adrie (2017): Why a data-based decision-making intervention works in some schools and not in others. In: *Learning Disabilities Research and Practice*, 32, pp. 32-45.
- Kingston, Neal/Nash, Brooke (2011): Formative assessment: A meta-analysis and a call for research. In: *Educational Measurement: Issues and Practice*, 30, pp. 28-37. doi: 10.1111/j.1745-3992.2011.00220.x
- Mandinach, Ellen B. (2012): A perfect time for data use: Using data-driven decision making to inform practice. In: *Educational Psychologist*, 47, pp. 71-85.
- Salaschek, Martin/Souvignier, Elmar (2013): Web-based progress monitoring in first grade mathematics. In: *Frontline Learning Research*, 2, pp. 53-69.
- Salaschek, Martin/Souvignier, Elmar (2014): Web-based mathematics progress monitoring in second grade. In: *Journal of Psychoeducational Assessment*, 32, pp. 710-724. doi: 10.1177/0734282914535719
- Salaschek, Martin/Zeuch, Nina/Souvignier, Elmar (2014): Mathematics growth trajectories in first grade: Cumulative vs. compensatory patterns and the role of number sense. In: *Learning & Individual Differences*, 35, pp. 103-112. doi: 10.1016/j.lindif.2014.06.009
- Schütze, Birgit/Zeuch, Nina/Souvignier, Elmar/Förster, Natalie (2015): Growth trajectories for mathematical competencies. Paper presented at the 16th Biennial Conference of the European Association of Research on Learning and Instruction (Limassol, August).
- Staman, Laura/Timmermans, Anneke C./Visscher, Adrie J. (2017): Effects of a data-based decision-making intervention on student achievement. In: *Studies in Educational Evaluation*, 55, pp. 58-67.
- Stecker, Pamela M./Fuchs, Lynn S./Fuchs, Douglas (2005): Using Curriculum-Based Measurement to Improve Student Achievement: Review of Research. In: *Psychology in the Schools*, 42, pp. 795-819.
- Yan, Zi/Keung Cheng, Eric Chi (2015): Primary teachers' attitudes, intentions and practices regarding formative assessment. In: *Teaching and Teacher Education*, 45, pp. 128-136. doi: 10.1016/j.tate.2014.10.002
- Zeuch, Nina/Förster, Natalie/Souvignier, Elmar (2017): Assessing teachers' competencies to read and interpret graphs from learning progress assessment. Results from tests and interviews. In: *Learning Disabilities Research and Practice*, 32, pp. 61-70. doi: 10.1111/ldrp.12126

V. Economization of Education

Section Editors:

Marcelo Parreira do Amaral, University of Münster

Paul Fossum, University of Michigan

Education Gone Global: Economization, Commodification, Privatization and Standardization

Marcelo Parreira do Amaral¹ and Paul R. Fossum²

1. Introduction

After many decades of neoinstitutionalist research emphasizing education as a central piece of the World Polity, an assertion that “education has gone global” does not have a new or unfamiliar ring. Most researchers in the field can relate – in one way or another – to the arguments and consequences it entails. Nevertheless, more recently researchers of the Global Education Industry (GEI) have argued that the globalization of education has taken on a different meaning, pointing to a new facet of this development: education has become an economic enterprise unto itself, in which myriad actors produce, exchange, and consume educational goods and services, often on a for-profit basis (Verger/Lubienski/Steiner-Khamsi 2016). Understanding the qualitative and quantitative influence the GEI exerts upon education calls for recognition of various (economic) rationales that undergird these processes. With that purpose in mind, this chapter takes a closer look at the themes of economization, commodification, privatization and standardization of education on a global scale. In conceptual perspective, it deliberates on the nature and meaning of each of these terms, probing questions as to their significance for education practice, policy and research. The chapter *first* discusses the global dimension of education, *and then* provides a conceptual discussion of mutually related concepts used to grasp the ongoing transformations in the field of education globally. The *last* section briefly deliberates on the (potential) consequences of the topic at hand for education practice, policy and research, also raising some questions for further consideration.

-
- 1 Marcelo Parreira do Amaral is Professor of International and Comparative Education at the University of Münster. Email: parreira@uni-muenster.de
 - 2 Paul R. Fossum is Professor of Educational Foundations at Michigan University-Dearborn, College of Education, Health, and Human Services. Email: pfossum@umich.edu

2. Education globalized

Historically, the development of mass education has come hand in hand with the evolution of the nation state. Since the eighteenth century, education has come to be a national concern – in economic, social and cultural terms – for which large organizational and administrative apparatuses were created, in most cases by the state. As education for the masses has developed during the last two and a half centuries, it has been predominantly state-sponsored and controlled, eventually emerging as a crucial instrument in nation-building efforts (Benavot/Resnik/Corrales 2006).

Education research also remained largely focused on national characteristics and developments, with even comparative and international education scholarship prone to assume an analytic stance recently criticized as ‘methodological nationalism’ (cf. Robertson/Dale 2017). Also, with the educational debates of the past three decades having concentrated on the relevance and implications of coinages like ‘globalization’ and ‘internationalization’ processes for national education systems, education research has in turn focused increasingly on the impact of globalization on the national character of education, implying a diminishing importance and/or ability of the state to control and steer education, formerly a state prerogative (Green 1997; Mitter 2006; cf. Dale 2015 for a critique). Sociological research to tap an additional perspective has emphasized the diffusion of universalist scripts about education as fixtures of what it termed World Polity. For neoinstitutional scholarship, education is a global phenomenon because it was disseminated and gained traction worldwide as part of rationalized world models as to the legitimate forms of organization and agency (nation state, formal organization e.g. schools, and individuals) (Meyer et al. 1992; Meyer et al. 1997).

More recently, scholarship on the global dimension of education has shifted the focus to the impact of economic, political and cultural globalization on education, highlighting the policy responses throughout the world (Mundy 2005; Steiner-Khamsi/Waldow 2012; Mundy et al. 2016). In an era characterized by globalization across numerous sectors, industries, technologies and social movements, the rise of an education industry – and one operating on a global scale – may occasion little surprise (Verger/Steiner-Khamsi/Lubienski 2016; Parreira do Amaral/Steiner-Khamsi/Thompson 2019). The rise of a GEI goes along with a rapid dissemination and adoption of a range of global education policies including accountability systems and common core standards (Hartong 2018; Hartong/Piattoeva 2019). These global developments concern not only the privatization of education’s provision but also the assumption of implementation and management roles with characteristic adherence to standards, accountability and quality

rationales. Education is even increasingly a locus of investment and profit making by the interests of (for example) philanthropic organizations, education businesses and technology companies on a global scale (cf. Ball 2019). This has been arguably accompanied by a changed role of the state, which, by allowing and even fostering the privatization of political decision-making processes and by devising education policies aimed at generating profit, now acts itself as a key player in paving the way for the economization of education (Erfurth 2019).

Central to the globalization of education at present, common logics and modes of operation,³ – all primarily economic in character – pervade in education reform and restructuring activities worldwide. The remainder of this introductory chapter thus examines these rationales – their meanings and scope – discussing concepts with which recent developments in the education sector may be examined, and pointing to the multifaceted and multiscale quality of the globalization of education.

3. Economization, commodification, privatization, and standardization

The concept of *economization* implies a broad and cross-cutting transformation that excludes no social sphere, organization or actor: health, politics, sports, media, religion and not least education. With rare exception, all such social sectors are valued and evaluated in economic terms.

Thus, in addition to the structural changes economization demands, it also asserts a new language as well as changed semantics, discourses and knowledge about education. In the field of education, this refers to the process of redescription or reformulation of educational processes in the language of economic transaction. This reformulation of education has been key in situating education within a global market environment. In times of fiscal austerity, application of new public management to education has invited and legitimized economic thinking, norms, and procedures in the provision, management, and evaluation of education. As a result, economization entails new modes of education provision and oversight and new models for accountability and quality assurance.

This reformulation of education was important in situating education in a market environment. And the emergence of new public management in times of education austerity has anchored economic thinking, norms, and procedures in the provision, management, and evaluation of education (see Hartong/Hermstein/Höhne 2018). We see, with the rise of the GEI the

3 For a further discussion of this theme, see: Parreira do Amaral/Thompson 2019.

expansion of the global reach and power of economic actors in the promotion and sale of their products. Accordingly, the *development and enactment of educational policies* is aptly seen as a field of *strategic interaction and trade* (Verger 2012).

In sum, economization may be seen as the symbolic/semiotic process by which education is made ready for the market (that is, its marketization). Ever since the beginning of modern political liberalism, the notion of the “(free) market” was linked to the idea (or ideology) of an impersonal and neutral institution that mediates social interests. In classic economic thinking, the market is the sphere where individual efforts can be transformed into individual wealth and social advancement. An operative and symbolic coalition within the imagery of the “market” has become the core of neoliberal market rationality, with the “market” serving as the sphere within which social prosperity and individual well-being is realized. To be sure, the role of education in this imagery cannot be overestimated, and *marketization*, on the level of the GEI, signifies the move toward *market readiness* of those educational goods, services, policies and people as well that are deemed indispensable for economic growth, public health, social, as well as individual well-being on a global scale. At the same time, the established market relations weaken former structures and infrastructures of education (Lawn 2013).

In sympathy with the concept of economization, *commodification* posits education as a tradable good – something that, like any other good or service, is appropriately subject to mechanisms of marketing and exchange. Under this rationale, education is justifiably subject to economic rationalities and values – a consumer good responsive to private preference, for example, and one to be traded, and, speculated in line with the competitive dynamic of a marketplace. Commodification thus subsumes not only the privatization of education’s provision and funding, but also the escalating influence in the spheres of education provision, management, and research, of vehicles of financial capitalization (loans, borrowing, student debt, impact investment, etc.) and of finance activities common to the marketing of products and services (brokering, investing, speculating, etc.).

To be sure, the construction of tradable commodities is of utmost importance for the economic penetration of the education sector. Commodification, precisely because it promotes education as a good that is appropriately fungible in nature, engages education in an exchange of values. Education’s commodification is evident in the quality assurance/accountability and evidence-based movements that have become increasingly ubiquitous in the past decade. For instance, while large scale initiatives such as the World Bank’s SABER profess the best possible provision of education as its main objective, ideological bias and open advocacy for private education is evident (cf. Klees et al. 2020; Bous/Farr 2019).

Thus, commodification, promoting and relying upon transformed meanings and understandings of education as a consumer good, draws education into a global mercantile.

Privatization involving the shift of public money into the private sector and the resulting reassignment of a service once provided by public actors – education – into the hands of private actors is not novel. Long a topic requiring contextualization vis-à-vis respective nation states, their traditions, and their institutional frameworks, now, however, the construct increasingly transcends such boundary. Verger, Fontdevila, and Zancajo (2016) have delineated six paths toward education privatization that discern the contextual dispositions, agents, and mechanisms of privatization, for example, “education privatization as a state reform” (as in Chile and UK), “education privatization in social democratic welfare states” such as in Nordic countries, “scaling up privatization” as with the school reform in the United States, “privatization by default in low-income countries” due to low-fee private schemes such as the ‘one-dollar’ a day schools, “historical public-private partnerships” such as in Benelux countries, or ‘privatization along the path of emergency situations’ after political or natural catastrophes (see Verger/Fontdevila/Zancajo 2016: 11). Thus, in the context of the GEI, the increased complexity accompanying the globalization of policy infrastructures as well as the global diffusion of privatization (ibid.) comes in tandem with the concentration of power and agenda setting capacities (for example, in the World Bank or the OECD). Privatization is a controversial topic not only because evidence is scant as to its capacity to improve efficiency or to offer better ‘value-for-money’, but also due to its problematic relationship with the widely articulated view of education as a fundamental human right (cf. Macpherson/Robertson/Walford 2014; Singh 2014). Further, though, privatization can occur within education systems that remain largely state-funded and controlled. Illustrating how privatization assumes differing forms in international context, shadow education has for instance, become a widespread phenomenon across the world (cf. Bray 1999; Bray/Kwo/Jokić 2015).

Standardization of education refers to implementation at scale of uniform indicators for levels, paces, paths and outputs of education. With the intention of making its content (curricula), operations (teacher proofing and certification) and quality assessment (student and professional performance, effectiveness/efficiency) comparable and accountable, standardization is productively viewed as possessing different aspects. Brunson and Jacobsson (2000: 4ff.) differentiate standards for being, from standards for doing and from standards for having something. Standards for being something specify what something is (e.g., a species of animals or plant), belong to a class of things or actors (e.g., primary school, pupil, teacher, etc.). They are also used to measure something in a standardized way (e.g., statistics of all kinds) or to establish the meaning and/or use of something (e.g., dictionary definition,

grammar, pronunciation). Standards for doing specify how a process, service or course of action is to unfold or what it is to include; they entail understandings of how processes, products and their effects are to be planned, implemented/produced, and controlled/evaluated. In education, standards articulate what should be included in curricula and how long a track is to last, for instance, but also how instruction is to be designed, implemented and controlled, and not least how it is to be accounted for. Standards for having something situate expectations as to what a legitimate state, organization or individual ought to possess (e.g. a constitution, democratic elections, an education or welfare system, organizational leadership/structure, or a plan for one's own educational and professional career). Regarding education specifically, this refers both to what an education system is to include, but in particular, at the individual level, which (standardized) life courses, qualifications, knowledge, skills and dispositions pupils and students are supposed to have acquired. Central aims of standardization include maximizing levels of quality, compatibility/comparability, and interoperability. Standardization thus leverages both the increase of commodification and privatization of education and is reflected in contemporary emphasis on data-based and digital infrastructures in the governance of education.

Related to the concepts discussed above, *digitalization*, a current hyper-trend in education, stands also as a key driver of the global market in education. Over the last decades, powerful imaginaries and objectives revolving around “digital technology and education” have gained currency in educational improvement discourses. For instance, digital learning environments and learning analytics stand for the optimization and individualization of learning. The establishment and the provision of access to the Internet is heralded as enabling access to knowledge and as promoting social participation. And the use of digital technology is said to reduce ‘frictional loss’ thereby putatively improving knowledge management: Along with growing computing capacities, the storage, analysis, and prognostic evaluation of data stands as a powerful instrument of educational governance.

In short, the digital transformation of the educational sector is driven by the innovation, optimization, and the increasing accessibility of learning and learning processes it fuels. And for the expansion of the GEI, the significance of digitalization is difficult to overestimate. In the coming years, the so-called e-learning market is expected to be valued in the hundreds of billions of US dollars. Furthermore, technological innovations in education – for example in the use of digital devices in classrooms – open up new markets and new customers. Further, the collection and management of large data infrastructures offer new modes of educational governance comprise additional aspects of digitalization that are highly relevant for the GEI (Lawn 2013). Data infrastructures thus complement the management and monitoring of educational institutions (see Hartong 2018), and represent a means of

translating and mediating the measurability of educational processes; they are an important ingredient of new public management.

4. Discussion and conclusion

In this article, the global dimension of education has been thematized using rationales common in education development and summed up under the umbrella of an expanding Global Education Industry. Concepts of economization, commodification, privatization and standardization have shaped the transformation of education across the globe. We have argued that a central feature of the global dimension of education at present are the mutual rationales, logics and modes of operation, but, more critically, that these concepts are built on prevailing economic footings, and that they have come to permeate education reform and restructuring across the globe.

The two chapters that follow illustrate extant research on the topics discussed above. *Sabine Hornberg* discusses in her chapter how schools labelled 'IB World Schools' have steadily proliferated during the past decades. These schools are authorized by the International Baccalaureate Organization (IBO) to offer an international university entrance qualification – the International Baccalaureate. While originating in the international private school sector, today over half of the schools offering IB programs or parts of them are public schools. All programs or other services offered by the IBO have to be paid for privately. Hornberg argues that in contrast to earlier times, the field of international education is nowadays dominated by the IB as a consequence of the standardization of curricula and examinations provided by IBO. The author also shows that due to globalization processes, not only the private sector schools, but also national, state-run education systems offer IB education programs and services in order to be able to compete when serving internationally oriented parents and students. In their chapter, *Alexandra Ioannidou* and *Annabel Jenner* focus on adult and continuing education as a less regulated and standardized part of education systems. As they argue, new actors operating in a space where state authority is disputed, international, private organizations are increasingly developing an agenda-setting capacity and gaining regulatory sway. Drawing on sociological neo-institutionalism, they examine the role of the International Organization for Standardization (ISO) in assuring quality and setting standards in non-formal education.

References

- Ball, Stephen J. (2019): Serial Entrepreneurs, Angel Investors and Capex Light Education Business Start-Ups in India: philanthropy, impacting investing and systemic educational change. In: Parreira do Amaral, Marcelo/Steiner-Khamsi, Gita/Thompson, Christiane (eds.): *Researching the Global Education Industry – Commodification, the Market and Business Involvement*. Cham et al.: Palgrave, pp. 23-46.
- Benavot, Aaron/Resnik, Julia/Corrales, Javier (2006): *Global Educational Expansion. Historical Legacies and Political Obstacles*. Cambridge, MA: American Academy of Arts and Science.
- Bray, Mark (1999): *The Shadow Education System: Private Tutoring and its Implications for Planners*. Paris: UNESCO.
- Bray, Mark/Kwo, Ora/Jokić, Boris (eds.) (2015): *Researching Private Supplementary Tutoring: Methodological Lessons from Diverse Cultures*. Hong Kong: Springer/CERC.
- Brunsson, Nils/Jacobsson, Bengt (2000): The Contemporary Expansion of Standardization. In: Brunsson, Nils/Jacobsson, Bengt/Associates: *A World of Standards*. Oxford: Oxford UP, pp. 1-20.
- Dale, Roger (2015): Globalisierung in der Vergleichenden Erziehungswissenschaft re-visited: Die Relevanz des Kontexts des Kontexts. In: Parreira do Amaral, Marcelo/Amos, S. Karin (eds.): *Internationale und Vergleichende Erziehungswissenschaft. Geschichte, Theorie, Methoden und Forschungsfelder*. Münster et al.: Waxmann, pp. 171-187.
- Erfurth, Marvin (2019): International Education Hubs as Competitive Advantage: Investigating the Role of the State as Power Connector in the Global Education Industry. In: Parreira do Amaral, Marcelo/Steiner-Khamsi, Gita/Thompson, Christiane (eds.): *Researching the Global Education Industry – Commodification, the Market and Business Involvement*. Cham et al.: Palgrave, pp. 181-202.
- Green, Andy (1997): *Education, Globalization and The Nation State*. New York: Palgrave.
- Hartong, Sigrid (2018): Towards a topological re-assemblage of education policy? Observing the implementation of performance data infrastructures and ‘centers of calculation’ in Germany. In: *Globalisation, Societies and Education*, 16, 1, pp. 134-150.
- Hartong, Sigrid/Piattoeva, Nelli (2019): Contextualizing the datafication of schooling – a comparative discussion of Germany and Russia. In: *Critical Studies in Education*. doi: 10.1080/17508487.2019.1618887.
- Hartong, Sigrid/Hermstein, Björn/Höhne, Thomas (eds.) (2018): *Ökonomisierung von Schule? Weinheim/Basel: Juventa*.
- Klees, Steven J./Ginsburg, Mark/Anwar, Hiba/Baker Robbins, Melanie/Bloom, Heidi/Busacca, Cameron/Corwith, Anne/Decoster, Brendan/Fiore, Amanda/Gasior, Stephanie/Le, Hang M./Primo, L. Heidi/Reedy, Timothy D. (2020): The World Bank’s SABER: A Critical Analysis. In: *Comparative Education Review*, 64, 1, Ahead of Print.
- Lawn, Martin (2013): A Systemless System: Designing the Disarticulation of English State Education. In: *European Educational Research Journal*, 12, 2, pp. 231-241.

- Macpherson, Ian/Robertson, Susan/Walford, Geoffrey (eds.) (2014): Education, Privatisation and Social Justice. Oxford: Symposium.
- Malouf Bous, Katie/Farr, Jason (2019): False Promises: How Delivering Education through Public-Private Partnerships Risks Fueling Inequality Instead of Achieving Quality Education for All. Oxford: Oxfam International.
- Meyer, John W./Boli, John/Thomas, George M./Ramirez, Francisco O. (1997): World Society and the Nation-State. In: American Journal of Sociology, 103, 1, pp. 144-181.
- Meyer, John W./Kamens, David/Benavot, Aaron (eds.) (1992): School Knowledge for the Masses. Washington DC: Falmer Press.
- Mitter, Wolfgang (2006): Bildungssouveränität und Schulträgerschaft in Europa in historisch vergleichender Sicht. In: Bildung und Erziehung, 59, 1, pp. 5-20.
- Mundy, Karen (2005): Globalization and Educational Change: New Policy Worlds. In: Bascia, Nina/Cumming, Alister/Datnow, Amanda/Leithwood, Kenneth/Livingstone, David (eds.): International Handbook of Educational Policy. Dordrecht: Springer, pp. 3-17.
- Mundy, Karen/Green, Andy/Lingard, Bob/Verger, Antoni (2016): Introduction: The Globalization of Education Policy – Key Approaches and Debates. In: Mundy, Karen/Green, Andy/Lingard, Bob/Verger, Antoni (eds.): The Handbook of Global Education Policy. New Jersey: John Wiley & Sons, pp. 1-20.
- Parreira do Amaral, Marcelo/Thompson, Christiane (2019): Changing Education in the GEI – Rationales, Logics, and Modes of Operation. In: Parreira do Amaral, Marcelo/Steiner-Khamsi, Gita/Thompson, Christiane (eds.): Researching the Global Education Industry – Commodification, the Market and Business Involvement. Cham et al.: Palgrave, pp. 273-290.
- Parreira do Amaral, Marcelo/Steiner-Khamsi, Gita/Thompson, Christiane (eds.) (2019): Researching the Global Education Industry – Commodification, the Market and Business Involvement. Cham et al.: Palgrave.
- Peters, Michael A./Besley, Tina (2015): Finance Capitalism, Financialization, and the Prospects for Public Education. In: Peters, Michael A./Paraskeva, João M./Besley, Tina (eds.): The Global Financial Crisis and Educational Restructuring. New York: Peter Lang, pp. 21-49.
- Robertson, Susan/Dale, Roger (2017): Comparing Policies in a Globalizing World. Methodological reflections. In: Educação & Realidade, 42, 3, pp. 859-876.
- Singh, Kishore (2014): Report of the Special Rapporteur on the Right to Education on Privatisation. New York: United Nations General Assembly. <http://www.right-to-education.org/node/371> [Last accessed February 5, 2020].
- Steiner-Khamsi, Gita/Waldow, Florian (eds.) (2012): Policy Borrowing and Lending in Education: World Yearbook of Education. London: Routledge.
- Verger, Antoni (2012): Framing and Selling Global Education Policy: The Promotion of Public-Private Partnerships for Education in Low-Income Contexts. In: Journal of Education Policy, 27, 1, pp. 109-130.
- Verger, Antoni/Fontdevila, Clara/Zancajo, Adrián (2016): The Privatization of Education: A Political Economy of Global Education Reform. New York: Teachers College Press.
- Verger, Antoni/Lubienski, Christopher/Steiner-Khamsi, Gita (eds.) (2016): World Yearbook of Education 2016: The Global Education Industry. New York: Routledge.

Agents of Privatization: International Baccalaureate Schools as Transnational Educational Spaces in National Education Systems

Sabine Hornberg¹

1. Introduction

National education systems worldwide are exposed to processes and effects of internationalization, globalization and transnationalization, such as international large-scale assessments (Maddox 2018), internationally compatible education offers and certificates. At the same time, a steadily growing private education market can be observed worldwide, materializing itself in many different ways. Publications tackling worldwide processes of privatization in national education systems are rare, both from a national as well as an international or global perspective. Against this background, Verger, Fontdevila and Zancajo (2016) have undertaken the challenge to shed light on some of the many facets of privatization in education from a global perspective. The complexity results not least from the diversity of interest groups advocating for private education. As they argue:

Privatization solutions are recommended and advocated by a broad spectrum of actors, from local interest groups to international organizations and private foundations. In some settings, even “strange bedfellows” (agents with apparently divergent interests, such as ethnic minority groups and conservative think tanks) end up advocating for similar forms of education privatization (Apple/Pedroni 2005). To all of these different actors, privatization is seen as a formula to expand choice, improve quality, boost efficiency, or increase equity (or all of these things simultaneously) in the educational system. (Verger/Fontdevila/Zancajo 2016: 3)

Given this widespread interest, it is perhaps not surprising to note the success of the private education sector worldwide. According to data provided by the UNESCO Institute of Statistics in 2015, between 1990 and 2012 the percentage of pupils enrolled in private primary education increased up to about 16% in most countries worldwide, “whatever their level of economic development – although this trend is not so marked in high-income and lower-middle-income countries” (ibid: 4). Looking at this trend focusing worldwide regions, only in sub-Saharan Africa this was not the case, while in 17 out of 21 OECD countries for which the respective data was available,

1 Sabine Hornberg is Professor of School Pedagogy and General Didactics at Dortmund Technical University. Email: sabine.hornberg@tu-dortmund.de

private expenditure on education has also increased since the mid-1990s (ibid: 5). Verger, Fontdevila and Zancajo (2016) do not speak of a private education market or sector, but of education privatization as a process, outlining that:

Education privatization can be defined broadly as a process through which private organizations and individuals participate increasingly and actively in a range of education activities and responsibilities that traditionally have been the remit of the state. (ibid: 3)

Drawing on Ball and Youdell (2008), Verger et al. (2016) furthermore distinguish between two privatization trends, the first one being of interest here:

[...] (a) privatization of public education, or ‘exogenous’ privatization, which involves ‘the opening up of public education services to private sector participation [usually] on a for-profit basis and using the private sector to design, manage or deliver aspects of public education’; and, (b) privatization in public education, or ‘endogenous’ privatization, which involves the ‘importing of ideas, techniques and practices from the private sector in order to make the public sector more like businesses and more business-like’” (Verger et al. 2016: 8).

In what follows, I will introduce what I consider to be a paradigmatic case of “privatization of public education” in the terms used by Verger and colleagues. More specifically, I discuss a form of education that can increasingly be observed worldwide since the turn of the new millennium in the field of international education, under the authority of the International Baccalaureate Organization (IBO), a non-profit organization founded in 1968 in Geneva, Switzerland. Since the beginning, the IBO has offered the International Baccalaureate (IB), an international university entrance qualification accepted by a steadily growing number of universities worldwide. In 2019 these amounted to more than 2,500 universities in 75 countries. Students have to complete the two-year IB Diploma Program in order to participate successfully, available since 1968 and complemented by accompanying K-12 education programs. All programs or other services authorized by the IBO have to be paid for privately.

With respect to this example of exogenous privatization, three aspects will be elaborated upon: First, I will argue that other than in the past, the field of international education is now dominated by the IB as a consequence of the standardization of curricula and examinations provided by this organization. Second, I will show that due to processes of internationalization, globalization, and transnationalization, not only the private school sector but also national, state-run school systems offer IB education programs and services in order to compete with other schools, while serving internationally oriented parents and students. Third, I will suggest that IB services, programs, etc. represent transnational educational spaces and agents of privatization in national school systems, thus putting new demands on national education systems.

2. Standardization and the rise and expansion of the International Baccalaureate

Since World War II, a steadily growing educational market has been unfolding worldwide, not only in tertiary education but also in public K-12 education systems. This market is complemented by a ramified network of educational organizations such as the International Schools Association (ISA), the International Schools Service (ISS) and the IBO, to name only some of the most influential organizations that have added to the growing number of international schools worldwide, especially since the turn of the millennium. Traditionally, international schools were primarily private schools, serving highly mobile families whose purpose for choosing these schools was to ensure some continuity to their children's education. These international schools often offered North-American or English curricula and university entrance qualifications, such as the English General Certificate of Secondary Education (GCSE) or the International General Certificate of Secondary Education (IGCSE) provided by Cambridge University.

Different reasons for the creation of international schools prevailed among those engaged in the international school sector. Some practitioners in the field of international schools argued that:

They have been created piecemeal, in response to immediate need, in answer to local pressure from globally mobile business enterprises, development aid agencies and diplomats for a (largely) English-medium education of sufficient quality to reduce the potentially negative impact of parental career moves on accompanying children, and to ease re-entry into national systems. Their driving force is pragmatic, not philosophical [...] (Bartlett 1998: 77).

Others argued normatively with reference to the UNESCO's Declaration on International Education (UNESCO 1974) by demanding an education supporting international dialogue and intercultural understanding. Having researched the international school sector for many decades, Hayden & Thompson (1995) stated, nearly a quarter of a century ago, that:

Many such schools have grown up in response to local circumstances on a relatively ad hoc basis and, although there are certainly subgroupings controlled by central organisations (such as the network of international schools supported by Royal Dutch Shell), for the most part the body of international schools is a conglomeration of individual institutions which may or may not share an underlying educational philosophy [...] (Hayden/Thompson 1995: 332).

Given the IBO and in particular the standardized services, educational programs and certificates offered by this organization, this is no longer the case today. In reaction to the significant increase in the number of international schools since the 1950s, an initiative for the establishment of an internationally compatible university entrance qualification developed in the

1960s, culminating in the introduction in 1968 of the International Baccalaureate and the IB Diploma Program. Influential representatives of national education systems took part in this development: for the Federal Republic of Germany, the then director of the Max Planck Institute for Human Development² and president of the Education Council, Hellmut Becker; for France, the former director of the University of Nancy; for Great Britain the former head of the “Department of Educational Studies” in Oxford, and for Belgium the director of the “Carnegie Endowment for World and Peace” at that time and later head of the “European Science Foundation”. Furthermore, the former director of the “US College Entrance Examination Board’s Advance Placement Program” as well as numerous teachers from international schools who are not cited individually here had a considerable impact on the arrangement of the international upper grades curriculum and on the IB (Fox 1991: 328). The predominance of western, at that time economically and politically influential states is thus reflected on the IB during its initiation period. It took another 24 years before the IBO added the IB Middle Years Program (MYP) in 1992, followed by the IB Primary Years Programme (PYP) in 1997 and by the Career-related program (CP) in 2012. All programs are offered in English, French, and Spanish; the IB Middle Years Program is also offered in ten other languages, with English as the predominant language of use in schools worldwide. The range of services provided by the IBO includes:

- The organisation, counselling, and certifying of schools as IB World schools,
- the organisation of congresses for schools and teachers and of teacher training courses,
- the development and provision of curricula, teaching, and learning materials,
- the collection of students’ data and certification of students’ achievement.

Thus, the IBO provides services for an international educational market, whose equivalents in national education systems fall under the authority of the state. Schools wanting to make use of services offered by the IBO have to pay for this privately. The IBO runs offices in different parts of the world, for example the IB Foundation Office in Geneva, the curriculum center in The Hague and four regional IB assessment centers in Cardiff (United Kingdom), The Hague (Netherlands), Washington, D.C. (USA) and Singapore. If a school aims at offering IB education programs, it has to successfully complete a cost intensive accreditation process for acquiring the right to carry the title

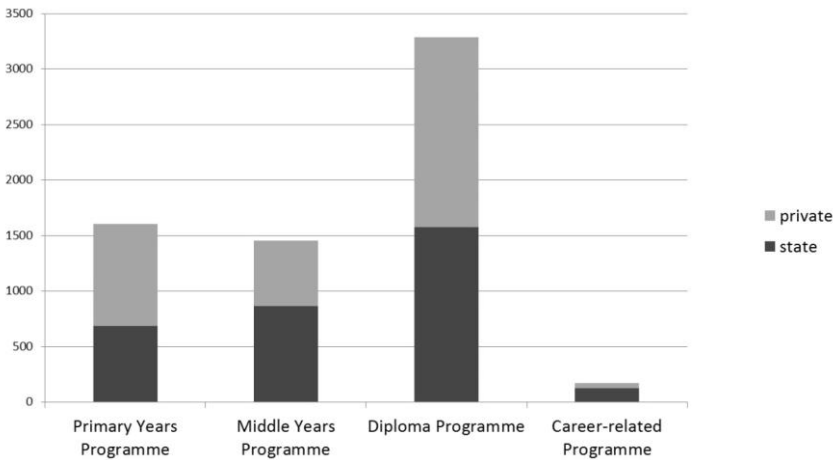
2 The German name of the institute and its English Translation differ: In German the institute is called *Max-Planck-Institut für Bildungsforschung*, e.g. educational research.

“IB World School”. Hence, a form of branding is exercised, with the IBO acting as a supplier on a global education market (Cambridge 2002; Resnik 2012: 259).

3. IB services and programs offered at state-run schools

At the outset, IB education programs were offered at private, often international schools catering for the children of highly mobile and privileged families. But since the turn of the millennium the number of single state schools or even whole school districts offering an IB education has constantly increased worldwide (Resnik 2012). The picture has changed to such a degree that today more than half of all schools offering IB education programs are state schools, as shown in Figure 1:

Figure 1. Number of IB programs worldwide at state-run and private schools (as of February 2018)



Source: Figure represents data by the International Baccalaureate Organization 2018: <http://ibo.org/programmes/find-an-ib-school/> [Last accessed February 27, 2018].

In 2018, Canada had the highest number in the world of state schools offering IB education programs. This is also an expression of the fact that the IB has different forms of partnerships, such as with governments, districts or groups of schools. In 2014, for example a partnership with the Canadian government and certain provinces was established covering a “Broad educational reform, Access to IB programs for all students, IB teacher support, Integration of IB into state systems, Linking the IB with state higher education” according to the IBO website³. Similar forms of partnership exist for some states in the United States of America. In Japan a dual language IB Diploma has been developed. In Germany, assessment and support services for some disciplines such as history, biology and theory of knowledge are offered in the German language. The Central Agency for Schools Abroad (*Zentralstelle für das Deutsche Auslandsschulwesen/ZfA*) enables interested schools outside Germany to have access to the IB, covering subjects in German. These are only a few examples of IB cooperation with governments, as shown on the IBO website. Hence, the overall picture reveals IB educational offers being adopted by public schools with funding from the state as well as by private, non-state institutions, such as associations or parents. The motives for this are often competitive advantages on a global educational market (Hornberg/Zipp-Timmer 2018: 10).

These perspectives correspond partly with a spread and increase of IB educational branches, especially since 2010. Such development was strategically planned by the IBO following a change in its top management. While experienced educationalists with a history of engagement in international schools had traditionally led the IB organization (for e.g. George Walker, from 1999-2005), a swift change happened with the arrival of Jeffrey R. Beard, the first director who was experienced in management and who thoroughly reformed the IBO as an organization during his period in office from 2005 to 2013 (cf. Tarc 2009). Since 2014, Dr. Siva Kumari has been the first woman to lead the IB organization. According to the IB website, she especially aims to spread a diverse range of IB offers in education systems in order to enhance the use of IB programs and services.

Hence, state education systems or single schools support IB education offers either directly, by paying student fees to attend IB education programs, or indirectly, by providing school campuses and facilities, state financed teachers and so forth. Thus, in cases where state schools offer IB education programs and certificates, a hidden and indirect, or an overt direct funding of private education takes place inside the state education system.

3 <https://www.ibo.org/benefits/ib-as-a-district-or-national-curriculum/government-partnerships/>

4. Transnational educational spaces in national education systems

In their monograph, Verger, Fontdevila and Zancajo (2016: 6) aim to:

[open] the black box of education privatization reform processes at an international scale. No other piece of research looks systematically at the scope of education privatization trends and scrutinizes the reasons, agents, and conditions behind the dissemination and adoption of privatization policies in educational systems from a comparative and global political economy perspective.

While a comparative and global political economy perspective is important to understand education privatization from the vantage point of policies and national education systems placed in their *international* context, what is suggested here is to look at the case of IB educational offers authorized by the IBO from a *transnational* perspective by referring to the concept of transnational educational spaces. This concept is suggested to be more adequate in terms of transcending the concept of the nation as a container and considering instead border-crossing aspects of education, which, as I argue, are the basis of privatization of education as exemplified by IB programs and other services offered by the IBO. The concept of transnational educational spaces will be outlined below.

The concept of transnational educational spaces was coined in the German-speaking context with reference to sociologists Ludger Pries and Thomas Faist on transnationalism, transnational social spaces, and transmigration more generally⁴. Following Pries (2001: 9), transmigration is: *“a modern type of a nomadic way of life [that gives rise to] transnational social spaces.”* Such spaces can extend across nations or continents and are constituted through the transmigrants’ conduct of life, with migration no longer understood *“as a singular or twofold changeover between two sites (areas of origin and arrival), but as a genuine component of definitely continuous biographies”* (Pries 2001: 49). Furthermore, with reference to Pierre Bourdieu and similarly to Faist (2000), Pries (2001) uses the term “space” and defines *“transnational social spaces”* as:

[...]a kind of ‘pluri-local interrelations’ (Elias 1986). Thus, transnational social spaces are relatively stable, condensed configurations of social daily routines, symbolism and artefacts, allocated to various sites or spread between multiple extended areas. Transnational social spaces emerge together with transmigrants (and transnational companies); both determine each other (Pries 2001: 53).

Here, the term “space” is not used in a conventional physical meaning as in the sense of a location such as a town or a country, but in the sense of a

4 For more on these concepts from studies of migration, see Glick Schiller/Basch/Blanc-Szanton (1992).

relatively stable relationship between protagonists, exceeding national borders. Taking up this perspective, Adick (2005: 262-266) and Hornberg (2010: 65-77; Hornberg 2014) suggest a concept of transnational educational spaces which links three previously separate, but parallel discourses:

1. socialization in transnational spaces,
2. transnational convergences in education, and
3. transnational education.

First, socialization in transnational spaces refers to the approaches as spelled out with reference to a sociological perspective on migration. Education research, for example, considers the question of to what extent multi-lingualism serves as a resource for transmigrants and/or transnational networks. Second, the term 'transnational convergences' is represented through worldwide isomorphism or structural similarity among institutions in education, as outlined from a neo-institutional perspective under the umbrella of the world polity theory (Meyer/Ramirez/Rubinson/Boli-Bennett 1977). These transnational convergences are, at the same time, a prerequisite for and the result of transnational educational spaces, because participation in transnational educational spaces relies to a certain extent on the connectivity and translatability of educational processes (Adick 2005: 263). Third, the term 'transnational education' refers to a definition put forward by the UNESCO and the Council of Europe in January 2002, when they drafted a Code of Good Practice for the Provision of Transnational Education. There, transnational education was defined as:

All types of higher education study programme, or set of course study, or educational services (including those of distance education) in which the learners are located in a country different from the one where the awarding institution is based. Such programmes may belong to the educational system of a state different from the state in which it operates, or may operate independently of any national system (Council of Europe 2001: 8).

According to this definition, transnational education takes place only in tertiary education and in classical 'private' realms of educational provision. However, developments in the public K-12 education system, such as the educational services offered by the IB organization which increasingly enter the public education realm, can also be examined referring to the concept of transnational educational spaces.

Today, the International Baccalaureate serves about a million students in 161 countries. As an alternative to national curricula and certificates authorized by the state, the IB has appealed to a steadily growing number of students and schools from the public system. The IBO is responding to, and at the same time supporting, this process by expanding educational services that satisfy the criterion of international and transnational compatibility that has become relevant for nationally-organized provision of education as well. This

increases the attractiveness of the IB to state schools even though the consumers, schools and students, have to pay for these educational offers themselves.

This raises the question whether transnational educational spaces serve as markers of distinction in state education systems, and which consequences this may have for school systems, schools and students. To date empirical evidence to tackle this question is rare, but some research has already been undertaken (for Latin America see e.g., Resnik 2012; for Germany e.g., Helsper/Dreier/Gibson/Kotzyba/Niemann 2015). While the aforementioned studies and others not mentioned here (e.g., Bunnell 2015; Sheveleva/Redkina 2013) shed some light on why state education systems or single state schools take on IB services and what consequences this has for teachers and students, there is still too little knowledge available to illuminate this mosaic tile of the privatization of education.

5. Conclusion

This chapter focuses on the IBO and its associated products, services, and qualifications as a case of “privatization of public education” or “exogenous” privatization in the terminology used by Verger et al. (2016). More specifically, I suggested that the IBO can be seen as an “agent of privatization” within the public education sector, and some of its particularities were closely examined in three argumentative steps.

In the first step (section 2), an explanation was offered for the popularity of the IB programs and their increasing adoption worldwide. This is a consequence of the standardization of curricula and examinations provided by this organization. Standardization in terms of the ‘product’ offered and ‘sold’ to schools who have to pay for it privately cannot be understood, however, without grasping processes of branding and marketing involved and hence illuminating the privatization dynamics at work. In the second step (section 3), the broader context in which the success of the IB model is located was outlined. No longer left unaffected by processes of internationalization, globalization, and transnationalization, state education systems are increasingly eager to support the spread of the IB education offers either directly or indirectly, as part of a perceived necessity to compete in the larger, global educational race. The motivations offered for adopting or choosing an IB educational program within public state-run schools often echo the privatization rhetoric placing a premium on competition and ‘staying ahead’. Therefore, states take active part in exogenous privatization of education via the IB. Finally, in the last step (section 4), an argument was made to extend the scope of discussing privatization of education by conceptualizing the

phenomena at hand beyond the *international* paradigm. Instead, a *transnational* perspective that crosses borders not only in terms of (nation-)states but also between private and public educational spheres was proposed. To this end, the concept of transnational educational spaces was offered as a possible device to further expand on IB-state-run schools as agents of privatization.

References

- Adick, Christel (2005): Transnationalisierung als Herausforderung für die International und Interkulturell Vergleichende Erziehungswissenschaft. In: Tertium Comparationis. In: Journal für International und Interkulturell Vergleichende Erziehungswissenschaft, 11, 2, pp. 243-269.
- Bartlett, Kevin (1998): International Curricula: More or Less Important at the Primary Level? In: Hayden, Mary/Thompson, Jeff J. (eds.): International Education. Principles and Practice. London,UK/Sterling, USA: Kogan Page, pp. 77-91.
- Bunnell, Tristan (2015): The rise and decline of the International Baccalaureate Diploma Programme in the United Kingdom. In: Oxford Review of Education, 41, 3, pp. 387-403.
- Cambridge, James (2002): Global product branding and international education. In: Journal of Research in International Education, 1, 2, pp. 227-243.
- Council of Europe (2001): The Committee of the Convention on the Recognition of Qualifications concerning Higher Education in the European Region. <http://www.unesco.org/education/studyingabroad/highlights/riga.doc> [Last accessed May 6th 2019].
- Faist, Thomas (2000): The Volume and Dynamics of International Migration and Transnational Social Spaces. Oxford: Oxford University Press.
- Fox, Elisabeth (1991): Das Internationale Bakkalaureate – Wegbereiter einer weltweiten Erziehung und Bildung. In: Bildung und Erziehung, 44, 3, pp. 327-349.
- Glick Schiller, Nina/Basch, Linda/Blanc-Szanton, Cristina (1992): Towards a Transnational Perspective on Migration: Race, Class, Ethnicity, and Nationalism Reconsidered. New York: New York Academy of Sciences.
- Hayden, Mary/Thompson, Jeff J. (1995): International schools and international education: A relationship reviewed. In: Oxford Review of Education, 21, 3, pp. 327-345.
- Helsper, Werner/Dreier, Lena/Gibson, Anja/Kotzyba, Katrin/Niemann, Mareike (2015): Exklusive Gymnasien und ihre Schüler. Passungsverhältnisse zwischen institutionellem und individuellem Schülerhabitus. Studien zur Schul- und Bildungsforschung. Wiesbaden: Springer.
- Hornberg, Sabine (2010): Schule im Prozess der Internationalisierung von Bildung. Münster: Waxmann.
- Hornberg, Sabine (2014): Transnational educational spaces. Border-transcending dimensions in education. In: Vega, Leoncio (ed.): Empires, post-coloniality and

- interculturality. New challenges for comparative education. Rotterdam: Sense Publishers, pp. 171-180.
- Hornberg, Sabine/Zipp-Timmer, Marie Elisabeth (2018): Voraussetzungen und Umsetzung des Internationalen Baccalaureates in Deutschland – eine Fallstudie an einem Gymnasium in Nordrhein-Westfalen. Unp. Report: TU Dortmund University.
- International Baccalaureate (2019): Government partnerships.
<https://www.ibo.org/benefits/ib-as-a-district-or-national-curriculum/government-partnerships/> [Last accessed May 17 2019].
- Maddox, Bryan (ed.) (2018): International Large-Scale Assessments in Education: Insider Research Perspectives. London: Bloomsbury Publishing
- Meyer, John W./Ramirez, Francisco O./Rubinson, Richard/Boli-Bennett, John (1977): The world educational revolution. In: *Sociology of Education*, 50, pp. 242-258.
- Pries, Ludger (2001): Internationale Migration. Bielefeld: transcript.
- Resnik, Julia (2012): The denationalization of education and the expansion of the International Baccalaureate. In: *Comparative Education Review*, 56, 2, pp. 248-269.
- Sheveleva, Marina/Redkina, Anastasiia (2013): International Baccalaureate Programme in Perm Region (Russia) as a government initiative: First results. In: *Mediterranean Journal of Social Sciences*, 4, 11, pp. 527-533.
- Tarc, Paul (2009): Global dreams, enduring tensions. International Baccalaureate in a changing world. New York: Peter Lang.

Regulation in a Contested Space: Economization and Standardization in Adult and Continuing Education

Alexandra Ioannidou¹ and Annabel Jenner²

1. Introduction

Across a variety of different countries, adult and continuing education (ACE) has been built bottom-up, on the initiative of labor and civil movements, and has traditionally stood outside formal, institutionalized education. It is generally less regulated and less standardized than K-12, higher education, or vocational education and training (VET), and is not mainly financed by the state. ACE differs considerably across countries, more than formal education does, which demonstrates institutional isomorphism around the world (Meyer/Ramirez 2003). What are the reasons for these differences?

Adult learning systems are embedded in specific economic and social arrangements, “they lie at the intersection of a variety of other systems including a nation’s education and training system, labor market and employment system and other welfare state and social policy measures” (Desjardins 2017: 21). ACE is also linked to a range of stakeholders (associations, chambers, communities of interest, industry) according to the historical origins of adult and continuing education in each country, the type of educational governance, and the type of skill formation regime. The state is *only one* amongst other actors in the policy field of adult and continuing education, and hierarchical order just one possible governance form amongst others. The dynamics that arose from the interaction of – state and non-state – policy actors at various levels (local, regional, national, international) and the variety of patterns of interaction among them (networks, coalitions, negotiations, mutual adjustment) are linked to certain characteristics of this policy field (Ioannidou 2007): the range of individual and collective stakeholders, multi-level structure, less regulation and standardization than other educational sectors. The scarce regulation by the state is a characteristic feature of ACE in many countries and leaves regulatory room to non-state actors, thus calling for concepts and strategies which state actors are unable or unwilling to develop or implement on their own.

1 Alexandra Ioannidou is Research Associate at the German Institute for Adult Education, Leibniz Centre for Lifelong Learning. Email: ioannidou@die-bonn.de

2 Annabel Jenner is Research Associate at the German Institute for Adult Education, Leibniz Centre for Lifelong Learning. Email: jenner@die-bonn.de

Adult and continuing education is not only *less regulated* but also *less homogeneous* than other education sectors regarding its institutional structure, function and target groups. Various organizations provide adult and continuing education, for example non-profit associations (environmental, political, confessional, etc.), adult education centers or community colleges, training departments of businesses, as well as commercial training institutes. The heterogeneity of institutional forms and the manifold regulatory structures and actors ensure the realization of a broad spectrum of general, vocational and employment-related ACE opportunities. In order to attract participants (whether paying out-of-pocket, sponsored by a company or mandated in the context of active labor market policies), ACE providers (whether publicly funded, not-for-profit, or commercial) compete for resources and legitimacy (Schrader 2014; see also section 2 in more detail). Thus, competition and market principles are part of the institutional and regulatory variety characterizing ACE.

Competing in a scarcely regulated space implies that ACE providers develop alternative strategies of standard setting. As we will show, building reputation by adopting widely recognized quality standards has become a necessity for competitiveness, as it secures resources and legitimacy. Therefore, in this contribution we ask from an international perspective whether and to what extent the implementation of quality standards through Quality Management Systems (QMSs) points to new actors beyond the state taking on standard setting functions in ACE. Drawing on a governance perspective, which emphasizes the dynamics arising from the interplay of different actors and their coordination principles in a multilevel system (Ioannidou 2014; Schemmann 2014), we argue that in a contested space of weak state authority, new actors emerge, taking on standard setting functions: international organizations and private actors. Drawing on neo-institutionalism, which highlights the embeddedness of organizations in their environments (focusing here on DiMaggio/Powell 1983), we discuss the argument that organizations within a shared context are likely to develop similar strategies in dealing with challenges and expectations regarding quality standards. Focusing on the role of the International Organization for Standardization (ISO) allows us to analyze the standard setting function of an international, private actor and to relate processes of standardization to economization. We finally discuss the consequences of economization for provision of and participation in adult learning. We address ‘provision’ in ACE because it is not normally regulated by national authority; and ‘participation’ because in most cases it is not mandatory. While provision of adult learning calls into question issues of quality assurance, participation in adult learning raises equity concerns.

In this paper, we first introduce the characteristics and distinctive features of ACE compared to other educational sectors, in particular as regards

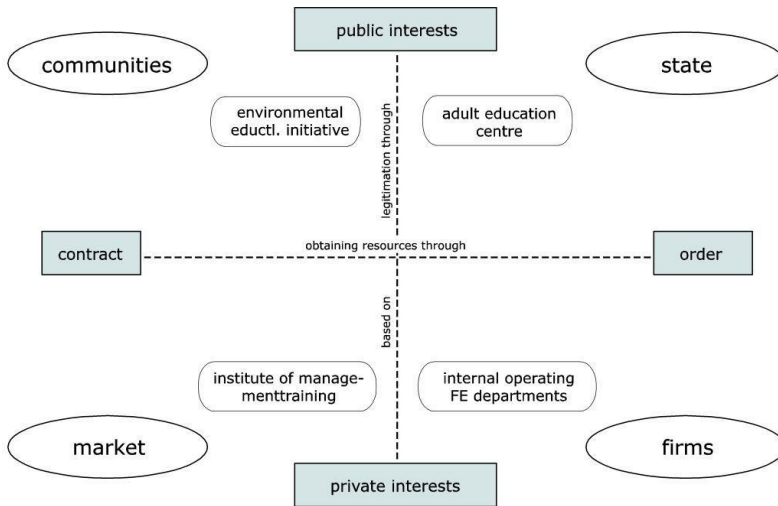
regulation and economization. Next, we outline the role of QMSs for ACE providers in securing resources and legitimation by taking on a standard setting function. Drawing on the role of ISO, a private actor, we discuss the international expansion of shared quality standards from a neo-institutional perspective, highlighting processes of standardization in terms of isomorphic processes and discussing their regulatory influence. We conclude with comments on the implications of economization for provision and equity and point out questions for further research.

2. The field of ACE: institutional heterogeneity between state and market

The institutional structure of adult and continuing education demonstrates a wide variety across countries. Referring to neo-institutionalism and social modernization theories, Schrader (2014) proposes a typology for describing the institutional variety of ACE providers, which captures the heterogeneity of the field. It draws on the basic assumption that ACE providers do not solely obtain material resources to ensure their existence. Rather, they also seek to obtain legitimacy from relevant actors in their environment. Depending on the modalities providers use to obtain resources (by hierarchical assignments or contracts) and legitimacy (towards public or private interests), Schrader distinguishes four “reproduction contexts” of ACE providers: 1) communities, 2) state, 3) firms and 4) market. These fields (‘contexts’) define the space in which ACE providers operate. Fig. 1 illustrates the positioning of exemplary paradigmatic organizations.

Taking this typology as an analytical frame to understand the way ACE providers operate, it is obvious that the state-regulated field is only one out of four. The majority of providers operates in fields where state actors are neither the only nor the most influential ones (e.g., in the context of the market or firms). Being dependent on acceptance and mainly voluntary participation, providers compete with one another to obtain resources and legitimacy. They have to commit themselves to serving public or private interests whilst providing innovative learning offers and flexible support structures. Thus, economic rationality and the logic of the market are familiar to ACE. Whereas competition and market principles are part of the institutional and regulatory variety characterizing adult learning systems, these principles can become problematic when they apply to learners as they enhance inequalities in access and participation in adult learning (see section 4).

Figure 1. Reproduction contexts – location of paradigmatic ACE providers



Source: Schrader 2014: 60

We refer to the term economization “as an increasing importance of economic considerations for financial profits and costs in particular societal sub-systems or even society-wide” (Schimank/Volkman 2012: 37). There again the term marketization implies an exposure of service providers to market principles (ibid.: 37), “reducing the impact of the state by initiating competition between non-profit and for-profit providers” (Ewert 2009: 24) and, thus, leading to an increased dependence on the articulated demand.

Economization as a description for long-term transformation processes in a variety of hitherto state-regulated policy areas such as education, health, or social protection (Höhne 2015) has become particularly prevalent in adult and continuing education and training since the 1990s (e.g.: Field 1994; Meisel 2008). The gradual withdrawal of the state from financial and political responsibility for ACE is supported by the ascendancy of the lifelong learning formula, which focuses on the learner’s responsibility for his or her employability and prosperity. At the same time, the introduction of global education markets (Komljenovic/Robertson 2017) and the privatization of educational goods and services under the General Agreement on Trade in Services (GATS) have added to further marketization in adult education (Lohr et al. 2013: 171-176; Meisel 2008). Turning to the case of Germany as an example, a recently published report reveals that in ACE private funding accounts for

over three-quarters of the total (Dobischat et al. 2019: 19)³ and that public spending for ACE in relation to the GDP, unlike all other educational sectors, has decreased during the past twenty years: from 0.32 percent of the GDP in 1995 to 0.21 percent in 2015 (ibid.: 26; 33).

The gradual withdrawal of the state and the increase of marketization has also pushed the systematic application of managerial and economic principles into adult education, particularly in the field of continuing vocational training: Terms like “supply and demand”, “competition”, “service provision” and “consumer protection”, which were for a long time absent from the adult education discourse, are now also used by publicly-funded providers (Meisel 2008: 242-243). ACE providers find themselves confronted with efforts to maximize their market share, even in the non-profit sector (ibid.: 245-247). Moreover, in the New Public Management paradigm, standards-based accountability (Green 2013: 204; 211), regulatory and control procedures grounded in performance indicators and external monitoring as well as audit and evaluation practices seem to apply both to public and commercial providers (Vater 2017; Lohr et al. 2013: 179-182).

The emphasis on market principles and economic rationality fuels conflicting arguments as to whether adult learning is a private or a public good (Ilieva-Trichkova/Boyadjieva 2018; Knauber/Ioannidou 2017).

In such a heterogeneous and scarcely regulated field and in the absence of a strong political agenda at state level, ACE providers oscillate between state regulation and market principles, whereas non-state and private actors both at national and international level are gaining regulatory power. We unfold this proposition in the following section by focusing on one actor that has internationally become quite influential in assuring quality and setting standards in non-formal education: the International Organization for Standardization (ISO). We draw on sociological neo-institutionalism to explain how the expansion of shared quality standards amongst ACE providers relates to the emergence of alternative regulatory forms beyond the state.

3. Regulation in a contested space: standard setting through QMSs

In the contested space of ACE regulation, the absence of a sovereign authority and common standards imply that (international) private organizations have considerable scope for setting the agenda. We argue that such an agenda setting function within ACE takes place through the implementation of

3 Indirect public funding through tax relief of private spending on ACE by individuals or firms has not been taken into account (Dobischat et al.: 19).

quality standards. Whilst measuring as well as displaying quality have become a central focus within the educational debate in general, in the context of ACE, quality management plays an important role especially due to the previously explained shift towards a more market-regulated field (Dollhausen 2008: 272; Käpplinger/Reuter 2017). QMSs aim at securing and improving the quality of processes within the organizations providing ACE. Besides, they have a signaling function because they disclose the provider's quality to other relevant actors in the organization's environment (Hartz 2008: 251-252) – such as potential learners, other ACE providers or financing bodies (Käpplinger 2017).

Against the background of this signaling function and according to Schrader's (2014) assumption that ACE providers can be classified by the conditions through which they obtain their resources and legitimization, QMSs can be understood as one possibility to secure and enhance these goods (Hartz 2009). Drawing on the example of Germany, the state has been actively engaged in promoting the introduction of QMSs by requiring certification when providers apply for certain public funding (Aust/Schmidt-Hertha 2012: 44; Ambos et al. 2018: 11-12).

Whilst some quality standards are only adapted at regional or national level, others have a global impact. For example, ISO, a non-governmental organization created in 1947 with the aim "to facilitate the international coordination and unification of industrial standards" (<https://www.iso.org/about-us.html>) has been engaged in the issue of developing global standards for (non-formal) education and training services, including adult education and training, since 2007.⁴ The underlying idea is that a universal set of standards helps safeguard quality across all types of non-formal education (Lynch 2009). One basic assumption is that education (here: ACE), regardless of who provides it, could be developed using the same tools, optimized in terms of efficiency, and evaluated against common standards (Höhne 2015: 27-28). In this context, the quality assurance discourse, imported from business administration and based on accountability and efficiency, plays a crucial role in the transfer of management principles to education organizations (ibid.).

ISO Standards⁵ claim to support providers of learning services to undertake quality assurance measures on a voluntary basis. The application of internationally recognized quality assurance systems promises significant competitive advantages in an increasingly internationalized education market.

4 Interestingly enough the idea to establish a technical committee (TC 232) within the ISO structure to deal with education and learning services came from Germany; in 2018 the ISO/TC 232 expanded its scope and changed its title from "Learning services outside formal education" to "Education and learning services", thus, covering the formal education sector as well (<https://www.iso.org/committee/537864.html>).

5 ISO Standards have been developed for instance for management systems of learning service providers, for distance learning services, for language learning services, and for educational assessment.

A few examples illustrate the international scope of diffusion regarding the implementation of ISO⁶:

- In Singapore, ISO standard 29990 (“Learning services for non-formal education and training – basic requirements for service providers”) is part of the formal requirements for the licensing of providers of training and continuing education since July 2017.
- ISO standard 29991 (“Language learning services outside formal education-requirements”) is part of the VET system in India; it is applied in the National Vocational Training Institutes and Advanced Training Institutes.
- In Germany, the percentage of adult learning providers who have at least one QM certificate was 80 percent in 2017 (Ambos et al. 2018: 12). The trend is towards multi-certification, which involves significant costs for the providers (Käpplinger/Reuter 2017: 16). ISO certification is by far the most widely used certification procedure in Germany. More than one third (35 percent) of all adult and continuing education providers in 2017 applied DIN EN ISO 9000ff⁷ (Ambos et al. 2018: 14).
- With 63 percent, DIN EN ISO 9000ff was particularly widespread in Germany among providers working predominantly for employment agencies or job centers – most of which are non-profit or commercially oriented private providers (ibid: 14-15).

Once established, the expansion of ISO standards may have serious implications. Whereas accountability and standards-based mechanisms allow for greater transparency, they also lead to a more centralized and unified control in defining the standards against which learning offers and learning outcomes are evaluated across countries regardless of the agent or location of provision.

How can this degree of expansion be possible, taking into account that ISO Standards are adopted on a voluntary basis? The theoretical perspective of sociological neo-institutionalism (here focusing on Di Maggio/Powell 1983) provides explanations to understand the expansion of shared expectations and the development of similar structures amongst different organizations, drawing attention to the interrelations of organizations and their environment. One basic assumption is that organizations operate in a field consisting of other relevant actors such as “key suppliers, resource and product consumers, regulatory agencies and other organizations that produce

6 Information is based on internal ISO documents (ISO/TC 232 N 370), if not otherwise indicated.

7 DIN EN ISO 9000ff standards form the process-oriented basis of quality management and are not geared to individual products, services or manufacturing methods. They enable standardization of work processes in a range of branches (from manufacturing to education and health services), both on a national and international level.

similar services or products” (ibid: 148) and that within these fields shared central concepts of society exist and influence social actions.

Sociological neo-institutionalism assumes that in a shared organizational field dynamics emerge amongst the actors that cause them to develop similarities to each other (ibid: 148). This process of increasing homogeneity is explained by the concept of institutional isomorphism and comprises three types that we outline briefly referring to DiMaggio & Powell (1983: 148-153): First, coercive isomorphism concerns structural homogeneity emerging due to pressures that occur as constraints, for example (but not only) through legal requirements. Second, mimetic processes take place especially in uncertain and ambiguous situations that may cause organizations to copy other organizations in aspects that appear to promise success. Third, normative pressures arise from shared professional understandings and normative rules, thus leading to similarities amongst different organizations.

The expansion specifically of ISO Standards following the dynamics of isomorph processes has been the subject of empirical studies in industry (Walgenbach/Beck 2003). The results have pointed out that organizations in industry initially decide to implement ISO standards primarily owing to the above-mentioned signaling function of QMS, thus contributing to structural isomorphism, rather than relating to ambitions of improving internal processes (ibid.: 503-505). Not focusing on ISO but on quality standards in general, for example Seyfried et al. (2019) have argued that the adoption of quality standards in higher education in Germany shows isomorphism. Hartz (2009) analyses the legitimating motives of ACE providers for adopting a German QMS developed specifically for the needs of ACE providers. Meeting the requirements for funding or of legislation (coercive isomorphism), copying the trend in the organizational field to show that a provider can meet the challenges of the market (mimetic isomorphism) and corresponding with professional standards in the field (normative pressure) are only some of the indicators revealing isomorph processes (ibid.: 144-146). The results suggest that isomorph processes occur in the field of ACE, and that they can play a role regarding competition over legitimacy and resources through the adaption of quality standards (ibid.). Against this background, the international adaption of ISO Standards in the field of ACE goes beyond an *expansion* of shared standards. Rather, this expansion can be characterized as a regulatory process that defines which standards are relevant for obtaining resources and legitimacy. In a contested space of weak government regulation, private international organizations like ISO gain scope to take over standard setting functions. The transnational certification of educational processes and services thus becomes a new regulatory form.

Whilst we have addressed issues of standardization primarily from a macro perspective, organizational research focusing on the *internal* processes within organizations in the field of ACE emphasizes the autonomy and self-

regulation of ACE providers. Research, for example, on the inner-organizational processes of ACE providers dealing with the simultaneity of pedagogical and economic premises points out that even under similar conditions for obtaining resources and legitimacy, organizations show a range of variation regarding their self-description, specific internal processes and decisions on their program planning (Dollhausen 2016: 244-247). Against this background, we conclude that standardizing processes take place whilst, at the same time, ACE providers have opportunities to preserve the heterogeneity of their internal processes and specific nature. Organization studies regarding the similarities arising among higher education organizations due to global standardization of quality criteria have also emphasized that local orders remain relevant, thus concluding that “[s]tandardization does not imply homogeneity” (Paradeise/Thoenig 2013: 215). These considerations make clear that greater standardization through isomorphism does not preclude differentiation. It is therefore necessary to take the internal *and* external conditions of ACE providers into account when empirically researching the dynamics of standardization.

4. Conclusion

In this paper, we have discussed standardization and economization in a heterogeneous and scarcely regulated educational sector. The gradual withdrawal of the state and the shift towards more market-type regulation in adult and continuing education leaves regulatory room to non-state actors. We pointed out the agenda setting capacity of an international private actor, the International Organization for Standardization, in the context of quality assurance. Our discussion demonstrated that providers in a contested space of weak authority are likely to develop alternative strategies of standard setting through the (mostly voluntary) adaption of quality standards in order to secure legitimacy and resources. To conclude, we bring our arguments into perspective with some critical comments and outline questions for further research.

Due to its analytical interest in the regulatory impact of standardization and economization on *organizations* in adult and continuing education, our paper so far has not yet reflected on the critical consequences that emerge *for adult learners* with regard to equality of opportunities. In an increasingly competitive global education market, the privatization of costs becomes more apparent, and inequality increases. Moreover, economic stagnation and austerity policies as imposed in some European countries affect adult learning opportunities and lead to significant cuts to publicly funded adult education programs (James/Boeren 2019: 8-9). These cuts as well as market-driven

orientation in adult education dismantle equality of access to publicly funded education, despite research findings pointing out that investment in providing opportunities for adult learning leads to increased participation, increased skills and improved employability; this also applies for focused investment on hard-to-reach groups (European Commission/ICF 2015: 46-47). In addition, data from the European Labor Force Survey show that there is a positive correlation between the level of public expenditure on education and participation in adult and continuing education, highlighting that the disadvantages of the low-skilled population decrease with increasing educational expenditure (Martin/Rüber 2016). Also, due to the dominance of economic rationality, traditional functions of adult education aiming to promote democratic citizenship and to compensate for educational inequalities such as civic education or basic adult education, are in danger of being side-lined in the contemporary discourse.

For further research, three main perspectives derive from our analysis: *First*, the discussion has revealed processes of standard-setting at an international level, thus calling for empirical research on legitimating motives as well as provider-specific strategies when referring to international standards. *Second*, our discussion has pointed to questions regarding the internal developments of ACE providers, i.e., the consequences for their organization and management. It might thus be necessary to probe the challenges they face by (not) opening up to (international) standards. One of these challenges emerges from the circumstance that meeting up to standards primarily addressing ACE providers' *organizational* processes leaves open the effects for the micro-didactical pedagogical quality of learning offers (Hartz/Meisel 2011: 103-104). *Third*, focusing on education provision includes asking how "the market-dependent production and distribution influence[s] the very service (the education provision) itself" (Fejes/Olesen 2016: 147). This also calls for enquiring which groups are especially dependent on high-quality publicly funded ACE programs, and, thus, are most affected by cuts in public funding and increasingly demand-driven ACE provision.

References

- Ambos, Ingrid/Koscheck, Stefan/Martin, Andreas/Reuter, Martin (2018): Qualitätsmanagementsysteme in der Weiterbildung. Ergebnisse der wbmonitor Umfrage 2017. Bonn.
https://wbmonitor.bibb.de/downloads/Ergebnisse_20180507.pdf [Last accessed 25 May 2019].
- Aust, Kirsten/Schmidt-Hertha, Bernhard (2012): Qualitätsmanagement als Steuerungsinstrument im Weiterbildungsbereich. In: Report – Zeitschrift für

- Weiterbildungsforschung 2, pp. 43-55. <https://www.die-bonn.de/doks/report/2012-weiterbildungssystem-01.pdf> [Last accessed 25 May 2019].
- Desjardins, Richard (2017): *Political Economy of Adult Learning Systems*. London: Bloomsbury.
- DiMaggio, Paul J./Powell, Walter W. (1983): The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. In: *American Sociological Review*, 48, 2, pp.147-160.
- Dobischat, Rolf/Münk, Dieter/Rosendahl, Anna (2019): *Weiterbildungsfinanzierung in Deutschland 1995-2015. Aktueller Stand, Entwicklungen, Problemlagen und Perspektiven*. Bericht der Bertelsmann Stiftung. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publicationen/GrauePublicationen/LL_Hintergrundstudie_Weiterbildungsfinanzierung1995-2015.pdf [Last accessed 10 April 2019].
- Dollhausen, Karin (2016): Wandel oder Kontinuität institutioneller Ordnungen im Weiterbildungssektor? Zur Bedeutung von organisationsspezifischen „Planungskulturen“. In: Leemann, Regula Julia/Imdorf, Christian/Powell, Justin J. W./Sertl, Michael (eds.): *Die Organisation von Bildung. Soziologische Analysen zu Schule, Berufsbildung, Hochschule und Weiterbildung*. Weinheim: Beltz Juventa, pp. 233-249.
- Dollhausen, Karin (2008): Steuerung der Angebotsentwicklung in Organisationen der Weiterbildung – Zur Bedeutung von Planungskulturen. In: Hartz, Stefanie/Schrader, Josef (eds.): *Steuerung und Organisation in der Weiterbildung*. Bad Heilbrunn: Klinkhardt, pp. 271-291.
- European Commission/ICF (2015): *All in depth analysis of adult learning policies and their effectiveness in Europe*. Luxembourg: Publication Office of the European Union.
- Ewert, Benjamin (2009): Economization and Marketization in the German Healthcare System: How Do Users Respond? In: *German Policy Studies*, 5, 1, pp. 21-44.
- Fejes, Andreas/Salling Olesen, Henning (2016): Editorial. Marketization and commodification of adult education. In: *European Journal for Research on the Education and Learning of Adults*, 7, 2, pp. 146-150. URN: urn:nbn:de:0111-pedocs-125339.
- Field, John. (1994): Market forces and continuing education. In: Lenz, Werner (ed.): *Modernisierung der Erwachsenenbildung*. Wien: Böhlau, pp. 85-102.
- Green, Jane (2013): *Education, Professionalism and The Quest for Accountability. Hitting the Target but Missing the Point*. New York: Routledge.
- Hartz, Stefanie (2009): Diffusionsprozesse in der Weiterbildung – eine Analyse aus der Perspektive des Neo-Institutionalismus. In: Koch, Sascha/Schemmann, Michael (eds.): *Neo-Institutionalismus in der Erziehungswissenschaft. Grundlegende Texte und empirische Studien*. Wiesbaden: Springer, pp. 133-159.
- Hartz, Stefanie (2008): Steuerung in und von Organisationen der Weiterbildung durch den Qualitätsdiskurs. In: Hartz, Stefanie/Schrader, Josef (2008): *Steuerung und Organisation in der Weiterbildung*. Bad Heilbrunn: Klinkhardt, pp. 251-270.
- Hartz, Stefanie/Meisel, Klaus (2011): *Qualitätsmanagement*. 3rd edition. Bielefeld: Bertelsmann.
- Höhne, Thomas (2015): *Ökonomisierung und Bildung. Zu den Formen ökonomischer Rationalisierung im Feld der Bildung*. Wiesbaden: Springer Essentials.

- Ilieva-Trichkova, Petya/Boyadjieva, Pepka (2018): Adult education as a common good. Conceptualisation and measurement. In: *International Journal for Lifelong Education*, 37, 3, pp. 345-358.
- International Organization for Standardization (ISO) (no date): No Title. <https://www.iso.org/about-us.html> [Last accessed 23 April 2019].
- International Organization for Standardization (ISO) (no date): ISO/TC 232. Education and learning services. <https://www.iso.org/committee/537864.html> [Last accessed 28 April 2019].
- Ioannidou, Alexandra (2014): The Adoption of an International Education Policy Agenda at National Level: Conceptual and Governance Issues. In: Zarifis, George/Gravani, Maria N. (eds.): *Challenging the 'European Area of Lifelong Learning': a Critical Response*. Dordrecht: Springer, pp. 203-215.
- Ioannidou, Alexandra (2007): Comparative Analysis of New Governance Instruments in Transnational Educational Space – Shift to Knowledge-based Instruments? In: *European Educational Research Journal*, 6, pp. 336-347.
- James, Nalita/Boeren, Ellen (2019): Adult Education in Austere Times: An Introduction. In: Boeren, Ellen/James, Nalita (eds.): *Being an Adult Learner in Austere Times. Exploring the Contexts of Higher, Further and Community Education*. Cham: Palgrave Macmillan, pp. 1-19.
- Käpplinger, Bernd/Reuter, Martin (2017): Qualitätsmanagement in der Weiterbildung. WISO Diskurs, 15. Bonn: Friedrich-Ebert Stiftung. <http://library.fes.de/pdf-files/wiso/13905.pdf> [Last accessed 18 April 2019].
- Käpplinger, Bernd (2017): Qualitätsentwicklung in der Weiterbildung: Fluch oder Segen? Eine mikropolitische Analyse. In: Bolder, Axel/Bremer, Helmut/Epping, Rudolf (eds.): *Bildung für Arbeit unter neuer Steuerung*. Wiesbaden: Springer Fachmedien, pp. 167-184.
- Knauber, Carolin/Ioannidou, Alexandra (2017): Grundbildungspolitik im Wohlfahrtsstaat: Akteure, Inhalte und Umsetzungsstrategien im internationalen Vergleich. In: Schmid, Josef/Amos, Karin/Schrader, Josef/Thiel, Ansgar (eds.): *Governance und Interdependenz von Bildung. Internationale Studien und Vergleiche*. Baden-Baden: Nomos, pp. 87-110.
- Komljenovic, Janja/Robertson, Susan Lee (2017): Making global education markets and trade. In: *Globalisation, Societies and Education*, 15, 3, pp. 289-295. doi: 10.1080/14767724.2017.1330140.
- Lohr, Karin/Peetz, Thorsten/Hilbrich, Romy (2013): *Bildungsarbeit im Umbruch. Zur Ökonomisierung von Arbeit und Organisation in Schulen, Universitäten und in der Weiterbildung*. Hans Böckler Stiftung. Berlin: Edition sigma.
- Lynch, Douglas E. (2009): ISO and the International Standardization of Adult Education. In: Strohschen, Gabriele (ed.): *Handbook of Blended Shore Education – Adult Program Development and Delivery*. Boston: Springer, pp. 63-70.
- Martin, Andreas/Rüber, Ina E. (2016): Die Weiterbildungsbeteiligung von Geringqualifizierten im internationalen Vergleich – Eine Mehrebenenanalyse. In: *Zeitschrift für Weiterbildungsforschung – Report*, 39, 2, pp. 149-169. <http://link.springer.com/article/10.1007/s40955-016-0060-2> [Last accessed 28 April 2019].

- Meisel, Klaus (2008): Steuerung öffentlicher Weiterbildungsorganisationen bei paradoxen Anforderungen. In: Hartz, Stefanie/Schrader, Josef (eds.): *Steuerung und Organisation in der Weiterbildung*. Bad Heilbrunn: Klinkhardt, pp. 233-250.
- Meyer, John/Ramirez, Francisco (2003): *The World Institutionalization of Education*. In: Schriewer, Jürgen (ed.): *Discourse Formation in Comparative Education*. Bern: Peter Lang GmbH, Internationaler Verlag der Wissenschaften, pp. 111-132.
- Paradeise, Catherine/Thoenig, Jean-Claude (2013): *Academic Institutions in Search of Quality: Local Orders and Global Standards*. In: *Organization Studies*, 34, 2, pp. 189-218.
- Schemmann, Michael (2014): *Handlungskoordination und Governance-Regime in der Weiterbildung*. In: Maag-Merki, Katharina/Langer, Roman/Altrichter, Herbert (eds.): *Educational Governance als Forschungsperspektive. Strategien. Methoden. Ansätze*. Wiesbaden: Springer VS, pp. 109-126.
- Schimank, Uwe/Volkman, Ute (2012): *Economizing and Marketization in a Functionally Differentiated Capitalist Society – A Theoretical Conceptualization*. Welfare Societies Conference Paper 2012.
https://pure.mpg.de/rest/items/item_1567285/component/file_1567283/content.
 [Last accessed 24 April 2019].
- Schrader, Josef (2014): *Strategies of modernization and their effects on configurations of adult education. Theoretical assumptions and empirical findings*. In: Käßpinger, Bernd/Robak, Steffi (eds.): *Changing configurations in adult education in transitional times. International perspectives in different countries*. Frankfurt am Main: Peter Lang GmbH, Internationaler Verlag der Wissenschaften, pp. 57-71.
- Seyfried, Markus/Ansmann, Moritz/Pohlenz, Philipp (2019): *Institutional Isomorphism, entrepreneurship and effectiveness: the adoption and implementation of quality management in teaching and learning in Germany*. In: *Tertiary Education and Management*, 25, pp. 1-15.
<https://doi.org/10.1007/s11233-019-09022-3> [Last accessed 25 March 2019].
- Vater, Stefan (2017): *Mehr Markt, mehr Management und alles wird (wieder) gut? In: Magazin Erwachsenenbildung.at*, 32, 4.
<http://www.erwachsenenbildung.at/magazin/17-32/meb17-32.pdf> [Last accessed 18 April 2019].
- Walgenbach, Peter/Beck, Nikolaus (2003): *Effizienz und Anpassung. Das Erklärungspotenzial der neo-institutionalistischen Organisationstheorie am Beispiel ISO 9000*. In: *DBW*, 63, 5. Stuttgart: Schäffer-Poeschel Verlag für Wirtschaft·Steuern·Recht GmbH, pp. 497-515.

VI. Challenges of Translation in Educational Research

Section Editors:

Norm Friesen, Boise State University

Rose Ylimaki, Northern Arizona University

The Necessity of Translation in Education: Theory and Practice

Norm Friesen¹

1. A missing dimension

Why translate in education? English-language educational research provides little evidence to suggest that translation is important or beneficial. If the object of such research is empirical and quantitative as it so often is then it is not even necessarily dependent on any single set of linguistic representations or language. It instead speaks through numbers, graphs and charts. In this context, translation appears as an exception, as a way of making contributions of uncommon figures like Paulo Freire or Jean Piaget accessible from Portuguese or French. This chapter argues, however, that the potential value of translation in education is not a matter of access or even of cross-cultural communication. Instead, it is arguably nothing less than a question of finding a language for education that is itself explicitly educational rather than one that is primarily psychological, sociological, philosophical or historical in nature.

In articulating this bold argument, this paper presents translation not as an exception for education, but as a central priority. It begins by explaining what is meant by a “language of education” that is explicitly educational and it continues by exploring how the educational vocabulary of the German language (together with its Scandinavian cousins) suggests what such a lexicon might look like. From there, it explicates a way of understanding translation (and the reading of translated texts) in terms of “alienness” and “ownness.” It also makes the case that translation can represent a way out of education as constructed both in and through English as the medium of a globally triumphant neo-liberal hegemony.

Speaking specifically of a “missing dimension” in Anglophone constructions of education, Gert Biesta writes: “One way to put it is to say that what is absent in the English-speaking world is the idea of a distinctively *educational* perspective on education” (Biesta 2015: 15; emphasis added).

When we look at education through the lens of what in the English-speaking world are known as the disciplines of education, we can say that the philosophy of education asks philosophical questions about education, the history of education asks historical questions, the psychology of education asks psychological questions and the

1 Norm Friesen is Professor in the Department of Educational Technology at the College of Education of the Boise State University Idaho. Email: normfriesen@boisestate.edu

sociology of education asks sociological questions, which then raises the question “Who asks the educational questions?” (Biesta 2015: 15).

But what exactly does it mean to ask educational questions about education? And what does this have to do with translation? Biesta explains that to understand education as specifically “educational” does not mean to adopt a particular (inter)disciplinary base of methods and knowledge. Instead, it means to embrace a particular passion or concern that provides the impetus for both educational research and practice. Biesta refers to this as

[...] the idea that there is such a thing as a distinctive educational interest, that is, a distinctive educational concern that provides a particular way of looking at and engaging with educational phenomena. This idea played a key role in the establishment of education as an academic discipline in the first decades of the 20th century, where proponents of what became known as “*geisteswissenschaftliche Pädagogik*” [human science pedagogy -NF] [...] established the discipline as what we might call an interested discipline [of pedagogy -NF] [...] that is, a discipline organized around a certain normative interest (Biesta 2015: 15).

And what is this founding interest, this normative concern or grounding that organizes the discipline and that underlies specifically *educational* questions about education? Biesta explains that it is one focused on the personal “emancipation of the child... [and the fact that] such emancipation was best served by an academic discipline that itself was emancipated from normative systems, such as the church and the state” (ibid.). Biesta’s characterization here is apparently based on two overviews of German *Erziehungswissenschaft* or *Pädagogik* from the 1970s (he cites Grootjohoff 1973 and Wulf 1978). However, increasingly, key passages from *geisteswissenschaftliche Pädagogik* have been appearing in English that expand upon Biesta’s paraphrase – including this bold characterization of the “new pedagogy” by the human science pedagogue Herman Nohl from 1926:

[The] basic stance or disposition of this new pedagogy is decisively characterized by the fact that its perspective is unconditionally that of the educand [or child -NF]. Its task, then, is not to act in service of objective powers, to draw the child towards the state, the church, law, the economy, towards a political party or an ideology that the child may be subjected to. Instead, it sees its goal in the subject [the child or young person -NF] and their physical and personal realization or unfolding. That this child here comes to his life’s purpose, that is the autonomous and inalienable task of the new pedagogy. This is what we call its autonomy, which equips it with a measure of independence from other cultural systems and [gives it -NF] the ability to observe them critically (Nohl 1926: 152, emphasis in the original).²

Besides buttressing Biesta’s observations about the emancipation of both the child and the discipline in human science pedagogy, Nohl is making a number

2 This quote, like others taken directly from the original German (as noted), is the author’s own translation. However, much of this passage (as well as others from Nohl’s writings) also appeared in Friesen 2017; further translated passages can be found in Horlacher 2016.

of additional points. First, he is saying that the principal interest of education is the child and his or her unfolding as a whole – extending beyond any narrow or political conception of “emancipation.” Second, Nohl is arguing that this interest (or in his own words, stance or disposition) is not directed to children in the abstract, but (inter)personally to “this child here,” in an engagement in which adult and child meet “each other” (Langeveld 1983: 6). Third, this concern is not just directed towards the children right now, to their current wishes and needs, but also to their future, to their “life’s purpose” or to who they will *become*.

It is precisely the translation of such theoretically and historically relevant passages and texts for which Gert Biesta advocates. But what is most important for this paper are the conclusions that Biesta arrives at about working across languages and traditions: Namely, his call for the English-speaking field of education to develop a particular kind of “academic bilingualism in education.” He adds: “The task of translation is, after all, never one of replacing words with other words but is about the transformation of one system of meaning into another system of meaning. It is a matter of semantics” (Biesta 2012: 21). It is, to summarize, an intricate semantic labor across systems and frames of meaning. And it is a kind of work, according to both Biesta and this paper, that is needed in contemporary educational thinking – whether it is explicitly recognized or not.

As mentioned, considering the challenges – and rewards – of this semantic work in the context of the literature and tradition of human science pedagogy (broadly defined) is precisely the focus of this paper. And the ultimate aspiration of such an effort is to contribute to something that can be cognitively very challenging, even for the translator him- or herself: This is the realization, as Wittgenstein noted, that the “*limits of my language* mean the limits of my world” (Wittgenstein 1974: 68; emphasis in original). It is the acknowledgement that one’s educational “world” can be considerably expanded by considering both carefully and sympathetically how it can be discussed and analyzed in other languages and linguistic traditions. In working towards this ambitious undertaking, I focus specifically on the theory and practice of translation within the human science pedagogical tradition referencing both Schleiermacher (1813/2012 who is its precursor) and Ricoeur (2004/2006; who developed the human sciences further in his own philosophy) to theorize the translator’s task. I begin by outlining the challenges inherent in translating some of the most basic terms used to talk about and analyze education in German and Scandinavian languages;³ I then discuss the challenges and possibilities presented by translation in theory and practice.

3 I say this while recognizing that Denmark, Finland, Norway and Sweden divide up the semantic field designed by “education” in slightly different ways. However, they all retain many of the key differentiations that are available in some form in German, but that do not exist in English.

2. Translating the basics: *Pädagogik*, *Erziehung*, *Bildung*, and *Didaktik*

In describing the German “language of education,” Biesta observes that

[w]hereas in the English language the word ‘education’ suggests a certain conceptual unity, the German language has (at least) two different words to refer to the *object* of study – ‘*Erziehung*’ and ‘*Bildung*’ – and (at least) two different concepts to refer to the study of *Erziehung* and *Bildung* – namely ‘*Pädagogik*’ and ‘*Didaktik*’ (Biesta 2011: 183).

Although three of these terms have ready equivalents in English (education, pedagogy and didactics), what is important to appreciate is that in each case, the German terms bring with them not only unique semantic fields, but also a particular tradition of landmark texts, interpretations and their history. Given Biesta’s (and the German language’s) ordering of these terms (into objects and their study), I begin this section by discussing education as an object of study, and end it with definitions of educational “phenomena” of *Bildung* and *Erziehung* themselves.⁴

Pädagogik has referred “from the earliest times to the teaching [*Lehre*] and the theory of human *Bildung* and *Erziehung*” (Böhm 2004: 750; emphasis in original). Biesta has already described *Pädagogik*, above, as a concept used to refer to the study of *Erziehung* and *Bildung*, and this primacy over other key terms in the German educational vocabulary is preserved in one *Historical Dictionary*, although it has been challenged in Germany since the 1970s.⁵ *Pädagogik*, both traditionally and today, however, refers to ways in which education and formation can be understood, both theoretically and pragmatically, in terms of what it is to educate and what it is for an individual to be formed (*Bildung*). This is clear from an 1876 definition quoted by Böhm:

4 In discussing the four terms – *Pädagogik*, *Didaktik*, *Erziehung* and *Bildung* – I rely heavily on separately authored contributions by Böhm, Wiggers, Oelkers, Benner and Brüggem from Benner and Oelkers’ dictionary, published in 2004, *Historisches Wörterbuch der Pädagogik*.

5 Following Biesta’s as well as Benner and Oelkers’ logic, one would expect faculties of education in Germany to be known as ones focusing on *Pädagogik*; however, this is not the case. Starting in the 1970s, these faculties adopted the name *Erziehungswissenschaft*, literally the science or study of education, although their individual departments still preserve reference to *Pädagogik* in their names to this day; e.g. *Sozialpädagogik* for social work or *Sonderpädagogik* for special education. Currently, in recognition that educational studies generally focus on development and social improvement in the broadest sense, some faculties are changing their titles from *Erziehungswissenschaft* to *Bildungswissenschaft*. Nonetheless, the word *Pädagogik* remains visible, for example, in the titles of introductions to the field of education.

Pädagogik does not just refer to the science (*Wissenschaft*) of education, but also includes the art of educating (*Erziehungskunst*), which arises from the fact that educational activity is not simply instinctual and habitual, but instead is anticipated in such a way that one proceeds from particular presuppositions, that one works towards a particular goal, and uses particular means to reach this goal, that in one word one is aware of particular basic principles (Baur, as quoted in Böhm 2004: 750).

Pädagogik is thus not necessarily entirely distinct from the broadest English sense of the term pedagogy – as “the art, science, or profession of teaching.” However, as Böhm’s quote (above) suggests, the German term also emphasizes a self-aware, reflective engagement in educative action and thought. Also, unlike the English term, *Pädagogik* cannot be reduced simply to an approach or method for teaching or instruction (e.g., choosing between or combining constructivist, critical and socio-cultural “pedagogies”).

Didaktik, as the second concept identified by Biesta to refer to the *study* of *Erziehung* and *Bildung*, has been defined as the “professional science of the teacher,” as “the study (and doctrine) of learning and teaching in general,” or as “the science of instruction” (Martial, as quoted by Wigger 2004: 245). In both everyday German and in specialized educational discourses, *Didaktik* refers to the study of teaching and instruction as largely instrumental processes – ones aimed at promoting *learning*. In this sense, English usage of the term “pedagogy” (e.g. in speaking of “constructivist” or “critical pedagogies”) is actually closer to the German *Didaktik* than to its cognate, *Pädagogik*. However, some English readers may know the term *Didaktik* in a different context; from the fairly steady stream of English-language publications comparing *Didaktik* with American curriculum studies published since the late 1990s (e.g., Westbury/Hopmann/Riquarts 1999: ix; Gundem/Hopmann 1998; Autio 2006; Uljens/Ylimaki 2017; Friesen 2018). In this context, *Didaktik* refers not so much to the study of teaching and learning as it does to a rather broad German and European tradition that goes by the same name, and that can be traced back to Comenius’ *Didactica Magna* (1659). It also refers to relatively contemporary ways of thinking that are frequently seen as having “a complex relation” to the study of curriculum in America (Westbury/Hopmann/Riquarts 1999a: ix). Just as curriculum in this context does not simply refer to literal school curricula and lesson plans, but to ways of understanding means-ends instrumentality in education,⁶ so *Didaktik* in this sense refers to an expansive tradition, reaching as far back as Johann Amos Comenius (1592-1670). This European tradition (which, as suggested, extends well beyond Germany and Scandinavia) is one that connects ideas and practices of instruction to understandings of what it is to be and become human, initially in religious terms, and later in ones more secular. In a more

6 E.g., see: Pinar, William (ed; 1975). Curriculum studies: The reconceptualization. Troy NY: Educator’s International Press.

practical sense, it is one that has also connected “teacher education, schooling and the teaching profession” (Westbury/Hopmann/ Riquarts 1999: 4) in ways that grant (but also guide) teacher autonomy in interpreting and transforming curricular prescriptions into instructional and classroom practices.

Definitions of *Erziehung* differ in clear and important ways from the ways that *education* is commonly defined in English. Here, I reference the entry provided in Böhm and Seichter’s Dictionary of Pedagogy (*Wörterbuch der Pädagogik*) from 2018, which defines education as both

a process and its result, an *intention* as well as actions (of the educator and the *educandus*), the situation of the child and the conditions that constitute it... [It can -NF] describe a particular class of activities, and [thus function as -NF] a descriptive-analytical concept, while at the same time offering criteria for particular activities, and thus [also work as a -NF] *normative* concept (p. 358; emphases added).

Erziehung, in other words (and as Biesta has pointed out earlier), is recognized in the German context in terms of a particular concern, an interest or intention that is (at least in part) *normative* in nature. It is in this sense that as both Biesta as well as Böhm and Seichter all suggest, the word *Erziehung* offers “critieria” and a “normative concept” – allowing us, for example, to designate some questions as educational and others as clearly not. Other definitions highlight various points of emphasis, but ultimately lead to a similar focus on the idea of an interest and intention. Betraying his Deweyan inclinations, Jürgen Oelkers (2004) for example speaks of *Erziehung* in terms of “moral communication,” a kind of communication that aims at “lasting influences and that presupposes a gap or lack” which these influences are to address (p. 303); and Wolfgang Brezinka, despite his aim to recast German educational studies as a positivist psychological enterprise, still defines *Erziehung* in terms of a normative influence that “seeks to improve the structure of psychological dispositions of another person” (as cited in Oelkers 2004: 339). To return to Biesta’s point, it is in all of these senses that in Germany and Scandinavia one is able to speak of education as a discipline defined by a *normative interest* to this day. Brezinka describes this normative influence when he speaks of “improving the [...] psychological dispositions of another person,” and as I’ve shown above, less positivistic approaches have interpreted this as the child’s autonomy and well-being. It is also in terms of this interest that one can also ask specifically “educational” questions about education *itself*. One can ask, for example, whether some experiences are actually educative (as Dewey 1938 does), for example, rather than being satisfied with the claim that we learn through all experience, regardless of its nature. Such questions, in short, are about how to do right by what is best for the child, both as expressed in the present and as anticipated for the future.

English definitions of the cognate term *education*, by contrast, are either flatly empirical, or focus on education as a kind of “ideal” attainment or state

of affairs. For example, the *Oxford English Dictionary* (2000) defines education as “the systematic instruction, teaching, or training in various academic and non-academic subjects given to or received by a child, typically at a school,” while *Merriam Webster* characterizes it as “the field of study that deals mainly with methods of teaching and learning in schools.” Alternatively, figures like R.S. Peters in the UK have defined education largely in terms of what it means to have been “educated” (e.g. see: Barrow 2014: 256-259), resulting in characterizations actually much closer to the German term *Bildung* (see below). Of course this also implies that such “idealizations” are rather distanced from everyday uses of the term “education” in phrases like “studying education” or “high school education.”

Bildung, finally, has no obvious English-language substitute. It has consequently been translated variously as education, edification, learning, culture, cultivation and literacy (Friesen 2007: 84-85). It was given canonical definition by Wilhelm von Humboldt as “the linking of the self to the world to achieve the most general, most animated, and most unrestrained interplay” (Humboldt 1792/1999: 58). In keeping with the breadth of this phrasing, Benner and Brüggem (2004) define *Bildung* as “the process of the forming (*die Formung*) of humans, as well as the determination (*Bestimmung*) of the goal and purpose of human existence” (Brenner/Brüggem 2004: 174) – further underscoring the vast, ill-defined semantic space that this term occupies in the German language.

Such lofty connotations, combined with *Bildung*’s general untranslatability, can be said to have led to considerable interest and also distortion in recent English-language scholarship. In English, *Bildung* is frequently defined in terms of specific, intellectual-historical moments, such as bourgeois and reactionary impulses in the German *fin-de-siècle* and Weimar periods (e.g. Pinar 2011: 2-5), of the gendered character of its neo-humanist constructions (Baker 2001, borrowing from Kittler 1983/1990), or of the goals of *Bildung* and the danger of their contemporary commodification (Autio 2003).⁷ Such narrow and largely critical accounts are then too often taken in English as interpretations of *Bildung* writ large. Although they might have been at home in the context of the critical theory dominant in West-Germany in the 1970s or 1980s,⁸ such definitions would be uncommon in both academic and quotidian German use today. In these living contexts, *Bildung* is not so much *confined* by its history and polysemy as it is *enabled* by them

7 These rather partial interpretations limit and reify the term *Bildung*, locating it in one or another historical situation, rather than reflecting multiple historical strands and descriptive accounts that are at play in its actual use. With the exception of Autio, these interpretations are based exclusively on English-language translations and accounts of German intellectual developments of relevance to *Bildung*.

8 In the 1970s and 1980s, studies critical of *Bildung* were undertaken by the likes of Kittler (e.g., 1983/1990), but especially by those developing concepts and structures from the Frankfurt School (e.g., Gernot Koneffke, Heinz-Joachim Heydorn).

(representing an exception to many historical German terms⁹). As Rebekka Horlacher points out, *Bildung* today is above all marked by a kind of semantic excess, a surplus of meaning that “transcends mere utility.”

In addition, *Bildung* signifies the ideal of the autonomous, self-determined, and self-reflected personality in its full realization, a “becoming [of] oneself [...]” [It -NF] signifies something that cannot be completely contained by terms such as “education,” “socialization,” “instruction,” or “schooling.” [But at the same time, it -NF] [...] signifies the aspiration of perpetual self-improvement in this life. It represents an unquantifiable excess value that [still -NF] ought to be administered [or realized -NF] in schools or at universities (Horlacher 2016: 1).

Bildung, in other words, identifies a kind of “becoming human” that spans biographical, collective, institutional and historical dimensions. As such it opens up the possibilities of a generative process through which we are formed by the world, form ourselves, and form the world (immediately) around us. As both fact and aspiration, Horlacher concludes, *Bildung*, in its broadly Humboldtian sense, still “sets the standard for today’s education policy issues” in German-speaking Europe (Horlacher 2016: 125).

3. Translation: an impossible task?

The four examples of *Pädagogik*, *Didaktik*, *Erziehung* and *Bildung* highlight the difficulty and complexity entailed in the kind of translation called for by Biesta – underscoring that it is never the simple replacement of one word with another, but rather, “the transformation... of one system of meaning into another system of meaning” (Biesta 2012: 21). In his famous text “On the Different Methods of Translating,” Friedrich Schleiermacher puts this as follows:

any language, despite the different concurrently and consecutively held views expressed in it, encompasses within itself a single system of ideas, which, precisely because they are contiguous, linking and complementing one another within this language, form a single whole, whose several parts, however, do not correspond to those to be found in comparable systems in other languages (Schleiermacher 1815/2012: 59-60).

The validity of this characterization is clear from the relatively confined system formed by the words *Pädagogik*, *Didaktik*, *Erziehung* and *Bildung* as just discussed: Their meanings are all interrelated, but as soon as these relationships are discussed, questions of connotation, polysemy, hierarchy, past versus contemporary meanings and more all crop up. And these can be

9 E.g., words like *Zucht* (discipline, breeding, guidance), *Geist* (spirit, mind) and *Volk* (people, nation).

addressed only by pointing to varying, sometimes contradictory realities, possibilities and ambiguities. Adequately capturing these in another language – as the case of *Bildung* highlights – sometimes seems impossible. Indeed, in his 2006 book *On Translation*, Ricoeur observes that translation, at least in theory, appears as an “impossible task” (2006: 13), and Schleiermacher, following analogous reasoning, similarly characterizes translation as an “utterly foolish undertaking” (Schleiermacher 1915/2012: 47).

One of the key structural, perhaps even ontological, challenges in translating any given term in German or English has to do with our fundamental confinement in a given language. In discussing the viability of any one translated text or term, one can do so only in one language at a time – either in the original or in the target language. And each language, of course, brings its own systems, shadings and histories which are often in no way aligned, but generally relate to those in another language only indirectly or orthogonally. As if to drive the impossibility of translation home, Paul Ricoeur characterizes this limitation as follows:

[...] there is no absolute criterion for good translation; for such a criterion to be available, we would have to be able to compare the source and target texts with a third text which would bear the identical meaning that is supposed to be passed from the first to the second (Ricoeur 2006: 22).

Translation offers us no such neutral third, whether it be a third text or third language, from which to understand and judge a translation as such. One can either work “within” one language or another. And generally, people cannot do both at the same time. Schleiermacher, sounding rather post-structuralist, also adds that a person is inevitably “in the power [*Gewalt*; also force or violence] of the language he speaks [...] he and all his thought are its products” (Schleiermacher 1915/2012: 46). Schleiermacher then goes on to explain that in this situation, the translator must humbly serve two masters, the author and the reader, and that in so doing, the translator can only choose between two possibilities: 1) demand more of the reader, and provide a relatively strict translation of the author, “retaining the feel of the alienness” from the author’s ideas and writing in the original; or: 2) demand more of the *author* – namely that their idiosyncrasies and foreignness be rendered in the familiar structures and turns of phrase of the reader’s tongue: “Either the translator leaves the writer in peace as much as possible,” Schleiermacher says, “and moves the reader toward him; or he leaves the reader in peace as much as possible and moves the writer towards him” (Schleiermacher 1915/2012: 46). The translator, in still other words, can stand in the original language of the author and reach out from this alien position to the reader, or take the position of the reader in *their* tongue and preserve from the author whatever can be rendered comfortably in *it*.

“Alien” or “alienness” are terms used regularly in Schleiermacher’s work “On the different Methods of Translating.”¹⁰ The mutually exclusive nature of “alienness” from what is familiar or what is “one’s own” has been analyzed more recently in broader terms in Bernhard Waldenfels’ “phenomenology of the alien.” As Waldenfels explains, things that are alien to us – “our” language versus another one, our culture compared to another, even our wakefulness versus the state of sleep – do not have an outside, a third position from which the two may be compared or evaluated. We can only “inhabit” one state or another (sleep or wakefulness), or one language and culture or another. A person cannot “act out” or “embody” both Japanese and American culture, habits and norms at the same time, just as most people (if they are bi- or multi-lingual) struggle to move from immersion in one language to another. To adopt Heidegger’s dictum that “language is the house of being” (Heidegger 1947/1993: 217) may be seen as cliché, but Heidegger’s words are certainly indicative of our dwelling in our one language or another.

Waldenfels’ two key terms, own and alien, name two separate “spheres” that are further separated, he adds, by a threshold.

The sphere of alienness is separated from my sphere of ownness by a threshold, as is the case for sleep and wakefulness, health and sickness, age and youth, and no one ever stands on both sides of the threshold at the same time [...]. There is no [...] cultural arbitrator to divide European and Far Eastern cultures from the outside, since Europeans must have distinguished *themselves* from Asians before such a division or comparison can be made. [...] the distinction between [...] ownness and alienness, cannot be reduced to two *terms*. Rather it refers to two different *topoi* (Waldenfels 2007: 7-8).

This relationship of own versus alien means that the two languages involved in any translation do not relate as the reversible, symmetrical “other” of each other. “My language” and a “foreign language” are not just different from each other, they are asymmetrical in their relation to “me,” and are, both in thought and experience, in many ways mutually exclusive. The one is “alien” to the other; its force or even “violence,” Waldenfels suggests, “arise from elsewhere” (Waldenfels 2007: 7). In keeping with Waldenfels’ (and Heidegger’s) characterization, the pragmatics of translation seem to be characterized very much by a type of labor and “dwelling,” first within the *topos* of one language – its particular force and demands – and then from within the *topos* of the other. Waldenfels explains that what is alien in this context is manifest not in some kind of “encounter” (as would be the case in engaging with an “other”), but through the alien’s *withdrawal*.

10 65 times to be exact.

In this author's experience,¹¹ translation begins with an attempt to simply capture, to whatever degree possible, the original (German) text in the target language, English. And at this beginning stage, to borrow Schleiermacher's characterization, the translator is "leav[ing] the writer in peace as much as possible and moves the reader toward" him or her (Schleiermacher 1815/2012: 49). This initial translation would confront any reader with all manner of foreignness – since it tends to reflect the grammar, idioms and other particularities of the author and his or her original language. The second stage, however, attempts to undo this. Here, the translator approaches the roughly-translated text from the world or *topos* of his or her own language or mother tongue, while not necessarily being directed or sustained by reference to the original. Here, the translator addresses points where the text takes the reader away from familiar idioms and turns of phrase, from words that may be correct in their denotation, but lead the mind astray through their connotative associations. In this case of German, of course, this is also where sentences must be rephrased to ensure that verbs and verbal phrases are rendered familiar to English-speakers in their formulation and positioning. The point, effectively, is at this stage to leave "the reader in peace as much as possible and mov[e] the writer towards him" (Schleiermacher 1815/2012: 49). The first and second stages are then repeated at least one more time, with the translator returning to the writer's original text and language and correcting the now "anglified" text on this basis, to ensure fidelity with the original: This "anglified" text is again checked and adjusted in an attempt to retain flowing and idiomatic English. The translator, in sum, goes back and forth across a "threshold" separating the two languages, inhabiting one while making the other alien. In this way, the translator works to present a text which is at once accessible, but also signals through degrees of unfamiliarity if not difficulty, that the reader is engaging with something different or alien. And in the case of German educational thought, engagement with the foreign is not just with one word or phrase in isolation, but with whole other systems of meanings and intellectual possibilities.

4. Conclusion: what is "most unlike ourselves"

Through the occupation of mutually alien *topoi* just described, a translated text can, however imperfectly, be oriented to the original author's alien world and language while not *unnecessarily* estranging the reader. However, the

11 E.g., in translating Mollenhauer, Klaus (2013): *Forgotten Connections: On Culture and Upbringing*. New York: Routledge. And co-translating Schleiermacher, Friedrich (forthcoming): *Outlines of the Art of Education, The 1826 Introductory Lecture*.

translator must still ask much of the reader, just as I have with the various “awkwardnesses” and “Germanicisms” in this chapter. As a translator, one has no choice but to stretch the possibilities of one’s own language and thus, to render it in some ways alien to the reader.¹² Again, in the vocabulary of both Schleiermacher and Waldenfels, this means cultivating a sensitivity for the “alien” in the reader, which as Waldenfels has said, is manifest only in its “withdrawal.” In the case of translation, this withdrawal can be said to happen, for example, through the reader’s distortion or diminution of the constitutive ambiguity of a word like *Bildung*. A similar withdrawal is also evident in the reader’s forgetting of the truncation of their sensitivity to emphasis, tone and style, their awareness of broader contexts of use that inevitably occurs when reading in translation. And as Gert Biesta indirectly suggests, such withdrawal may even result in the failure to sense that there might be something missing in one’s own constructions of education – both as a field and a practice.

Needless to say, there is much at stake in these moments of withdrawal. After all, the English language as the modality of our reading, reflection and expression is hardly an indifferent medium, an innocent embodiment of or messenger for educational thought and its dissemination. Besides its imperial history, English is currently *the* linguistic embodiment of a globally-triumphant neo-liberalism of benchmarks, testing and universal efficiency (e.g., see Phillipson 2010). Historically, English language *education* bears a particularly heavy burden – as a result of both its colonial and ongoing neo-colonial projects, and through its continued “colonization of the consciousness” of academia internationally, as Tsuda has noted (Tsuda 1997: 22). In this context, there is not a great deal of difference between moments of the withdrawal of alien meanings and possibilities for thought and their relegation to historical and cultural oblivion – a process not altogether different from the ongoing diminution of global linguistic diversity (e.g., Anderson 2012).

Returning to Wittgenstein, we can no longer be satisfied with the limits of our primary (or only) language being the effective limit of our world. Instead, what is important are not only redoubled efforts at translation, but a return to a particular kind of reading. As Schleiermacher observes, “if rules for [such reading -NF] are to be given, they would have to be such as to produce a purely moral state of mind [*sittliche Stimmung*] in which the spirit remains receptive even to that which is most unlike itself” (Schleiermacher 2004: 44). Now “*sittlich*” and “*Stimmung*” as readers of Hegel¹³ and Heidegger¹⁴ may

12 An example of this is provided by the term “alien” itself: “It doesn’t bring to mind the ontological ethics of Derrida or Levinas as much as it suggests invasive life-forms from another planet or undocumented crossings of borders” (Friesen 2014: 69).

13 For a discussion of *Sitte* and *Sittlichkeit* in general and in the works of Hegel, see: Carritt 1936.

know, each present their own challenges for translation. However, the point here is that in both a translation and its interpretation, we are always encountering that which is “most unlike ourselves.” In this context, we are called to sensitize ourselves, to be receptive or open, to even engage in a kind of passivity or submission to a text – an exercise in humility which is still sometimes seen as the hallmark of good reading (e.g., Gadamer 2006). However, to thus submit oneself, to engage in this passive reception, of course, is not to undertake some type of selfless sacrifice. This contact with the elusive alien does not mean stopping one’s own thought; indeed, as Waldenfels argues, exposure to the alien it is constitutive of the very movement of thought itself.

In this light, I hope this chapter has helped to show that what Schleiermacher wrote of his own early 19th century German language can apply to English as well: Namely, that “we must [...] realize that much in our language that is beautiful and strong was developed, or restored from oblivion, only through translation;” that even our own language “can most vigorously flourish and develop its own strength only through extensive contact with the alien” (Schleiermacher 2004: 52). At a time when the “medium” of our own language is the “message” of an educationally and environmentally destructive neo-liberal globalism, these possibilities are more important than ever.

The following contributions discuss the topic of translation from three perspectives. *Kathrin Berdelmann* deals with some of the challenges that arise in the translation of conceptual terms from German to English in historical educational research. Focusing on the pedagogy of Enlightenment in particular and tracing the various ways that these terms have been rendered in English, Berdelmann examines the cross-lingual derivation of the German terms *Bildsamkeit* and *Vervollkommenung*. *Inés Dussel* reflects on three aspects of translation in the context of academic practices in the social sciences and humanities. She emphasizes that translation is a material practice embedded in particular conditions – specifically within an academic geopolitics in which English is perforce becoming the international academic lingua franca. *Britta Upsing and Musab Hayatli* discuss the challenges faced in the translation of international large-scale educational assessments such as PISA (Programme for International Student Assessment) and PIAAC (Programme for the International Assessment of Adult Competencies). Using examples of real-life errors and challenges, Upsing and Hayatli provide an overview of methods used to ensure quality in translation and the special difficulties presented by tests for plurilingual populations.

14 For a discussion of the meaning of *Stimmung* in general, see: Krebs 2017: 1420.

References

- Anderson, Stephen (2012): *Languages: A Very Short Introduction*. Oxford: Oxford University Press.
- Autio, Tero (2003): Postmodern paradoxes in Finland: The confinements of rationality in curriculum studies. In: Pinar, William F. (ed.): *International handbook of curriculum research*. Mahwah NJ: Lawrence Erlbaum, pp. 301-328.
- Baker, Bernadette (2001): *In perpetual motion: Theories of power, educational history, and the child*. New York: Peter Lang.
- Barrow, R. (2014): Concept of Education. In: Philips, D.C. (ed.): *Encyclopedia of Educational Theory*. Thousand Oaks, CA: Sage, pp. 256-259.
- Benner, Dietrich/Brüggen, Friedhelm (2014): *Bildsamkeit/Bildung*. In: Benner, Dietrich/Oelkers, Jürgen (eds.): *Historisches Wörterbuch der Pädagogik*. Weinheim: Beltz, pp. 174-215.
- Biesta, Gert (2011): Disciplines and theory in the academic study of education: a comparative analysis of the Anglo-American and Continental construction of the field. In: *Pedagogy, Culture & Society*, 19, 2, pp. 175-192.
- Biesta, Gert (2012): Wanted, dead or alive: educationalists – On the need for academic bilingualism in education. In: Aubry, Carla/Geiss, Michael/Magyar-Haas, Veronika/Miller, Damian (eds.): *Positionierungen: Zum Verhältnis von Wissenschaft, Pädagogik und Politik: Festschrift für Jürgen Oelkers*. Weinheim, Basel: Beltz/Juventa.
- Biesta, Gert (2015): On the two cultures of educational research, and how we might move ahead: Reconsidering the ontology, axiology and praxeology of education. In: *European Educational Research Journal*, 14, 1, pp. 11-22. doi: 10.1177/1474904114565162
- Böhm, Winfried (2004): *Pädagogik*. In: Benner, Dietrich/Oelkers, Jürgen (eds.): *Historisches Wörterbuch der Pädagogik*. Weinheim: Beltz, pp. 750-782.
- Carritt, E.F. (1936): Hegel's Sittlichkeit. *Proceedings of the Aristotelian Society* 36 (1935-1936), pp. 223-236.
- Comenius, John Amos (1657/1907): *The Great Didactic (Didactica Magna)*, trans. Keatinge, M.W. London: Adam and Charles Black.
- Dewey, John (1938): *Experience and education*. New York: Macmillan Company.
- Erziehung (1968): *Brockhaus Enzyklopädie*, 17. ed., vol. 5. Wiesbaden: Brockhaus, p. 707.
- Friesen, Norm (2007): (Micro)Didactics: A tale of two traditions. In: Hug, Theo (ed.): *Didactics of microlearning*. Muenster: Waxmann, pp. 83-97.
- Friesen, Norm (2014): Waldenfels' Responsive Phenomenology of the Alien: An Introduction. In: *Phenomenology & Practice*, 7, 2, pp. 68-77.
- Friesen, Norm (2017): The pedagogical relation past and present: experience, subjectivity and failure. In: *Journal of Curriculum Studies*, 49, 6, pp. 743-756. <https://doi.org/10.1080/00220272.2017.1320427>
- Friesen Norm/Brinkmann, Malte (in progress): *Tact and the pedagogical relation*. Boston: Palgrave.
- Gadamer, Hans-Georg (2006): *Truth and Method*. New York: Continuum.
- Groothoff, Hans-Hermann (1973): *Theorie der Erziehung*. In: Groothoff, Hans-Georg (ed.): *Pädagogik Fischer Lexikon*. Frankfurt am Main: Fischer, pp. 72-79.

- Gundem, Bjorg/Hopmann, Stefan (eds.) (1998): *Didaktik and/or Curriculum. An international dialogue*. New York et al.: Peter Lang.
- Hopmann, Stefan/Westbury, Ian/Riquarts, Kurt (eds.) (1999): *Teaching as a reflective practice: the German didaktik tradition*. Mahwah NJ: Erlbaum.
- Horlacher, Rebekka (2016): *The educated subject and the German concept of Bildung: A comparative cultural history*. New York: Routledge.
- Kittler, Friedrich (1990): *1800/1900: Discourse Networks*. Stanford: Stanford University Press.
- Krebs, Angelika (2017): *Stimmung: From Mood to Atmosphere*. In: *Philosophia* 45, 4, pp. 1419-1436.
- Langeveld, Martinus Jan (1983): *Reflections on Phenomenology and Pedagogy*. In: *Phenomenology + Pedagogy*, 1, 1, pp. 5-7.
- Oelkers, Jürgen (2004): *Erziehung*. In: Benner, Dietrich/Oelkers, Jürgen (eds.): *Historisches Wörterbuch der Pädagogik*. Weinheim: Beltz, pp. 303-340.
- Phillipson, Robert (2010): *Linguistic imperialism continued*. New York and London: Routledge.
- Pinar, William (2011): *The character of curriculum studies: Bildung, currere, and the recurring question of the subject*. Boston: Palgrave.
- Ricoeur, Paul (2006): *On translation*. New York: Routledge.
- Schleiermacher, Friedrich (2012): *On the different methods of translating*. In: Venturi, Lawrence (ed.): *The Translation Studies Reader*. New York: Routledge, pp. 43-63.
- Tsuda, Yukio (1997): *The hegemony of English and strategies for linguistic pluralism: Proposing the ecology of language paradigm*. In: Smith, Larry E./Forman, Michael Lawrence (eds.): *World Englishes 2000. Selected Essays*. Honolulu HI: College of Languages, Linguistics and Literature. Univ. Of Hawaii and East-West Center, pp. 21-31.
- Uljen, Michael/Ylimaki, Rose M. (eds.) (2017): *Bridging educational leadership, curriculum theory and Didaktik: non-affirmative theory of education*. New York: Springer.
- von Humboldt, Wilhelm (1792/1999): *Theory of Bildung*. In: Hopmann, Stefan/Westbury, Ian/Riquarts, Kurt (eds.): *Teaching as a reflective practice: The German didaktik tradition*. Mahwah NJ: Erlbaum
- Waldenfels, Bernhard (2011): *Phenomenology of the alien: Basic concepts*. Evanston, IL: Northwestern University Press.
- Wigger, Lothar (2004): *Didaktik*. In: Benner, Dietrich/Oelkers, Jürgen (eds.): *Historisches Wörterbuch der Pädagogik*. Weinheim: Beltz, pp. 244-278.
- Wittgenstein, Ludwig (1974): *Tractatus Logico-Philosophicus*. London and New York: Routledge.
- Wulf, Christoph (1978): *Theorien und Konzepte der Erziehungswissenschaft*. München: Juventa.
- Zirfas, Jörg (2014): *Pädagogischer Takt: Zehn Thesen*. In: Gödde, Günter/Zirfas, Jörg (ed.): *Takt und Taktlosigkeit: Über Ordnungen und Unordnungen in Kunst, Kultur und Therapie*. Bielefeld: Transcript.

When Dictionaries are not Enough: Translational Challenges of Conceptual Historical German Terms in Educational Research¹

Kathrin Berdelmann²

1. Introduction

Researchers in the history of education often face a sort of detective work: find sources that are rare, trace down past relations and contextual factors, and reconstruct nexuses out of very few puzzle pieces. Often it is about finding the needle in a haystack. It seems that a large part of this kind of work needs to be done a second time when papers with historical terminology are translated into English. Basically, the options for translating historical German terms into English seem to be either to translate these terms into current English or into historical English.³ In both cases there will often be displacements, shifts and possibly even losses of meaning or precision.

In this paper, I shall outline some of the difficulties of transporting meaning of historical terminology from German into English language, notably, when translating conceptual terms that are utilized in the context of specific historical-cultural and local practices. This is, for example, the case with historical sources produced by practitioners within and for an everyday practice, where language is applied differently than in printed historical sources of professional discourses. Educational terminology in those documents is rather a practically operationalized version of certain notions and concepts, a specific application of terms that is anchored in a local practice and thus challenges translation to a particularly high degree.

Leaving aside the general question as to whether translation can succeed at all when challenged by historical terminology, I want to underline the productivity of problems of translation and the potential of ‘almost untranslatable’ terms – and argue against the search for single-word-equivalents in

1 A first version of this paper was published in *Jahrbuch für Historische Bildungsforschung* Vol 25, 2019: 160-168.

2 Kathrin Berdelmann is Head of the Research Library for the History of Education – Research Unit at the DIPF | Leibniz Institute for Research and Information in Education, Berlin. Email: berdelmann@dipf.de

3 In a growing number of publications, authors handle these issues by leaving historical terms in the original language and circumscribing and explaining them in footnotes. However, with some sources explanatory footnotes for the terms might become too dominant as they are getting too lengthy in comparison to the main text.

dictionaries or computer assisted translators, even in historical dictionaries or lexicons. Especially when it comes to those special types of historical sources in which language is used in a different way, terms that are hard to translate stimulate considerations of the historical and cultural backgrounds that contextualize a special local usage of terms. Such considerations can result in a deeper understanding of what is currently or originally meant with the term – in both languages – and they offer the opportunity of shaping a term or rather circumscribing it with more precision in the target language of the translation. I want to demonstrate this by using educational terminology of the pedagogy of Enlightenment as an example – in particular the terms *Bildsamkeit* and *Vervollkommnung*.

My example originates from a research project on the history of pedagogical observation in schools (see Berdelmann 2018). The sources within this project are mostly handwritten ones taken directly from everyday school-practice, in which a specific historical but very practical language is used and some basic terms of Enlightenment are ‘pedagogically operationalized’. The following is an extract from documents created by teachers at the end of the 18th Century in a well-known pietistic school⁴ that was committed to Enlightenment pedagogy at that time. This quarterly evaluation of each student described his behavior and subject-specific progress, and was based on the observations by several teachers within the previous three months. We find typical Enlightenment pedagogical terms in the following two protocols:

Teachers write about the student Julius Goldhagen:

Actually he demonstrates too little *Bildsamkeit*⁵ and is disobedient to all reminders that aim for it. He cites our greatest and most emphatic advice to educate oneself for the world – [but] is silent and lives on as before. Will this obstinacy also lead him through the world? We have often posed this question to him. He leaves it unanswered and remains as he was. [...] Whether this is intentional or merely habit, we do not know.⁶

About the student Carl von Madai we learn:

What is honorable with the *Vervollkommnung*⁷ can validly be judged by the purity of intentions and amount of effort which, as far as humans can observe it, go along with it.⁸

The terms *Bildsamkeit* and *Vervollkommnung* in original German stood for important educational concepts and illustrated a new way of thinking in that

4 Pädagogium Regium of the Francke Foundations in Halle, Germany.

5 *Bildsamkeit* is the mere capability to learn, to build oneself by learning processes and by having a certain plasticity.

6 AFSt/S A I 199, Bl. 93, Schularchiv der Franckeschen Stiftungen, Halle, translated into English by K.B.

7 Which means here ‘perfectibility’ that is achieved by perfecting ourselves.

8 AFSt/S A I 199, Bl 195.

particular period of time when modern school evolved. When looking for adequate translations into English, translators and dictionaries offer for *Bildsamkeit*: “ductility” or “plasticity” as well as “perfectibility”⁹. In more specialized literature one can additionally find the word “educability” (Siljander 2012: 87f.) and “capable of learning” (English 2013: 14f.). Translations given for *Vervollkommnung* include: “improvement” or “completion”, “perfectioning” or simply “perfectibility”.¹⁰ These are examples of how English professional educational literature (and thus not practitioner’s language) refers to and translates the German concepts of *Vervollkommnung* and *Bildsamkeit*. It is surprising, however, that the term perfectibility is occasionally applied for both German terms, *Bildsamkeit* and *Vervollkommnung*, as they are two very different concepts, although historically they grew out of a single concept, as I will show.

With regard to the pedagogical context at that time, this original concept was strongly linked to Jean-Jacques Rousseau (1712-1787), who stated that perfectibility (“la perfectibilité”) is the fundamental difference between humans and animals (Rousseau 1754/1995: 183f.). Perfectibility is the capability to perfect oneself, a faculty that is able to develop all other skills one after another. The mere possibility of perfecting oneself, and thus the possibility of advancement of humans towards humanity and morality is inherent to every single human being who progresses slowly to what is good, what is better. For Rousseau an animal, in contrast, is after a few months already what it is going to be throughout its whole life, and what its species still will be in a thousand years.

Within the German reception of Rousseau, specifically in the professional anthropological and pedagogical discourse by the end of the 18th Century, two German terms referred to the word “la perfectibilité” (and the reflexive verb “se perfectionner” – to perfect oneself), namely the terms *Vervollkommnung*¹¹ and *Bildsamkeit*. The German suffix “-sam” of the adjective “*bildsam*” means that “something can be done with a person or thing”¹² and is a compound of something passive as well as active¹³. Although Johann

9 See www.dict.leo.org (31.08.2018), www.dictcc.com (31.08.2018), Langenscheidt 2017; Pons 2015.

10 See www.linguee.de (31.08.2018), www.dictcc.com (31.08.2018), Langenscheidt 2017; Pons 2015.

11 See the German translation of Rousseau’s Discourse, p.22: the original verb „se perfectionner“ was translated into „vervollkommnet“ or, in the first translation of Rousseau’s “Emile” in the historically important publication of “Allgemeine Revision des gesamten Schul- und Erziehungswesens”: the original word „perfectionnent“ was turned into „vervollkommen“ 1789: 216.

12 Originally: “mit der beschriebenen Person oder Sache (kann KB.) etwas gemacht werden” Gesellschaft für Deutsche Sprache e.V.: <https://gfd.s.de/bedeutung-und-herkunft-von-sam-z-b-in-einsam/> [Last accessed October 17, 2019].

13 Deutsches Wörterbuch von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Leipzig 1854 (edition of 1967).

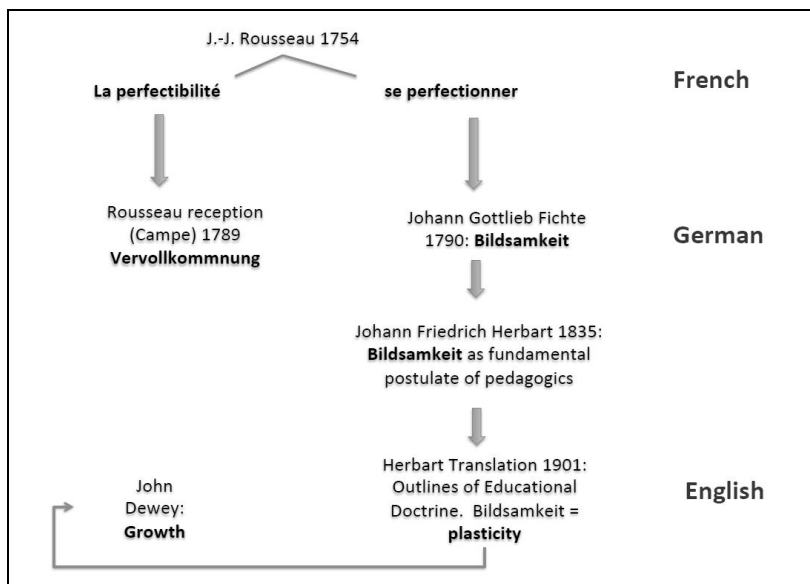
Gottlieb Herder's (philosopher, theologian, translator and poet, 1744-1803) early understanding of *Bildsamkeit* (Herder 1967) was already connected to Rousseau's *perfectibilité* (Ricken 2012: 332), it was, Johann Gottlieb Fichte, who (philosopher, 1762-1814) translated "la perfectibilité" into *Bildsamkeit* (Fichte 1796/1960: 79-80; Giesinger 2011: 894) and shaped *Bildsamkeit* into a fundamental anthropological term (ibid.; Ricken 1999: 358). *Bildsamkeit* according to Fichte implies a principle openness and indefiniteness, and is a fundamental human condition. It is in this condition the human was handed over to herself or himself (Fichte 1771: 86f). *Vervollkommnung* in contrast, as the German reception of how Rousseau defined it by the end of the 18th Century, is directed at a target: the act of and the striving towards a morally, civilly educated person through the unfolding or development of nature and predispositions (Ehlers 1789¹⁴; Rohbeck 2018: 118f).

So the German Rousseau reception referred to *perfectibilité* in a twofold way. The translations of the notion of *perfectibilité* show that the concept was divided into different parts: firstly, it was referred to as *Vervollkommnung*, and secondly, as *Bildsamkeit*, particularly with Fichte, who formed it into an anthropological concept. Educational theory at that time took up Fichte's *Bildsamkeit* and rendered it more precisely as moldability by education in relation to societal requirement and through self-active involvement (Villaume 1785: 40).

Much later, in 1844, the psychologist and pedagogue Johann Friedrich Herbart declared that *Bildsamkeit* is the fundamental postulate of pedagogics. He differentiated the concept further in his theoretical writings about education as a scientific discipline in its own (Herbart 1844). Since then, in pedagogical discourse, the notion of *Bildsamkeit* marks the precondition for the mere possibility and the potential of education (Benner/Brüggen 2004; Ricken 1999: 331). Again sixty years later Herbart's original text was translated into English. From the first edition on, *Bildsamkeit* was translated as "plasticity" (Herbart 1901: 1). Around the same time John Dewey, in turn, addresses Herbart's notion of *Bildsamkeit* as a form of growth within learning from experience (English 2013: 98; Siljander 2012; Prange 2006).

14 See a comment from the philanthropist Martin Ehlers in the translation of Rousseau's "Emile", published in "Allgemeine Revision des gesamten Schul- und Erziehungswesens" (translated by C.F. Cramer, edited by Joachim Heinrich Campe 1789): "Selbst bei einer unvollkommenen Erziehung hat man doch den zweifachen Endzweck, den zu erziehenden Menschen selbst vollkommen zu machen und in ihm der menschlichen Gesellschaft ein nützliches Mitglied zu liefern [...]", p. 42.

Figure 1. Reception of the term “perfectibilité” and its translations



This brief and certainly incomplete outline of a part of the reception of the term “perfectibilité” in German pedagogy of Enlightenment illustrates how complex the problem for translation is: there is no original or true meaning of *perfectibilité* and its German translations, but with every reference other specifications and slight transformations occur.

Turning from the professional discourse to the contextualized use of terms by practitioners, how can *Bildsamkeit* be translated in texts like the student evaluations quoted before/at the beginning of this paper? These were written for practical purposes, and the meaning of the term becomes distinct for its practical local and specific cultural as well as national context. Translations such as “plasticity” and also “educability”, as they appear in recent educational research, capture only a small part of what *Bildsamkeit* meant in these sources. When taking a deeper look into how the terms are applied within the teachers’ evaluations, we find that they gain their specific meaning directly within their pedagogical context. More precisely, when used in student-evaluation practice, *Bildsamkeit* gains typical pedagogical connotations: it is not only a mere human condition or capability, but appears as something that can be large or little, and has to be demonstrated by the child. When a student shows too little *Bildsamkeit* – as was the case with Goldhagen – the evaluation reminded him of this, and thus was an appeal to show more in the future. Accordingly, *Bildsamkeit* in this pedagogical

practice was something that had to be learned or learned to be demonstrated. The available translations, however, relate to a rather passive construction and withhold the strong and, for the German Enlightenment pedagogy, central aspect of self-activity and self-building.

This is also the case with *Vervollkommnung*. In the example above, the term *Vervollkommnung* implies that one can only truly perfect oneself (*sich vervollkommen*) when one's intentions are pure and when it is not too easy, but hard work. It is not only the case that this aspect of *Vervollkommnung* has to be understood within the pietistic context of the school, but within this pedagogical practice *Vervollkommnung* also requires that certain moral components are developed, and that effort is required to overcome obstacles in developing them.

It is quite obvious that these are typical pedagogical interpretations of Rousseau's and Fichte's concepts, as both of their works were important references for the teachers of Enlightenment pedagogy. However, within the local educational practice, the teachers transformed these concepts, and the terms are applied differently. The meaning of those terms emerges within practical contexts and situations where they are utilized, and this brings along slight displacements of meaning.¹⁵ As is very often the case with translations, they refer to the (theoretical) concepts and transfer only some aspects of the practical meaning and abandon others.

I want to propose the view that historical sources from practice, as discussed here, generate problems of translation that are productive because they require a deeper analysis of what the original notion meant within a specific practical usage and its historical context. What references did a given term have and what was its purpose within the educational setting? How exactly was it applied by the practitioners at that time? Initially, those translational difficulties make terms and their backgrounds accessible by opening a gap between the available familiar terms in the commonly used target language, and the otherness of the term in the source language that is not quite captured by the options available in the target language. Thinking about these problems as ones of a transcultural and transnational nature that cannot be resolved by simply matching terms of different languages suggests that they call for – and thus open up possibilities for – a more precise language, and a more adequate explanation of what terms meant at a particular time and in a particular place. In this paper I have shown that, for historians, challenges like this surface especially with unprinted sources, those that were generated within a historical practice, where something was

15 Although generally, all notions tend to transform in discourses over time, from a praxeological perspective it can be assumed that practice itself influences and slightly shifts meanings of notions while performing them (see e.g. de Certeau: *The practice of everyday life*, 2011: 131ff.).

done in a specific way at a specific time. Thus, as Walter Benjamin has famously noted, in translations “the life of the originals attains latest, continually renewed, and most complete unfolding” (Benjamin 2002: 225), and this means that translation proves to be a method of gaining knowledge. This seems to be especially true for translations of historical documents of practical pedagogy.

References

- Benjamin, Walter (2002): The Task of the Translator. In: Bullock, Marcus/Jennings, Michael W. (eds.): Benjamin, Walter: Selected Writings Volume 1, 1913-1926. Harvard University Press, pp. 253-263.
- Benner, Dietrich/Brüggen, Friedhelm (2004): Bildsamkeit/Bildung. In: Benner, Dietrich/Oelkers, Jürgen (eds.): Historisches Wörterbuch der Pädagogik. Weinheim: Basel Verlag, pp. 174-215.
- Berdelmann, Kathrin (2018): Individuality in Numbers. The Emergence of Pedagogical Observation in the Context of Student Assessment in the 18th Century. In: Alarcon López, Cristina/Lawn, Martin (eds.): Assessment Cultures. Historical Perspectives. Bern: Peter Lang, pp. 57-86.
- de Certeau, Michel (1988): The Practice of Everyday Life. Berkley: University of California Press.
- English, Andrea R. (2013): Discontinuity in Learning. Dewey, Herbart and Education as Transformation. New York: Cambridge University Press.
- Fichte, Johann G. (1790): Grundlage des Naturrechts nach Prinzipien der Wissenschaftslehre (1971) In: Fichte, Immanuel Hermann (ed.): Fichtes Werke. Bd. III. Zur Rechts- und Sittenlehre I. Berlin: de Gruyter.
- Fichte, Johann G. (2000): Foundations of Natural Right according to the principles of the Wissenschaftslehre. Neuhouser, Frederick (ed.): Cambridge University Press. (Translation of Grundlage des Naturrechts, 1796.)
- Giesinger, Johannes (2011): Bildsamkeit und Bestimmung. Kritische Anmerkungen zur Allgemeinen Pädagogik Dietrich Benners. In: Zeitschrift für Pädagogik, 57, 6, pp. 894-910.
- Herbart Johann Friedrich (1844): Umriss pädagogischer Vorlesungen. Göttingen: Verlag der Dieterischen Buchhandlung.
- Herbart, Johann Friedrich (1901): Outlines of Educational Doctrine. New York: Macmillan.
- Herder, Johann Gottfried (1967): Briefe zur Beförderung der Humanität. Sammlung 1-6 (1793-1796). In: Suphan, Bernhard: Sämtliche Werke. Vol 17. Hildesheim: Olms.
- Prange, Klaus (2006): Herbart und Dewey: Pädagogische Paradigmen im Vergleich (Herbartstudien). Jena: Garamond.
- Ricken, Norbert (1998): Subjektivität und Kontingenz. Markierungen im Pädagogischen Diskurs. Würzburg: Königshausen & Neumann.

- Ricken, Norbert (2012): Bildsamkeit und Sozialität. Überlegungen zur Neufassung eines Topos pädagogischer Anthropologie. In: Ricken, Norbert/Balzer, Nicole (eds.): Judith Butler: Pädagogische Lektüren. Wiesbaden: Springer VS, pp. 329-352.
- Rohbeck, Johannes (2018): Perfektibilität und Teleologie in der Geschichtsphilosophie der Aufklärung. In: Baron, Konstanze/Soboth, Christian (eds.): Perfektionismus und Perfektibilität. Theorien und Praktiken der Vervollkommenheit in Aufklärung und Pietismus. Studien zum 18. Jahrhundert, vol. 39. Hamburg: Felix Meiner Verlag.
- Rousseau, Jean Jacques (1754/1992): Discourse on the Origin and Basis of Inequality Among Men. Indiana: Hackett Publishing.
- Rousseau Jean Jacques (1754/1995): Discours sur l'origine et les fondements de l'inégalité parmi les hommes. Paris: Edition GF Flammarion.
- Rousseau, Jean Jacques (1789): Emil oder über die Erziehung. Erster Teil. Aus dem Französischen übersetzt von Carl Friedrich Cramer. Mit erläuternden, bestimmenden und berichtigenden Anmerkungen der Gesellschaft der Revisoren aus dem Revisionswerk besonders abgedruckt und herausgegeben von Joachim Heinrich Campe. Braunschweig: Verlag der Schulbuchhandlung.
- Siljander, Pauli (2012): Educability and Bildung in Herbart's Theory of Education. In: Siljander, Pauli/Kivelä, Ari/Sutinen, Ari (eds.): Theories of Bildung and Growth. Connections and Controversies Between Educational Thinking and American Pragmatism. Rotterdam: Sense Publishers, pp. 78-105.
- Villaume, Peter (1785): Allgemeine Theorie, wie gute Triebe und Fertigkeiten durch die Erziehung erweckt, gestärkt und gelenkt werden müssen. In: Allgemeine Revision des gesammten Schul- und Erziehungswesens von einer Gesellschaft praktischer Erzieher, vol. 4, Hamburg: Bohn, pp. 3-604.

Unprinted Sources

- AFSt/S A I 199, Bl. 93, Schularchiv der Franckeschen Stiftungen Halle an der Saale.
- AFSt/S A I 199, Bl. 195, Schularchiv der Franckeschen Stiftungen Halle an der Saale.

Translating Research: Tensions and Challenges of Moving Between and Through Research Practices

Inés Dussel¹

1. Introduction

Translation has been a central part of research since it emerged as a scholarly practice, an emergence that can be traced back to medieval universities but also to earlier forms of observation and recording of social and natural events. Research has always involved some kind of transit between languages (i.e., between Latin and the vernacular languages of teachers and professors in medieval universities, Le Goff 1993) and between modes of thinking, seeing, touching, or listening and the production of records and inscriptions in research practices (Daston/Lunbeck 2011). These transfers speak well to the etymology of the noun *translatio*, which associates it to a wide range of practices: metaphor, transport, transmission, transposition, transplant, displacement in space (de Libera 2016).

Even if it has been constitutive of research, until recently translation had not received much attention beyond literary or religious studies. A parallel can be constructed between the notion of ‘situated knowledges’ (the title of Donna Haraway’s seminal essay from 1989), with its emphasis on localizing practices, and Bruno Latour’s *Reassembling the Social* (2005), which points to connections, travels and translations. Movement and transferability, more than situation and location, appear as key concepts of social theory and also of scholarly practices, as can be seen in higher education with the increased cash value of internationalization and knowledge mobilization. There is an increased awareness that scholars produce amidst multiple movements in and through linguistic and epistemic practices.

It should be noted that this awareness is not only due to pressures to become global actors and perform in international arenas – which, for lack of a better word, can be ascribed to the ‘neo-liberal academia’ (Gill 2010) – but also because of post-colonial challenges to the claims of a universal knowledge and language (Mignolo 2000). The possibility of having a conversation among scholars from different geopolitical regions without reinstating the coloniality of knowledge remains a cherished yet elusive ideal, but there is a growing dialogue about how we understand the world from different locations and about the languages we use to talk about it.

1 Inés Dussel is Researcher and Professor at the Department of Educational Research CINVESTAV. Email: idussel@cinvestav.mx

In this article, I would like to reflect on three aspects of translation as part of academic practices in the social sciences and the humanities. The first one is related to the very notion of translation, for which I will go back and forth in history to understand it as a material practice embedded in particular conditions. The second is related to English becoming a *lingua franca* in contemporary academia, and the geopolitics of knowledge this institutes. The third one is related to a project that, together with a group of colleagues, was launched in 2018, and that intends to promote more reflection on research as translation. Even if it is a minor initiative, it points in the direction of making our research practices more visible as performed in and through translation, and makes the claim that we need to engage much more seriously in scholarly conversations about the languages in which we work and communicate.

2. Research as translation: textual practices and beyond

Producing knowledge in the social sciences and humanities involves several types of translation: from the oral or the visual to the written language, from events to records, from one language to another. Yet the history and the peculiarities of this work have only recently been subjected to scholarly scrutiny, together with a growing reflexivity on the subjective and material dimensions of this practice.

In writing this history, some scholars have insisted on the need to approach it from a decentered perspective that challenges the primacy of the Eurocentric views on the movements of languages, peoples and artifacts that tend to privilege some fluxes and marginalize others. An example of such an approach was shown in an exhibit on translation at the MUCEM (Musée des Civilisations de l'Europe et de la Méditerranée) that took place in Marseille in 2016. One of its displays was a PILI (Luminous Indicative Itinerary Plan, as the one used in metro plans) of the routes of translation of five basic authors or works: Aristotle, Euclid, Galen, Ptolemy and 1001 Nights.² The map made it evident that Baghdad and Cairo were more centrally connected to the flux of texts and people than Paris or Rome. Even compared to Cordoba and Toledo, major intercultural and interlinguistic sites in the Middle Ages, Baghdad and Cairo outperformed the rest in the amount of circuits of translation that went through them.

Yet despite this previous centrality, in the 12th century Renaissance the balance started to shift towards Europe when there was an institutional creation, the university, which not only held libraries but basically made

2 The map was done by Labex TransferS and Julien Caverio, and was included in the exhibit's catalogue (Cassin 2016: 96-97).

studium, study, its central *métier* (de Libera 2016). Universities were considered in medieval times as *translatio studiorum* (translation or transfer of studies), the other *translatio* being *imperium*, related to the transfer of power. These scholarly institutions were central in transporting arts – such as medicine – and ancient texts. But the transfer shifted its contents too: European universities created a new sense of Babel that was secular, Latinized, Aristotelian, and later Baconian. The fact that scholars had to converse and write in Latin meant that they were constantly translating from their vernacular languages to another idiom that granted them access to a wider understanding (Caruso 2014). This new Babel granted knowledge a special place, distinct from power, law and religion. Their institutional mandate was to produce a tradition to be passed on, which could also be understood as a common way of reading, a vocabulary, and a set of references. Considered under this lens, this was certainly a major institutional creation, with long-lasting consequences in terms of how knowledge was produced, stored and circulated and of the practices that became associated with it (Schildermans 2019).

Throughout the centuries, this Babelic tradition of shifting between languages, disciplines and geographical spaces moved beyond the universities. It was inscribed into the cosmopolitanism spread by the Enlightenment and its dreams of an educated people that would participate in the public sphere through some scholarly competences such as reading, writing, and debating the public good (Popkewitz 2008). This cosmopolitan self was to be educated through a national school system that sought to impose monolingualism, erasing the traces of foreign languages in the national language and instituting a standard version that strictly policed popular idioms and regional, natives and migrants' languages (Balibar 1985).³

However, this imposition of a monolingual standard national language was not done uniformly or successfully everywhere. Far from it: In the Latin American history of education it is clear that this process went through several negotiations and adaptations that involved both readings of the colonial past (i.e., ambivalent relationships with Iberian Spanish) and affirmations or anticipations of a future culture, for example in the efforts to create creole grammars, most notably Andrés Bello's *Spanish Grammar for the use of Americans* (1847), and to expand the school system. In these adaptations there was a close reading of European philosophical texts and pedagogical treatises, which Latin American nineteenth century liberal educators read avidly and in several languages. But somehow these translations were studied as part of a history of ideas that was disconnected from the materiality of the

3 The route to monolingualism was also paved by early modern translations of the Bible such as Luther's or King James'; it is not a coincidence that modern schooling first emerged in Protestant countries interested in disseminating this singular version of the Bible. I thank Norm Friesen for this nuance.

voyages of books and people that made these readings possible, and from the linguistic and epistemic negotiations that these intellectuals produced.

Let me delve briefly into one fascinating example of these adaptations: the life and works by Domingo Faustino Sarmiento (1811-1888), considered the “founding father” of Argentinean education, and one of nineteenth century Latin America’s leading liberal intellectuals (Puiggrós 2017). A cosmopolitan traveler, in the 1840s he was exiled in Chile, and was commissioned by the Chilean government to Europe and North America to report on their educational systems.⁴ Having been born in a province on the outskirts of the main cultural and political hubs of the former viceroyalty of the Rio de la Plata (itself on the outskirts of the Spanish Empire), and claiming to have been self-taught (although he had some clergy members on his maternal side), Sarmiento nonetheless had a stunning career as a politician and as an intellectual, and part of his success was his ability to read and translate some key European notions (such as civilization, barbarism, or republicanism) and turn them into ordering concepts to understand local contexts (Amante 2012).

His reading of European culture and epistemic traditions was not only metaphorical but had a very concrete presence in his life. Sylvia Molloy, a renowned literary historian, has studied Sarmiento’s relationship with foreign languages, and her work captures well all the ambivalence that liberals like Sarmiento felt towards European traditions. Molloy states that Sarmiento was very proud of his foreign language proficiency and even claimed that he had got “a learning machine to learn languages” from one of his relatives.⁵ In his autobiography, *Recuerdos de Provincia*, he recounted how he learned French:

In 1829, while under house arrest in San Juan, I took up the study of French as a pastime. I had planned to study it with a Frenchman, a soldier of Napoleon, who knew neither Spanish nor his own grammar, but the sight of don José Ignacio de la Rosa’s library made me greedy and, with a borrowed grammar and a dictionary, I translated twelve volumes, including [empress] Joséphine’s Memoires, one month and eleven days after beginning my solitary apprenticeship. Let me give a concrete example of my devotion to that task. I kept my books on the dining room table and just put them aside so that breakfast, lunch, then dinner might be served. My candle would go out at 2 in the morning but, when I was too absorbed in the reading, I would spend as much

4 In these trips his attention was caught by the U.S. experience, which he found much more advanced and progressive than what he saw in France, Germany and England. He became friends with Horace and Mary Mann, and since then he exchanged correspondence with Mary Peabody. At his return to Argentina, he became the head of the Education Department of the province of Buenos Aires (1856-1862), the largest of the country. From 1868 to 1874, he was President of the Argentine Republic; when he retired, he came back to the Department of Education of the province of Buenos Aires (1880-1884). He died in Paraguay in 1888, in a self-imposed exile.

5 His method of learning languages was taught by his uncle: “Oro urged the boy to translate recognizing the differences, and then to wander away from the text: ‘he enlivened the reading with digressions on the geographic canvas of the translation’ (Sarmiento p. 71)” (Molloy 1996: 26).

as three days at a stretch leafing through the dictionary. It took me 14 years to learn how to pronounce in French, for I did not really speak the language until 1846, after I had been to France (Quoted in Molloy 1991: 25).

The long paragraph speaks of the material practices and spaces through which he organized his relationship to French. In his detailed description one can almost feel how this process of translation occurred, which involved textual practices but also particular artifacts, spaces, intellectual climates, and affective involvements: Sarmiento seems to be devoured by the flames of translation. Sylvia Molloy observed that even when he admitted being a novice in the French language, the Argentinean claimed to have translated twelve books in a month and eleven days, that is, only three days per volume, not counting consultations of grammar and the dictionary that this presumably required. His learning of English was no less frantic. In Chile, he spent half his salary to pay an English teacher named Richard, and paid also the night watchman to wake him up at 2 am to study what he referred to as *my English*. After a month and a half, his instructor told him he already mastered the language except the pronunciation, which apparently he never learned. He moved to another city, where he said he translated, at a pace of one per day, 60 volumes by Walter Scott – his complete works –, while working at a mine in Copiapó.

One obvious conclusion is that Sarmiento was exaggerating, which might be true. However, Sylvia Molloy makes a more significant remark: for Sarmiento, to read was also to translate, but freely or with a difference. In a way close to what Walter Benjamin would later write about translation going beyond fidelity and the mere reproduction of meaning (Benjamin 1968), Sarmiento thought that to translate was not to read well but, from a conventional point of view, “to read very badly” (Molloy 1991: 38). Sarmiento cannibalized texts, “quoted and misquoted, borrowed and adapted” (p. 32). His was not a submissive way of reading, and did not defer to European authorities; it was from very early on a disrespectful reading, a reading entitled to read “expansively, digressively, even perversely” (p. 27). Molloy goes on: “[t]his seemingly cavalier attitude towards the European canon on Sarmiento’s part was denounced, is even denounced today, in the name of knowledge. Sarmiento, claim his opponents, does not *know*; what they fail to see is that he *knows differently*” (p. 27). His creative distortion was symptomatic of how he positioned himself in relation to Europe’s linguistic traditions: he felt entitled to “go on a rampage” through its available ideas and technologies.

Molloy’s study of Sarmiento as a kind of looter of the European intellectual practices points to the political and personal trajectories that intersected at these acts of translation, as well as to the fact that translation is “a privileged way of inventing between languages”, according to Barbara Cassin, the curator of the MUCEM exhibit on translation. In this her view in-

vention is another way of speaking of what is known differently, of what takes place or is done in-between languages, peoples, places (Cassin 2016: 12).

Sarmiento's case makes it evident that translation is always caught up in a web that includes politics and identities, and that creates new languages and cultures in its way. But in the MUCEM exhibit there is another fascinating example of the politics of translation that shows that paths are never individual and that their in-between inventions contain different possibilities. Xiaquan Chu studied the work done by the Bureau of Translators of the Chinese Communist Party, officially called the "Central Bureau for the compilation and translation of the works by Marx, Engels, Lenin and Stalin" – the four last names having been recently ditched (Chu 2016: 131). Founded in 1938, before the 1949 revolution, this Bureau began to centralize and uniform the translations of Marxist texts that had circulated in Chinese since 1905, cleansing them of the traces of their "bourgeois, anarchist, opportunist" early translators (p. 133). The sequence of the editions is indicative of the Party's political priorities: in 1958, the Bureau published the translation of the 13 volumes of Stalin's works, closely followed in 1959 by the publication of Lenin's 39 volumes. But it was only in 1983 that Marx and Engels' 50 volumes' work appeared, after the hiatus of the Cultural Revolution. During those decades, the Bureau monopolized the license to translate these authors, and no other versions were allowed to circulate. On the other hand, since the early 1960s most of the Bureau's energy was devoted to translating Mao's works into as many languages as possible. After Mao's death, the Bureau was in charge of translating the leaders of the party. It is only recently that public debates on translations – particularly of *The Communist Manifesto* – have emerged, showing a timid diversification of Chinese official language politics (Chu 2016: 141).

Chu's essay makes a parallel between the Bureau's efforts to control and centralize translations and the Egyptian King Ptolemy II's gathering of 64 Jewish sages around 270 BC who asked to translate the Hebrew Bible into Greek. The legend goes that all wise men produced an identical text, a miracle that had the trace of God (Chu 2016: 131). Yet the story also shows the contrary: only by divine intervention an identical translation could be achieved. Human beings must remain in the messiness of the Babelic world, unless other forces are called to intercede, as will be seen in the next section.

3. Contemporary monolingualisms

Translators have been present in scholarly work for some thousand years: the translation of texts, the passing on of languages and the ability to use these texts to enter into different conversations, was their major task. Why has it become so marginalized? Why it is not as central as it used to be? I would like to present some research and reflections on the invisibility of translation practices that make up academic tasks and the drive towards a less Babelic academia.

In today's academic field, most scholarly conversations are happening in English – and this very piece is an example of it. English has become the academic *lingua franca* due to the pressures to internationalize and join global rankings, and also because of the growth of international associations and congresses as venues to disseminate knowledge and construct legitimacy (a trend that is not new – see Lawn 2008).

Another source of the Anglicization of academic languages is linked to the concentration of the academic publishing industries: English monolingualism is a convenient move for transnational publishing houses that allows them to expand their readership and lower their production costs. In relation to books, Françoise Benhamou underscores that most European publishing houses have joint associations to publish in a second language, but this language is almost invariably English. Particularly in the scientific journal subsector the key players are a handful of big giants (Benhamou 2014). The losses in linguistic diversity and pluralism are remarkable.

These patterns are more evident when one considers the flux of translations. For the most part, work done in the English language in dominant centers is translated and exported to readers in other languages elsewhere in the world. Some recent research on the global flows of translation of books in the world shows that in the last three decades English has not ceased to grow as the main language from which books are translated into other languages: it went from 44.2% of the total in the decade 1980-1990 to 59.01% in 2000-2010. In that time, there has been, not surprisingly, a sharp decline in the translation of Russian books, but also fewer translations of French and German books. Translated Spanish books are slowly growing their share, going from 1.69% to 2.64% – still a very minor figure (Sapiro 2016).

Also, there are institutional changes that contribute to the increased Anglicization of academic languages. In several countries, higher education institutions are evaluating academic productivity with measures such as the Impact Factor, usually reduced to citations in exclusively English-language databases. This has been strongly debated in the scientific field (see for

example the Declaration On Research Assessment – DORA – from 2012), and its consequences for the quality of education have been put into question.

It must be said that these requirements for internationalization are felt more heavily in educational institutions outside the Anglo sphere than in the US universities, where the prevalence of English monolingualism has been described as “appalling” (Anderson-Levitt 2011). Bi- or tri-lingualism is a requirement in many academic fields in the world, except for most of the English-speaking fields. Of course, languages are never monolingual; they carry traces and signs of foreign ones, sediments of webs of travels and appropriations that go far back in history (Spivak 2012). Yet Anderson-Levitt’s characterization remains important, as it speaks of the disdain in certain research centers for connecting and understanding different linguistic systems, and the pressure on other centers, generally on the margins of the global research network, to conform to a standard language that isolates them from their local communities.

These pressures run through the old and new channels of imperial cultural and economic fluxes and seem to be deepening global inequalities rather than alleviating them. The research ecologies that this globalized, increasingly monolingual system produces are dangerously unequal. When citations are “more likely to be counted when they are in English or when an author has a conventional English name” (Cope/Kalantzis 2014: 47), and when Impact Factors have a disproportionate weight in academic evaluations, researchers who are working on non-English speaking countries are pushed to publish in English, with the consequence that “the work of some of the most talented and best recognized Latin American and Caribbean scientists [is] exclusively available in English” (Delgado-Troncoso/Fischman 2014: 389). These authors state that there are forceful arguments that:

Latin Americans cannot abandon the expression of local scientific and technological developments in their own language because to do so would run the risk of alienating their own research and development community, as well as public support for that community. Our concern is that the rationale to use incentives to publish solely in English is not adequate because it does not consider the need to train the next generation of local talent, and may even contribute to creating a more serious problem (Delgado-Troncoso/Fischman 2014: 389-390).

This more serious problem is already being experienced in some countries, such as Brazil, Argentina and Mexico, where public funding for Research & Development is being cut, and there is little public debate beyond the circles of researchers about the consequences that these cuts will have on future growth and income distribution.

For Delgado-Troncoso and Fischman, “the real challenge is to find feasible ways of reaching international audiences where English is the *lingua franca*, as well as having a more local scope” (2014: 390). National governments and research agencies can play a role in taking up these challenges; for

example Gisèle Sapiro (2014) has been studying the policies developed by the French government and cultural agencies in promoting French books on social sciences and humanities, financing translations, training translators, and other initiatives such as research stays for authors and researchers, seminars, or book fairs in different countries. As a limitation, these actions seem to be reserved for relatively wealthy countries and do not always stimulate cross-cultural interchanges.

But I would also like to claim that translation needs to be rethought not only across linguistic traditions, balancing the trend towards Anglicization, but also between research and policy fields, public education, and other arenas in which research can become more relevant and more open to dialogues with other knowledges, whose plurality highlights the polyphonic nature of knowledge production (Burke 2012). A recent study done by Barata, Shores and Alperin (2018) showed that during the Zika crisis in Brazil, relevant scientific information about the virus circulated in social media platforms such as Twitter or Facebook mostly in English, and was thus inaccessible for the public in the areas most affected. Linguistic politics are inscribed within other dynamics that affect the availability and utility of research for local communities. I will move now to the final section in which I would like to introduce a project that, together with other colleagues, was started in 2018.

4. Research in translation: an editorial project for internationalizing educational research

As part of the editorial board of the journal *International Studies in the Sociology of Education*, we have started a project that is set to promote scholarly conversations on the challenges and obstacles for translation in educational research. Based on the arguments presented above about the linguistic imbalances in published research, we created a section called *Research in Translation* that attempts to bring attention to the differential flows and directions in the translation movements.

The section aims to make accessible to English language readers work done in other languages and regions in the field of educational theory, research and practice, which is of direct interest to sociologists of education internationally. Multiple formats for submissions are accepted, be it dialogues, reviews, research papers, essays. The idea is to contribute towards more pluralism in the educational research community, and to undermine what Dipesh Chakrabarty (2002: 2) has called “asymmetric ignorance”, where the margins are conscious of the center, but the center is ignorant of the

margins. As Anderson-Levitt says, it is often the case that ‘southern’ scholars read foreign languages and travel abroad for their education, while it is much more rare that US academics undertake the same movement in the other direction, causing them to “suffer from a huge ‘blind spot’ by missing most of the literature originating outside their language zone” (Anderson-Levitt 2011: 19).

Yet, I wouldn’t like to imply that the road has to be traveled in one way only, from the North to the South. There are many occasions in which Southern scholars also suffer from ‘blind spots’ and reify the same kind of traffic of theory that has already been described, only quoting research published in English and neglecting the contexts and languages in which this research has been produced. Moreover, the notion of borders has to be further problematized, avoiding a clear-cut distinction between the North (now considered guilty of all sins) and the South (apparently free from them). It is clear for many of us that ‘northern theories’ have made it possible to think otherwise, to pose new questions in the South or in the North, to challenge our own traditions and givens, and seeing them under a new light. Speaking (also seeing, feeling, listening) across borders implies challenging these borders, particularly in this age of strong transnational flows.

The section also welcomes contributions that reflect on the practices of translation involved in scholarly practices. When we translate, we move across and through research practices. For example, moving from the oral to the written brings about several challenges (methodological and ethical) that are not always discussed in research articles or papers. Valeria Luiselli, in her essay on working as a translator for refugee children seeking asylum in New York’s courts, points to these difficulties:

[...] nothing is ever that simple. I hear words, spoken in the mouths of children, threaded in complex narratives. They are delivered in hesitance, sometimes distrust, always with fear. I have to transform them into written words, succinct sentences, and barren terms. The children’s stories are always shuffled, stuttered, always shattered beyond the repair of a narrative order. The problem with trying to tell their story is that it has no beginning, no middle, and no end (Luiselli 2016: 7).

Can these reflections find a home in an academic journal? Our argument is that they can and they should. They help understand that translation is not a unilineal process. It has zigzags, detours, returns, u-turns; it involves textual practices but also affective involvements, hearing technologies, protocols for records, among many other things. There are roads traveled in translation and roads not traveled. So, going back to the figures mentioned before, and again taking Dipesh Chakrabarty’s ideas, with the section we want to help de-provincialize translation, to think it across and beyond languages. Translation is more about opaqueness than about transparency; it is about making an intimate connection with otherness, and struggling with it, not reducing it to the known and the safe place (Spivak 2012).

The project also wants to discuss specific cases of untranslatability of research concepts and frameworks, and to open up dialogues that bridge these limitations. Perhaps the most interesting and challenging translations are those that question the translatability of cultures, simultaneously striving for common grounds (following on the logic of equivalence) and accepting that ultimately some parts of a foreign culture might remain irreducible alterity (on the logic of difference) (Donald 1992). As Sanford Budick asserts, “even if we are always defeated by translation, culture as a movement toward shared consciousness may emerge” from this defeat (Budick 1997: 22).

This defeat might ultimately be a source of strength for cultural and political renewal. As Gayatri Spivak (2012) says, translation is impossible yet necessary; it is “an intimate act of reading” (p. 251), an act that engages with the other, but that has no guarantee of achieving its ends. On the contrary, it needs to be thought as an active site of conflict, that wants to be a trace of the other, of history, of class, of differences and one not be subsumed under a generalized law of equivalences. Translation is what makes us able to think of “a mutual future in language”, of being able to imagine oneself “using the categories of the other”, which is the basis of human beings’ imaginary (Das 2001: 107 and 105).

Maybe we should engage much more openly and consciously with these kinds of debates: how do gender or equality issues translate into other cultures, not necessarily national but transnational, subnational, non-national? What is lost in translation? What is gained, for whom, and from which perspective? And how should we speak about ‘Northern’ and ‘Southern’? I use the quotation marks to denote some uneasiness about these categories, yet in the end I continue using them to point to colonial differences that need to be spoken to and revised. If these conversations are to be maintained in English, the challenges and limitations of translations should not be rendered invisible, and it should be hoped that the routes of translations will travel in several directions and not only to just one language, increasingly standardized to conform to corporate rules.

This article is just a minor gesture amidst a complex problematic, and it is clear that there is much work to do. Yet it makes it evident that the question of languages and categories remains an important one in our academic practices, and that it should be taken more seriously by our institutions and ourselves.

References

- Amante, Adriana (ed.) (2012): *Historia crítica de la literatura argentina*. Sarmiento. Buenos Aires: Emecé.
- Anderson-Levitt, Kathryn M. (2002): *Teaching Cultures: Knowledge for Teaching First Grade in France and the United States*. Cresskill, NJ: Hampton Press.
- Anderson-Levitt, Kathryn M. (2011): Introduction. *Anthropologies and Ethnographies of Education Worldwide*. In: Anderson-Levitt, Kathryn M. (eds.): *Anthropologies of Education. A Global Guide to Ethnographic Studies of Learning and Schooling*. New York: Berghahn Books, pp. 1-28.
- Balibar, Renée (1985): *L'institution du français. Essai sur le colinguisme des Carolingiens à la République*. Paris: PUF.
- Barata Germana/Shores, Kenneth/Alperin, Juan Pablo (2018): Local chatter or international buzz? Language differences on posts about Zika research on Twitter and Facebook. In: *PLoS ONE*, 13 (1), e0190482. <https://doi.org/10.1371/journal.pone.0190482>
- Benhamou, Françoise (2014): *Le livre à l'heure numérique. Papiers, écrans vers un nouvel vagabondage*. Paris: Éditions Du Seuil.
- Benjamin, Walter (1923/1968): The Task of the Translator. In: Arendt, Hannah (ed.): *Walter Benjamin. Illuminations*. New York: Schocken Books, pp. 69-82.
- Burke, Peter (2012): *A Social History of Knowledge. Vol. II: From the Encyclopédie to Wikipedia*. London: Polity Press.
- Caruso, Marcelo (2014): Within, between, above, and beyond: (Pre)positions for a history of the internationalisation of educational practices and knowledge. In: *Paedagogica Historica*, 50, 1-2, pp. 10-26. doi: 10.1080/00309230.2013.872678
- Chakrabarty, Dipesh (1992): Postcoloniality and the Artifice of History: Who Speaks for "Indian" Pasts? In: *Representations*, 37, Special Issue: Imperial Fantasies and Postcolonial Histories, pp. 1-26.
- Chu, Xiaoquan (2016): Marx et le bureau des traducteurs en Chine. In: Cassin, Barbara (ed): *Après Babel, Traduire*. Marseille: MUCEM-Actes du Sud, pp. 131-141.
- Cope, Bill/Kalantzis, Mary (2014): Changing knowledge ecologies and the transformation of the scholarly journal. In: Cope, Bill/Phillips, Angus (eds.): *The Future of the Academic Journal: Second Edition*. Oxford, UK: Elsevier, pp. 9-83.
- Das, Veena (2001): Violence and Translation. In: *Anthropological Quarterly*, 75, 1, pp. 105-112.
- Daston, Lorraine/Lunbeck, Elizabeth (eds.) (2011): *Histories of Scientific Observation*. Chicago: The University of Chicago Press.
- De Libera, Alain (2016): *Translatio Studiorum*. In: Cassin, Barbara (eds.): *Après Babel, Traduire*. Marseille: MUCEM-Actes du Sud, pp. 101-110.
- Delgado-Troncoso, Jorge Enrique/Fischman, Gustavo (2014): The future of Latin American academic journals. In: Cope, Bill/Phillips, Angus (eds.): *The Future of the Academic Journal: Second Edition*. Oxford, UK: Elsevier, 379-400. doi: 10.1533/9781780634647.379
- Gill, Rosalind (2010): Breaking the silence: The hidden injuries of neo-liberal academia. In: Ryan-Flood, Roisin/Gill, Rosalind (eds.): *Secrecy and silence in the research process: Feminist reflections*. London: Routledge, pp. 228-244.

- Haraway, Donna (1988): *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspectives*. *Feminist Studies*, 14, 3, pp. 575-599. doi: 10.2307/3178066
- Latour, Bruno (2005): *Reassembling the Social. An Introduction to Actor-Network-Theory*. Oxford, UK: Oxford University Press.
- Lawn, Martin (2008): *A Europeanization of Education Research? The Work of the International Examination Inquiry, its Researchers, Methods, and Influence*. In: Lawn, Martin (ed.): *An Atlantic Crossing? The Work of the International Examination Inquiry, its Researchers, Methods and Influence*. Oxford, UK, Symposium Books.
- Le Goff, Jacques (1993): *Intellectuals in the Middle Ages*. London: Blackwell.
- Luiselli, Valeria (2017): *Tell Me How it Ends. An Essay in Forty Questions*. London: 4th State.
- Mignolo, Walter (2000): *Local Histories/Global Designs: Coloniality, Subaltern Knowledges, and Border Thinking*. Princeton, NJ: Princeton University Press.
- Molloy, Sylvia (1991): *At Face Value. Autobiographical writing in Spanish America*. New York, Cambridge University Press.
- Puiggrós, Adriana (2017): *Adiós Sarmiento. Educación pública, Iglesia y mercado*. Buenos Aires: Colihue.
- Popkewitz, Thomas S. (2008): *Cosmopolitanism and the Age of School Reform. Science, Education, and Making Society by Making the Child*. New York: Taylor and Francis.
- Sapiro, Gisèle (ed.) (2014): *Sciences humaines en traduction. Les livres français aux États-Unis, au Royaume-Uni et en Argentine*. Paris: Institut Français-Centre Européen de Sociologie et des Sciences Politiques.
- Sapiro, Gisèle (2016): *La traduction à l'heure de la mondialisation: état des lieux*. In: Cassin, Barbara (ed.): *Après Babel, Traduire*. Marseille: MUCEM-Actes du Sud, pp. 83-87.
- Schildermans, Hans (2019): *Making a University. Introductory Notes on an Ecology of Study Practices*. Ph.D. Dissertation, Faculty of Psychology and Educational Sciences, KU-Leuven.
- Spivak, Gayatri Chakravorty (2012): *An Aesthetic Education in the Era of Globalization*. Cambridge, MA: Harvard University Press.

The Challenges of Test Translation

Britta Upsing¹ and Musab Hayatli²

1. Introduction

Test translation can easily go wrong. Just to give a few examples: In one PISA study the term ‘space suit’ was rendered as ‘special suit’ in the Spanish version and the item had to be dropped; in another higher-education study, the translated rubric talks about a ‘goal scorer’ instead of ‘scorer’, and in a school test ‘early agrarian society’ was rendered ‘a society with agrarian industry’. These errors were detected before the tests were actually conducted as a result of translation quality control checks.³ These examples show how important it is to have professionals do the translations, using rigorous methodologies. While these examples may lead some to believe that it would be easier to simply write the tests in the language of the respondents with no translation involved, this is not an option for international tests or surveys, particularly in many countries that have more than one national language.

In the past two decades, international large-scale assessment studies like PISA (the Programme for International Student Assessment) or PIAAC (the Programme for the International Assessment of Adult Competencies) have become prevalent and their political impact is not to be underestimated. Studies like PISA are under much scrutiny: The translation of the tests is easily criticized as no objective measures exist by which to easily evaluate translations (for some examples of this criticism, see: Arffman 2012; Dolin 2007; Ercikan 1998; Karg 2005; Wuttke 2007). Also, doubts arise whether it is even possible to conduct fair tests across different cultures and languages (cf. Arffman 2007; Asil/Brown 2015; Bonnet 2002; El Masri/Baird/Graesser 2016; Hamilton/Barton 2000; Puchhammer 2007). Language plays an important role in these tests. If a respondent struggles with the language of a test, these difficulties will probably interfere with his or her ability to answer the test items correctly. At this point, the validity of the test may be at stake. These issues become even more complicated with increasing global population migrations and internal diversification within nation states, which raise the question into which languages any test should be translated. Still, it

-
- 1 Britta Upsing is Researcher at the Technology Based Assessment Centre at the DIPF | Leibniz Institute for Research and Information in Education. Email: upsing@dipf.de
 - 2 Musab Havatli is General Manager at cApStAn Linguistic Quality Control (LQC) Inc. Email: musab.hayatli@capstaninc.us
 - 3 Examples drawn from one of the author’s [Hayatli] experience in managing translation quality control for tests and assessments.

is not possible to conduct these tests or studies without translation. Even though there have been advocates for writing different tests in each of the languages within a given country or across countries (cf. for example Bonnet et al. 2001 to learn more about this approach), translation has become the norm for international tests.

The goal of this article is to illustrate the challenges of test translation and to describe some of the measures that have been implemented to deal with these challenges. We will first explain what international large-scale assessment studies (iLSA) are: We will give a brief outline of their history; describe their contents, goals and their political impact. Next, we will use an actual test item from the PIAAC study as an example to illustrate which questions and difficulties come up when test items are translated. We will then describe the strategies that have been developed to deal with these translation challenges. Here we will mostly draw on strategies for the PISA- and PIAAC-tests. In the final section, we will discuss the remaining challenges, with a focus on the role of language in diverse societies.

2. International large-scale assessment studies

The PISA study is probably the most famous international large-scale assessment study. The term international large-scale assessment (iLSA) “refers to national or international assessments that serve to describe population characteristics with respect to educational conditions and learning outcomes, e.g. the competence level in a particular population” (Upsing/Gissler/Goldhammer/Rölke/Ferrari 2011: 44f.). These assessment studies “are used for monitoring the achievement level in a particular population, for comparing assessed (sub)populations, and also for instructional program evaluation” (Upsing et al. 2011: 45). In the end, “such assessments may form the basis for developing and/or revising educational policies” (Upsing et al. 2011: 45) – and this is one of the reasons why iLSA and the processes for setting them up are under such scrutiny.

In the case of PISA, comparisons are made between the levels of competencies of 15 year old students across countries. The first PISA cycle was administered in 2000, but the very first iLSAs were already administered in the 1960s when twelve countries participated in the “Pilot Twelve-Country Study” by the *International Project for the Evaluation of Educational Achievement*, the precursor of the *International Association for the Evaluation of Educational Achievement* (IEA) (also see Wagemaker 2013). For this study, 13 year-olds were tested in different subjects and the study examined whether similar research across countries would be feasible. Further research focused on test implementation and methodological issues,

but the next big milestone was set only in 1995, when the *Third International Mathematics and Science Study* (TIMSS) (today *Trends in International Mathematics and Science Study* – TIMSS) was conducted in 46 countries to test mathematical and science competencies of students. A literacy study (*Progress in International Reading Literacy Study*, PIRLS) was set up by IEA closely thereafter. PISA, the first extensive iLSA by the OECD, followed in the year 2000. Work on implementing this study already began in the middle of the 1990s. Since then, international large-scale assessment studies have been on the rise, with more countries participating, more age groups being targeted (there are also studies like PIAAC which target adults), and more domains being tested. Or, as Wagemaker (2013: 18) puts it: “Today, the work of IEA and, in particular, its TIMSS and PIRLS assessments, along with OECD’s PISA, are characterized by participation that is truly worldwide”.

Most studies contain a survey to explore the respondent’s socio-economic background and education, as well as a performance-oriented test to evaluate the respondent’s competencies (such as literacy, problem-solving, numeracy). When these tests are created, it first has to be decided which construct (such as literacy or numeracy) is to be measured and what kind of test items are needed to measure this construct. These decisions are documented in the so-called “assessment framework”, which is then used as a basis for the development of the test (for more information: Upsing et al. 2011: 47). Each test contains test items of varying difficulty:

[. . .] an item is the smallest assessable entity of a test. It consists of a stimulus that serves to evoke an observable response from the test taker; this is the material that the subject [or test taker] uses to answer the question. Individual differences in the response are assumed to reflect individual differences in the assessed ability or competence (Upsing et al. 2011: 45).

A test contains multiple items to assess the same ability, which “allows [one] to measure individual ability levels reliably” (Upsing et al. 2011: 45). The responses given across the items of the test are used as empirical basis for estimating the subject’s ability level (Upsing et al. 2011: 45).

The survey and the test will have to be translated after their creation. The goal of the translation is to have comparable international tests. The psychometric properties of the item should not be touched by the translation, which means that a test item should not become easier or more difficult because of translation. But how is this done and what makes test translation so difficult?

3. Why is test translation so difficult?

At first glance, translation seems to be an easy task: Transferring a text from one language, the source language, into another language, the target language. Still, translation is not as easy as it seems, and not as straightforward as might be expected.

The difficulties that arise with translation can be explained with an example from Prunč (2012: 157-160). Prunč asks about the adequate translation for the greeting [“Dear Sir”] of an English business letter into German. There are several ways of translating this English greeting into German: For one thing, it is possible to translate this greeting by just transcoding the literal meaning of the English words into German [“*Lieber Herr*”]. Secondly, it is possible to use the German greeting that would be expected in a German business letter [“*Sehr geehrter Herr X*”]. But which one of the two is correct? According to Prunč, the purpose of the translation determines which one of the two translations is the adequate one. The first translation will feel awkward for a native speaker of German when read as a part of a German business letter. That greeting will not feel authentic (as “*Lieber Herr*” would not be used as a greeting in a German letter). Still, if the purpose of the translation is informative, e.g. explaining what formulations are used in English business letters, then the first translation suggestion is adequate. If the purpose is to have a translation to be used as a part of a German business letter, then the second suggestion is adequate – even though the translation is not at all literal. So the purpose of the translation determines which approach to translation should be used and which translation suggestion is deemed adequate.

In test translation, it is entirely feasible to think of circumstances for which an informative approach should be taken. So for example a Korean test that was given to Korean respondents may later on be translated into English to inform non-Korean speaking researchers about the content of this test or survey. In all likelihood, these translations will not be adequate to be used as test items for English speaking test takers. Still, the purpose of most test translations is not informative. The purpose of the majority of test translation will be the creation of tests in different languages to enable testing populations with different native languages. The question here is how to translate test items to make sure that the resulting translations will allow for fair testing across languages. So the main goal of the translation process for test items in international large-scale assessment studies is that “a person of the same ability will have the same probability of answering any assessment item successfully independent of his or her linguistic or cultural background” (Thorn 2009: 9). So a respondent is not supposed to be advantaged or disadvantaged for taking a test in a certain language. Or, as Ferrari et al. put

it, the goal is “to retain the cognitive equivalence of tasks as much as possible, so that each item examines the same skills and invokes the same cognitive processes as the original version, while being culturally appropriate within the target country” (Ferrari/Wäyrynen/Behr/Zabal 2013: 10f.).

The difficulties that arise with this high standard can be exemplified using a test item from the PIAAC study (also see Upsing 2017 regarding this example). Test respondents for the field test of PIAAC were asked to read the text in Figure 1 (the stimulus material) to answer the question: “By what time children should arrive at preschool?”

Figure 1. Excerpt from an English test item

<p style="text-align: center;">Preschool Rules</p> <p>Welcome to our Preschool! We are looking forward to a great year of fun, learning and getting to know each other. Please take a moment to review our preschool rules.</p> <ul style="list-style-type: none">• Please have your child here by 9:00 am.• [...] • Please sign in with your full signature. This is a licensing regulation. Thank you.• Breakfast will be served until 7:30 am.• Medications have to be in the original, labeled containers and must be signed into the medication sheet located in each classroom.• If you have any questions, please talk to your classroom teacher or to Ms. Marlene or Ms. Tree.
--

Source: Organisation for Economic Co-operation and Development [OECD] 2013: 1

Before the test can be translated (for example into German), a decision will have to be made on how faithfully the structure of the translated text should follow the structure of the source test: that is, how important is it that the translation reads like the source or that it reads like an authentic text – e.g. like a text parents or teachers may encounter in a “German preschool”? Authenticity may only be achievable at the expense of changing the literal meaning of the translation substantially from the source version.

To illustrate some of the challenges that arise during the translation task, here are some of the questions that translators working on the text above may ask themselves regarding several text elements when attempting to translate the text above into – for example – German:

- What kind of signature is asked of parents when bringing their kids?
Would a comment like this be included in a note to parents in a preschool

in Germany – and if not, should a translation still be rendered for the German text for this paragraph?

- What is a “licensing regulation” in this context?
- Is it realistic for a preschool in Germany to have breakfast served until 7.30 if most preschools only open at that time?
- How should an expression like “classroom” or “teacher” be rendered into German if the literal translations of these words (*Klassenzimmer* and *Lehrer/-in*) are only used in compulsory schooling situations?
- How formally should the reader be addressed?
- What is a “medication sheet”?

For the German translation, the following solutions were found:

Figure 2. Excerpt of a translated test item

Kindergartenregeln

Willkommen in unserem Kindergarten! Wir freuen uns auf ein großartiges Jahr mit viel Spaß, Lernen und gegenseitigem Kennenlernen. Bitte nehmen Sie sich einen Augenblick Zeit, um unsere Kindergartenregeln durchzusehen.

- Bitte sorgen Sie dafür, dass Ihr Kind bis 10.00 Uhr hier ist.
- [. . .]
- Bitte tragen Sie sich mit Vor- und Zunamen ein. Dies ist eine Zulassungsvorschrift. Vielen Dank.
- Frühstück gibt es bis 8.30 Uhr.
- Medikamente müssen sich in beschrifteten Originalverpackungen befinden und in den Medikamentenbogen eingetragen werden, der in jedem Gruppenraum ausliegt.
- Falls Sie irgendwelche Fragen haben, wenden Sie sich bitte an die Erzieherin Ihrer Gruppe oder an Frau Mahler oder Frau Baum.

Source: GESIS, Leibniz-Institut für Sozialwissenschaften 2014: 1

The German translation shows that a mixed approach was used. The translation uses source text structures like “licensing regulations” (*Zulassungsvorschrift*), “medication sheet” (*Medikamentenbogen*) or the complete literal rendering of the sentence “Please sign in with your full signature. This is a licensing regulation. Thank you.” It is questionable whether these translations will be understood by a German reader. Still, the translation also shows some signs of adaptation to the German context: “Teacher” is rendered as *Erzieherin* (literal translation: “educator”), “classroom” is rendered as *Gruppenraum* (literal translation: group room). Also, the arrival and breakfast

times have been adapted to better fit the German context. To sum up though, the German translation has not resulted in a text that will feel authentic to a German reader (at least not to a German parent or a German teacher), and which reads clumsily and with little fluency. For test translation the question though is whether the lack of authenticity will impact the difficulty of the item for a German test taker.

There are even more poignant examples to show what difficulties translators face when doing test translation: Questions like “How many sides are there to a triangle?” or “A dentist is what kind of doctor?” – may seem like perfectly reasonable items in English or French (in OECD studies like PISA these two languages are used as source languages). In German though (and also in Arabic, Finnish or Hungarian), the only possible translation of these items give away the answer as these languages do not use a Latin or Greek basis for these words. Instead, they have a reference to the number “3” (“*Dreieck*” – “*Drei*”, “*Kolmio*” – “*Kolme*,” “*muthallath*” – “*thalathah*” = 3, etc.), or the literal translation of “dentist” in these languages is “tooth doctor” (“*Zahnarzt*”, “*hammaslääkäri*”, “*fogorvos*”, “*Tabiib Asnaan*” respectively).

These examples also show the following additional difficulty: tests like PISA or PIAAC are not only translated into one language, but into dozens of languages. The questions that arise when a text is translated into one particular language will with all likelihood also arise for other languages. Some questions will arise in some languages, but not in all. So on the basis of what kind of information are these translation decisions made? How do translators with other target languages deal with the challenges? Are test results still comparable if a text only seems authentic in some of the translations (and the source text)? These questions are important as comparability of test items is a prerequisite for international studies. The next section will explain what measures are taking for the PISA test to deal with these challenges.

4. Strategies for test translation – an example from PISA

If translating a simple greeting or basic terms poses such challenges even for languages that are closely related to each other as German and English, then it would be fair to argue that achieving equivalence across languages when translating a test is very difficult if not impossible. And indeed, an exact equivalence might not be an attainable goal. However, aiming for an unbiased translated test version for all languages would seem a reasonable goal to aspire to. And the task of such translation is certainly one where a number of procedures can be taken “to ensure that the instruments used in all participating countries to assess students’ performance provide reliable and

comparable information” (OECD 2017b: 92). In the PISA study, these steps, or procedures include the following (OECD 2017b: 92):

- “optimising the English source version for translation through translatability assessment
- development of two source versions of the instruments, in English and French [...]
- double-translation design [two independently-generated translations]
- preparation of detailed instructions for the localization [translation] of the instruments [...]
- preparation of translation/adaptation guidelines
- training of national staff in charge of the translation/adaptation of the instruments
- validation of the translated/adapted national versions [...]

These processes can be divided into three main steps:

- Upstream work: work taken before translation starts, which is aimed at the production of a text that is fit for translation and allows for achieving multiple target versions with reduced bias (translatability assessment would be one such an approach),
- Translation process: measures taken during the translation work itself; for example a rigorous approach to translation (double-translation and reconciliation of multiple translations would be an example); and
- Quality control and other measures: measures taken after the translation itself is completed (like verification/review or auditing).

4.1 Upstream work

This process targets the source item, the material in the source language that is intended to be translated into the target language versions. In OECD or IEA studies, the source languages are English, French or both (in PISA, English and French are used as sources simultaneously; the creation of the French version from the English version is also used to detect problems with the source). As one means of upstream work, PISA uses the so-called ‘translatability assessment’ (cf. OECD 2017b: 92-93 for a detailed description). Here translators of different language family groups – to maximize exposure – look at the draft source version of the items and evaluate how transferable and adaptable the text would be into their language, taking into account syntactic, semantic, and cultural issues. Their feedback could be that there are no issues to be seen, that an item should be dropped or rewritten, or that notes to translators should be provided to help them with translation. These so-called item specific guidelines provide help for specific translation challenges by

explaining difficult phrases or by giving hints on how to adapt an item to the national context (OECD 2017b: 94). These various types of feedback are then consolidated into one single report to give feedback to item developers (cf. Figure 3).

Figure 3. An example for a report generated by advance translation, taken from the PISA 2015 field trial

English	Translatability Evaluation	Linguist's Comments	Suggestions for alternative wording	Suggestion for translation or adaptation
One 8 oz. glass of milk is packed with vitamins, minerals, and a wealth of health benefits.	Redundancy issue	1. Adding the equivalent of '8 oz.' doesn't add to the meaning of this sentence, at least not in Dutch. If kept, adaptation to local measurement units is necessary. 2. '... and a wealth of health benefits' asks for a non-literal translation.	[if measurement considered redundant – please note that the same amount is also indicated in the second stimulus] One glass of milk is packed with vitamins, minerals, and a wealth of health benefits.	[If measurement maintained] Please adapt and convert to local measurement units.

This feedback, which results in possible item changes, notes for translators and possibly dropped items, is part of the quality control process. Upstream work is complemented through the training of translation personnel, and making them aware of the general traps to be avoided when doing test translation.

4.2 Translation process

In PISA, the so-called double-translation approach is used. Here, two translators work independently of each other on the same text and the same target language while utilizing the translation notes. A third translator then

reconciles the two texts by selecting the better of the two translations without making any preferential stylistic changes, while trying to ensure consistency in terminology and style. This approach allows for multiple inputs from various translators thus hopefully enriching the target text and limiting idiosyncratic stylistic preferences (OECD 2017b: 93-94).

4.3 Quality control

In PISA, all translations are verified: For each language a translator examines each translated text segment taking into account any existing item notes, assessing the quality of the translation, offering rough back-translation in case of errors, and then offering a remedy to any problems they identify. This feedback is discussed between the different parties involved in the process.

4.4 Other measures

It has to be added that the translation quality control process as described above (which is rather of a qualitative, not quantitative nature) is not the only measure. After the translation of the test items, the items are field-tested to rule out bias: “an item is biased if persons with the same standing on the underlying construct (e.g., they are equally intelligent) but coming from different cultural groups, do not have the same expected score on the item“ (van de Vijver 2015). This means that bias does not occur at random, but shows that a systematic error exists which would occur again if the study was to be repeated (van de Vijver/He 2016: 232f.). Item bias is detected by administering the test items to a representative sample of the target population and by analyzing the results via so-called differential item functioning analysis (DIF analysis) (Malda et al. 2008: 452). These statistical measures help to detect “whether items have an equal probability of a particular response for examinees from different language groups who have equivalent measures” (He/Wolfe, 2010: 81). So if two test-takers from different populations are equally able, they should give comparable answers for each item. If they do not, then there should either be a good reason for this behavior or the item should be deleted from the test (cf. Hambleton 2005: 29). The PISA and the PIAAC study both delete these kinds of items from their final test after translation. Still, one problem remains when this approach is used: “If a test is biased consistently in favor of one language, DIF will not detect this and it will appear that all is well. Consistent bias in a specific direction could be due to poor translation quality where all items in one version are systematically biased.“ (El Masri et al. 2016: 11). This would mean that if a translated test is – on the whole – easier or more difficult than other translated versions, then this difference may not be detected.

5. Conclusion

As we have shown, many steps are taken in international large-scale assessment studies like PISA to ensure high quality translations. The measures have come a long way considering that the importance of translation is easily underestimated, and that a method like back translation⁴ has long been used as the standard practice for quality control for translated tests (OECD 2017b: 93). This multi-step approach only makes sense in the way that it allows non-speakers of the target language to achieve a sense of the translated text. However, back translation, in addition to being more costly and less practical, tends to have a bias towards literal translation. It is costly as it requires full translation costs and time-frames; and it is less practical as it only diagnoses problems rather than offering solutions (for a detailed critique see Behr 2017).

Measures used in PISA offer an advantage to back translation by involving several translators who look at both the source and target versions, and by providing these translators with guidelines, training and an optimized source text (OECD 2017b: 93). International large-scale assessment studies are apparently successful at creating tests with similar psychometric properties across languages (and therefore having fair tests): This is achieved by using quality control procedures in translation and by selecting items on the basis of their psychometric properties after the field tests (for example, the item “preschool rules” from above did not make it into the final item selection of PIAAC).

Nonetheless, there are many situations where it would be possible for one translated version to be – on the whole – easier or harder than other translations. One of the sources of this challenge may not be the translation or the test itself, but a systematic difference between how different populations deal with languages. A monolingual population may deal differently with a test than a multilingual population. The translation approaches described here seem to have in mind an “ideal recipient” of the translation of the test: This person is expected to be monolingual with exactly one dominant language: the language spoken at home is the same as the language of instruction, which is the same as the national language, and so on. Or, as it is put in the “PISA Technical Standards”: “It is assumed that the students tested have reached a level of understanding in the language of instruction that is sufficient to be able to work on the PISA test without encountering linguistic problems” (OECD 2017a: 8).

4 Back translation involves translating the source text into the target language, translating the target text back into the source language, comparing the two “source” language texts to identify discrepancies and possible problems with the target language translation.

But the situation may be more complicated than this. First of all, there is “diglossia”, where a language community uses “two or more varieties of the same language [...] under different conditions” (Ferguson 1959: 325). Mostly, one variety (usually the standard language) is restricted to formal situations (and often constitutes the written language), whereas the other one is used for everyday situations and interactions (cf. Ferguson 1959). Or, as Ferguson puts it: “it is typical behavior to have someone read aloud from a newspaper written in H [standard language] and then proceed to discuss the content in L. [dialect]” (Ferguson 1959: 325). So the standard language is effectively the second language of these speakers and is acquired by formal education (Ferguson 1959: 331). A prominent example of this situation is Arabic, where Modern Standard Arabic (MSA) is the written version of the language, which no Arabic speaker (no matter what social background) learns as his or her first language. All Arabic language speakers are raised speaking a local dialect, which effectively is their first language. Their exposure to MSA starts when they learn to read and write and rarely does it become the language of everyday communication except in its written form. Even that is being challenged these days with the advance of social media where unstandardized written forms are rather prevalent.

In the case of tests this means that the language in which the test is written is not the language spoken as such; it is instead a language that is considered a standard. The question now is whether a test in Modern Standard Arabic is on the whole more difficult for an Arabic speaker – as Modern Standard Arabic is not this person’s native language. This question also arises for other diglossic speech communities, for example in the German speaking part of Switzerland, with Creole in Haiti, or Greek in Greece, to mention a few (Ferguson 1959: 325).

A similar situation exists in countries that have more than one official language, and where one of these languages is dominant at the expense of other languages spoken in that country. This means that this dominant language is not the native language of the majority of the population, but used in formal settings and/or in education. For example, while English and Tagalog are both official languages of the Philippines, neither may be the first language of large parts of the population (Smolicz/Nical 1997). A similar situation may arise with English and Hindi in India, or English in Singapore, as well as in many African countries, to give a few more examples. Problems may arise when – as is the case for instance in many sub-Saharan African countries – the national language (often English or French) is not the same as the native language(s) of the majority of the population, and the language of instruction at school is not the same as the native language of the students (Smolicz/Nical 1997).

Several of these countries are furthermore shaped by multilingual population. Papua New Guinea constitutes an extreme example (where

approximately 870 languages are spoken by a population of a few million people), and here situations are described in which a child speaks one language at home, another one at the market place, and English if schooling is continued (Cenoz/Genesee 1998: 4). This situation is similar in India, as well as in many African countries. So in these and other cases multilingualism predominates and children (and adults) are frequently exposed to different languages.

In the context of international large-scale assessment studies, such cases raise some very important questions: Should the test be in the native language, the language of instruction or in both languages? Is the test going to be harder for the respondents because of any one of these possibilities?

Similar questions arise in seemingly monolingual countries but with substantial to large minority language speaker populations. For example, the US government sought, through the Elementary and Secondary Education Act (ESEA) 2001, better known as the 'No Child Left Behind' (NCLB) Act, to enact measures allowing, among other things, for school students to take exams in their mother tongues. The difficulties remain, nonetheless. For some migrants it may be hard to discern which language actually is the 'native language', e.g. which language is the dominant one (and why) (Parameswaran 2015). If the first language (or language of origin) is only spoken at home and all other interactions are in the national language of that country, then a test in the first language of that person may not be very helpful as most situations are experienced in the second language of that person. For others, although Spanish, Arabic, or Haitian Creole might be their first language at home, it is doubtful that, as students in US schools, they do their course work in any of these languages, raising the issue, once again, of the comparability of the test results.

In all these cases, even the best translation may not overcome the difficulties raised for these persons. More research is needed to find out more about whether the assumed difficulties actually arise, and if so, how to overcome them. A possibility (which is technically feasible with computerized tests) may be to allow for switching between languages in a test. Tests like PISA which are used as a proxy to compare different education systems may also increasingly try to experiment with different approaches. Already today, participating countries are encouraged to construct test items and submit them for review to be included in the test. These efforts will not eliminate the cultural challenges, but help to make sure that the origin of the test items becomes more and more diverse. Meanwhile, it is important to ensure that the results of iLSA tests are not over-interpreted. When discussing results, it is also important to keep in mind that – despite all the best efforts – the tests may still be biased. Further, in the case of individual diagnostics, when the competencies of individuals are compared – and the results carry consequences for this individual and his or her future

career – it is even more important than in the case of iLSA to make sure that test language does not disadvantage this individual. Thus, great efforts should be made when constructing and translating a test; and the efforts put into this process in PISA or PIAAC may serve as an example here. Speaking more generally, it would be very welcome if multilingualism was regarded not as a problem to be solved – as happens increasingly in some national and international contexts – but rather as an asset to be recognized and affirmed.

References

- Arffman, Inga (2007): The problem of equivalence in translating texts in international reading literacy studies. A text analytic study of three English and Finnish texts used in the PISA 2000 reading test. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Arffman, Inga (2012): Unwanted literal translation: An underdiscussed problem in international achievement studies. In: *Education Research International*, 2, pp. 1-13. doi: 10.1155/2012/503824
- Asil, Mustafa/Brown, Gavin T. L. (2015): Comparing OECD PISA Reading in English to Other Languages: Identifying Potential Sources of Non-Invariance. In: *International Journal of Testing*, 16, 1, pp. 71-93. <https://doi.org/10.1080/15305058.2015.1064431>
- Behr, Dorothée (2017): Assessing the use of back translation: The shortcomings of back translation as a quality testing method. In: *International Journal of Social Research Methodology*, 20, 6, pp. 573-584. doi: 10.1080/13645579.2016.1252188
- Bonnet, Gérard/Braxmeyer, Nicole/Horner, Susan/Lappalainen, Hannu-Pekka/Levasseur, Jacqueline/Nardi, Emma/Rémond, Martine/Vrignaud, Pierre/White, Janet (2001): The use of national reading tests for international comparisons: ways of overcoming cultural bias. Paris: Ministère de l'Éducation Nationale.
- Bonnet, Gerdard (2002): Reflections in a Critical Eye: On the Pitfalls of International Assessment. In: *Assessment in Education*, 9, 3, pp. 387-399. doi: 10.1080/0969594022000027690a
- Cenoz, Jasone/Genesee, Fred (1998): *Beyond Bilingualism: Multilingualism and Multilingual Education: Multilingual Matters Limited*. Clevedon: Multilingual Matters.
- Dolin, Jens (2007): PISA – An example of the use and misuse of large-scale comparative tests. In: Hopmann, Stefan Thomas/Brinek, Getrude/Retzl, Martin (eds.): *Schulpädagogik und Pädagogische Psychologie: Vol. 6. PISA zufolge PISA – PISA According to PISA: Hält PISA, was es verspricht? – Does PISA Keep What It Promises?* Wien: LIT Verlag, pp. 93-125.

- El Masri, Yasmine H./Baird, Jo-Anne/Graesser, Aart (2016): Language Effects in International Testing: The Case of PISA 2006 Science Items. In: *Assessment in Education*, 23, 4, pp. 1-29. doi: 10.1080/0969594X.2016.1218323
- Ercikan, Kadriye (1998): Translation effects in international assessments. In: *International Journal of Educational Research*, 29, 6, pp. 543-553. doi: 10.1016/S0883-0355(98)00047-0
- Ferguson, Charles A. (1959): Diglossia. In: *WORD*, 15, 2, pp. 325-340. doi: 10.1080/00437956.1959.11659702
- Ferrari, Andrea/Wäyrynen, Laura/Behr, Dorothee/Zabal, Anouk (2013): Translation, adaptation, and verification of test and survey materials. In: Organisation for Economic Co-operation and Development (OECD) (ed.): *Technical report of the Survey of Adult Skills (PIAAC)*. Pre-publication copy. https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf
- GESIS, Leibniz-Institut für Sozialwissenschaften (GESIS) (2014): *PIAAC Beispielaufgaben Deutschland*. http://www.gesis.org/fileadmin/piaac/Downloadbereich/PIAAC_BspAufgabenDeutschland.pdf
- Hambleton, Ronald K. (2005): Issues, Designs, and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. In: Hambleton, Ronald K./Merenda, Peter F./Spielberger, Charles D. (eds.): *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Mahwah, NJ: Taylor & Francis, pp. 3-38.
- Hamilton, Mary/Barton, David (2000): The International Adult Literacy Survey: What Does It Really Measure? In: *International Review of Education*, 46, 5, pp. 377-389. doi: 10.1023/A:1004125413660
- He, Wei/Wolfe, Edward W. (2010): Item Equivalence in English and Chinese Translation of a Cognitive Development Test for Preschoolers. In: *International Journal of Testing*, 10, 1, pp. 80-94. doi: 10.1080/15305050903534738
- Karg, Ina (2005): *Mythos PISA. Vermeintliche Vergleichbarkeit und die Wirklichkeit eines Vergleichs*. Göttingen: V-&-R-Unipress.
- Malda, Maike/van de Vijver, Fons J. R./Srinivasan, Krishnamachari/Transler, Catherine/Sukumar, Prathima/Rao, Kirthi (2008): Adapting a Cognitive Test for a Different Culture: An Illustration of Qualitative Procedures. In: *Psychology Science Quarterly*, 50, 4, pp. 451-468. http://www.psychologie-aktuell.com/fileadmin/download/PsychologyScience/4-2008/01_Malda.pdf
- Organisation for Economic Co-operation and Development (OECD) (2013): *PIAAC sample items literacy*. <http://www.oecd.org/site/piaac/Literacy%20Sample%20Items.pdf>
- Organisation for Economic Co-operation and Development (OECD) (2017a): *PISA 2015 Technical Standards: Annex*. <https://www.oecd.org/pisa/pisaproducts/PISA-2015-Technical-Standards.pdf>
- Organisation for Economic Co-operation and Development (OECD) (2017b): *PISA 2015. Technical report. Programme for International Student Assessment*. <http://www.oecd.org/pisa/data/2015-technical-report/>
- Parameshwaran, Meenakshi (2015): Explaining intergenerational variations in English language acquisition and ethnic language attrition. In: Heath, Anthony (ed.): *Migrants and their children in Britain: Generational Change in Patterns of Ethnic Minority Integration*. London: Routledge, pp. 27-45.

- Prunč, Erich (2012): Entwicklungslinien der Translationswissenschaft. Von den Asymmetrien der Sprachen zu den Asymmetrien der Macht (3rd ed.). TransÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens: Vol. 43. Berlin: Frank & Timme.
- Puchhammer, Markus (2007): Language-Based Item Analysis – Problems in Intercultural Comparison. In: Hopmann, Stefan Thomas/Brinek, Getrude/Retzl, Martin (eds.): Schulpädagogik und Pädagogische Psychologie: Vol. 6. PISA zufolge PISA – PISA According to PISA: Hält PISA, was es verspricht? – Does PISA Keep What It Promises? Wien: LIT Verlag, pp. 127-137.
- Smolicz, Jerzy J./Nical, Illuminado (1997): Exporting the European Idea of a National Language: Some Educational Implications of the Use of English and Indigenous Languages in the Philippines. In: International Review of Education, 43, 5, pp. 507-526. doi: 10.1023/A:1003098223423
- Thorn, William (2009): International adult literacy and basic skills surveys in the OECD region. OECD Publishing. Advance online publication. doi: 10.1787/221351213600
- Upsing, Britta (2017): Lokalisierung in internationalen Bildungsvergleichsstudien aus informationswissenschaftlicher Sicht: Herausforderungen, Übersetzungen, Prozesse (Dissertation). Stiftung Universität Hildesheim, Hildesheim. <http://d-nb.info/1145140017/34>
- Upsing, Britta/Gissler, Gabriele/Goldhammer, Frank/Rölke, Heiko/Ferrari, Andrea (2011): Localisation in international large-scale assessments of competencies: Challenges and solutions. In: Localisation Focus, 10, 1, pp. 44-57. http://www.localisation.ie/sites/default/files/publications/Vol10_1UpsingGissleretal.pdf
- Van de Vijver, Fons J. R. (2015): Cross-Cultural Research Methods in Sociology. In: Wright, James D. (ed.): International Encyclopedia of the Social & Behavioral Sciences. New York: Elsevier, pp. 322-326. doi: 10.1016/B978-0-08-097086-8.44014-6
- Van de Vijver, Fons J. R./He, Jia (2016): Bias Assessment and Prevention in Noncognitive Outcome Measures in Context Assessments. In: Kuger, Susanne/Klieme, Eckhard/Jude, Nina/Kaplan, David (eds.): Assessing Contexts of Learning. Cham: Springer, pp. 229-253.
- Wagemaker, Hans (2013): International Large-Scale Assessments. In: Rutkowski, Leslie/von Davier, Matthias/Rutkowski, David (eds.): Statistics in the Social and Behavioral Sciences Series. Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis. Boca Raton, NJ: CRC Press, pp. 11-36.
- Wuttke, Joachim (2007): Uncertainties and bias in PISA. In: Hopmann, Stefan Thomas/Brinek, Getrude/Retzl, Martin (eds.): Schulpädagogik und Pädagogische Psychologie: Vol. 6. PISA zufolge PISA – PISA According to PISA: Hält PISA, was es verspricht? – Does PISA Keep What It Promises? Wien: LIT Verlag, pp. 241-263

Index

A

Achievement gap 23, 156, 195f, 198, 205ff
Adult Education 24, 322, 324ff, 329f
AERA (American Educational Research Association) 9, 20ff, 33, 155, 230

B

Benjamin, Walter 358, 364
Big Data 24, 237f, 241, 251f, 263, 265ff
Bildung 52, 340f, 343ff, 348

C

Common Core Standard 12, 302
Comparative education 11, 17ff, 33, 38f
Curriculum 11ff, 15, 69, 71f, 162f, 186, 226, 254, 268, 270, 284, 287, 291, 313, 315, 341

D

Democracy 14ff, 21, 122
Dewey, John 12, 355
Didaktik 340f, 344
Digital data 24, 225, 227, 230f, 237, 242f, 257, 283, 296

E

Educational achievement 9, 19, 103f, 171, 223, 231, 374
Educational administration 34, 164, 225
Educational data 226, 232, 268f, 272f
Educational discourse 14f, 341
Educational governance 306, 321
Educational leadership 34, 37f, 40f, 49, 62, 67, 69f, 73, 217
Education policy 12, 23, 51, 155, 161, 171f, 179f, 188f, 223f, 230f, 235ff, 242, 253, 256, 260ff, 268, 273, 344
Europe 12f, 17, 23, 34, 95, 101, 120, 145, 155, 226, 234, 317, 344, 361, 363f
European Commission 120, 232, 235, 330
European Union 116, 232, 235, 276

F

Fichte, Johann Gottlieb 355, 357

G

Germany 9ff, 31, 33f, 36f, 41, 47ff, 51ff, 62ff, 70f, 73f, 80f, 84ff, 97, 102, 112, 120f, 127ff, 132, 155, 157ff, 165, 172ff, 176, 187, 191, 197, 217, 226, 230ff, 235, 238, 242f, 247, 284, 313, 315, 318, 324, 326ff, 340ff, 378
Global Education Industry 21, 24, 301, 307
Globalization 16, 301ff, 305, 307, 310f, 318
Governance 19, 21f, 24, 37, 42, 84f, 163ff, 171f, 182, 187, 224f, 231, 235, 237, 242ff, 247f, 250, 253ff, 257f, 260, 306, 321f

H

Humboldt, Wilhelm von 11, 343

I

IBO (International Baccalaureate Organization) 307, 311ff
ICCS (International Civic and Citizenship Study) 23, 121, 123ff, 127ff
IEA (International Association for the Evaluation of Educational Achievement) 9, 19, 155, 171, 231, 374f, 380

K

KMK (Standing Conference of the Ministers of Education and Cultural Affairs) 10, 157, 162f, 174f, 231

L

Large-Scale assessment 11, 21, 23, 41, 123, 155ff, 162, 171, 174, 217, 223, 231f, 235, 238, 310, 373ff, 383, 385
Learning analytics 229f, 277, 306

N

New public management 31ff, 157,
303, 307, 325
No Child left behind act 13

O

OECD (Organisation for Economic
Co-operation and Development) 9, 19,
24, 67f, 101, 123, 155ff, 162f, 171f,
182f, 185f, 188ff, 196, 216, 218, 233,
238, 253ff, 258ff, 268, 305, 310, 375,
379ff

P

Pädagogik 338, 340ff, 344
Pedagogical 9, 12, 15, 20, 32, 34, 37f,
41f, 68, 85f, 329f, 339, 252ff, 362
PIAAC (Programme for the
International Assessment of Adult
Competencies) 349, 373ff, 377, 379,
382f, 386
PIRLS (Progress in International
Reading Literacy Study) 267, 375
PISA (Programme for International
Student Assessment) 11, 16, 19, 23f,
123, 155ff, 161ff, 171ff, 178ff, 182f,
185ff, 195, 217, 223, 227, 231, 233,
235f, 238, 253ff, 267, 273, 349, 373ff,
379ff, 385f
Privatization 24, 301ff, 310f, 316,
318f, 324, 329
Public Education 13, 21f, 24, 33, 96,
98f, 104f, 148, 311, 317ff, 368

R

Rousseau, Jean Jaques 354f, 357
Russel, James E. 18

S

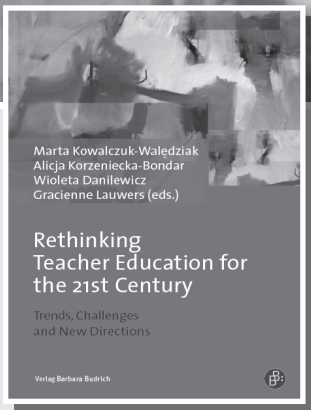
Schleiermacher, Friedrich 339, 344ff
School culture 53, 56, 236
School development 32, 42, 162f
School monitoring 24, 41, 242ff, 247f,
250
School reform 10, 13, 82, 86, 176, 305
School system 10ff, 21, 42, 47, 50,
67f, 85, 87, 173, 176, 235, 258ff, 311,
318, 362
Student achievement 9, 13, 19, 41, 69,
223, 228, 232, 236, 271, 273, 283,
291, 296

T

Teacher education 50, 162, 231, 342
Teacher training 31, 162f, 224, 290ff,
294ff, 313
TIMSS (Trends in Mathematics and
Science Study) 11, 23, 155, 157f, 161f,
171, 195ff, 208, 216ff, 231, 267, 375
Transatlantic 9, 13f, 20, 237
Translation 15, 21, 25, 171f, 175, 179,
187f, 191, 313, 337ff, 343ff, 352ff,
360ff, 368ff, 373ff, 385

U

UNESCO (United Nations
Educational, Scientific and Cultural
Organization) 19, 103, 137, 198, 310,
312, 317
UNHCR (United Nations High
Commissioner for Refugees) 96ff,
137f, 140ff, 148ff
USA 9f, 17ff, 31, 38, 47ff, 51, 54ff,
65, 67f, 71, 73, 88, 254f, 262, 274,
279, 313



Marta Kowalczyk-Wałędziak
Alicja Korzeniecka-Bondar
Wioleta Danilewicz
Gracienne Lauwers (eds.)

Rethinking Teacher Education for the 21st Century

2019 • 402 pp. • Pb. • 76,00 € (D) • US\$105.00 • GBP 67.00
ISBN 978-3-8474-2241-9 • eISBN 978-3-8474-1257-1

This book focuses on current trends, potential challenges and further developments of teacher education and professional development from a theoretical, empirical and practical point of view. It intends to provide valuable and fresh insights from research studies and examples of best practices from Europe and all over the world. The authors deal with the strengths and limitations of different models, strategies, approaches and policies related to teacher education and professional development in and for changing times (digitization, multiculturalism, pressure to perform).

The book is an **Open Access** title (DOI: 10.3224/84742241) , which is free to download or can be bought as paperback.

www.barbara-budrich.net



Petra Strehmel | Johanna Heikka
Eeva Hujala | Jillian Rodd
Manjula Waniganayake (eds.)

Leadership in Early Education in Times of Change

Research from five Continents

2019 • 308 pp. • Pb. • 36,00 € (D) • US\$50.00 • GBP 32.00
ISBN 978-3-8474-2199-3 • eISBN 978-3-8474-1224-3

The collection brings together the latest work of researchers from Australia, Africa, Asia, and Europe focusing on early childhood leadership matters. It covers different aspects of leadership in early education: professional education and development, identity and leadership strategies as well as governance and leadership under different frame conditions.

The book is an **Open Access** title, which is free to download or can be bought as paperback.

www.barbara-budrich.net

International Perspectives on School Settings, Education Policy and Digital Strategies

In the field of education, transatlantic discourse has been shaped by an exchange of ideas for a long time. Over the past two decades, education science has increasingly become internationally networked. Since 2015, the Office for International Cooperation in Education at DIPF | Leibniz Institute for Research and Information in Education has organized international sessions on education research at the Annual Meetings of the American Educational Research Association, thus providing a floor for transatlantic exchange on current research topics. The volume gives an overview of the transatlantic activities in education research with regard to these sessions. It represents a collection of topics ranging from school development to the use of large-scale assessment and digital data in education to questions related to migration and public education, or the economization of education. Moreover, the volume offers a reflection on the gains respectively obstacles linked to international exchange.

The editors:

Dr. Annika Wilmers and Dr. Sieglinde Jornitz are both academic staff at DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany.

ISBN 978-3-8474-2299-0



www.budrich.eu

Picture credits: photo: www.shutterstock.com