Educational Monitoring, Comparative Studies, and Innovation

From evidence-based governance to practice

OECD/CERI regional seminar for the German-speaking countries

in Potsdam (Germany) from September 25 to 28, 2007

Editors:
State Institute for School and Media Berlin-Brandenburg
(LISUM, Germany)

Federal Ministry for Education, the Arts and Culture
(bm:ukk, Austria)

Swiss Conference of Cantonal Ministers of Education
(EDK, Switzerland)

IMPRINT

The following articles are a translation of selected contributions which have been published in the conference publication of the OECD/CERI regional seminar of the German speaking countries in 2007. The book "Bildungsmonitoring, Vergleichsstudien und Innovationen. Von evidenzbasierter Steuerung zur Praxis" has been published by the Berliner Wissenschafts-Verlag GmbH in August 2008.

Cover photo:
Lutz Hannemann,
Internet: http://www.foto-hannemann.de

Edited by:
Agentur für Projekt- und Netzwerkmanagement, Berlin
Internet: http://www.agpronet.de

Translated by:
Susannah Goss and Sorcha O'Hagan

The regional seminar, the conference publication, and this translation of a selection of contributions to the publication was patronized by the German Federal Ministry of Education and Research (BMBF).

The documentation of the regional seminar can be downloaded from
www.bildung-brandenburg.de/ceri.html.

The conference publication was published by the Berliner Wissenschafts-Verlag GmbH, www.bwv-verlag.de, ISBN 978-3-8305-1550-0

INDEX

PREFACE 4

Petra Stanat
Development and Implementation of Educational Innovations as a Consequence of Educational Monitoring, Educational Reporting, and Comparative Studies of Student Performance – Opportunities and Limitations 6

Olaf Köller
When Science Allows Itself To Be Instrumentalized – PISA and the German Debate on School Structure 15

Stefan C. Wolter
Purpose and Limits of a National Monitoring of the Education System Through Indicators 26

Hans Döbert
Educational Reporting in Germany – Or: How Can Indicators Inform Innovation in the Education System? – 27

Werner Specht
The National Educational Report – A Step in the Direction of Evidence-Based Policy in Austria 41

Norbert Maritzen
Educational Monitoring in Germany – A System Innovation to Safeguard Quality Standards 52

Kurt Reusser & Ueli Halbheer
Consequences of Student Assessment Studies and Educational Monitoring,– The Outlook for Educational Standards and their Implementation 63

Jan Hofmann
Reflecting on Developmental Problems in Education Systems:Thirty Years of OCED/CERI Regional Seminars for Germany, Austria, and Switzerland, 1977-2007 73

Contributors 79

PREFACE

From 25th to 28th September 2007, more than 80 representatives of the scientific community, the education authorities, and school practice in Germany, Austria, and Switzerland met in Potsdam to discuss the development and implementation of educational innovations.

Since 1977, OECD/CERI regional seminars have been conducted every two years in the German-speaking OECD countries. A six-year cycle of three seminars, one in each country, previously addressed a single thematic area.

The Potsdam seminar, which was the 16th of its kind, broke with this tradition and was hosted by Germany, which commissioned the state of Brandenburg with the seminar's organization, as a stand-alone event. The reasons for this decision were pragmatic. As a result of the reform of federalism and the dissolution of the Bund-Länder Commission for Educational Planning and Research Promotion (BLK), Germany was undergoing various structural changes that cast temporary doubt on the future of the seminar series. A solution was chosen to continue the successful cooperation among the German-speaking countries, while at the same time leaving the future direction of the seminar series open. The Potsdam seminar will therefore go down in the history of the OECD/CERI regional seminars as a bridge between tradition and the future.

Under the somewhat wordy title "Development and Implementation of Educational Innovations as a Consequence of Educational Monitoring, Educational Reporting, and Comparative Studies of Student Performance – Opportunities and Limitations," the seminar focused on questions of knowledge transfer – from the findings of evaluation and monitoring activities to implementation in schools and classroom practice.

Three main outcomes of the seminar can be identified:

1. Empirical educational research is a prerequisite for quality improvement in all areas of education. Data alone, no matter how extensive, do not tell the whole truth about the system; rather, there is a need for careful interpretation and for reasoned conclusions and recommendations.

2. The importance of networks and supraregional cooperation is growing. As a result of its history and geographical location, Austria is a member not only of the tri-national CERI organization, but also of a group of south-eastern European countries that cooperate in addressing questions of education. Following the reform of federalism in 2006, which shifted more of the responsibility for education policy to the federal states, Germany faces the challenge of cultivating national and international cooperation within the new structures and of redefining the responsibilities of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder and its various bodies. In Switzerland, a similar process has been triggered by the revision of the provisions on education in the Federal Constitution, one effect being the most widespread harmonization of cantonal legislation on compulsory schooling since the Federation was founded.

3. It is becoming increasingly important for the methods and objectives of innovative educational concepts to be properly communicated. Research findings must not only be disseminated to policy makers and used to generate policy-relevant knowledge, but also communicated to practitioners in schools. It is vital that teachers recognize the value of concepts such as “standardization,” “monitoring,” and “evaluation” (e.g., for improving student learning outcomes and making their own work more easily manageable).

This selection of articles from the seminar’s documentation contains the country-specific reports prepared for the seminar, selected presentations, and a general report on the seminar and its outcomes.

We thank all those involved in preparing, conducting, and documenting the seminar. A particular debt of thanks is owed to the Federal Ministry of Education and Research in Berlin, which funded the Potsdam seminar and which has made it possible for the findings to be made accessible to a broader audience.

Hans Ambühl,

Secretary General, Swiss Conference of Cantonal Ministers of Education (EDK), Berne, Switzerland

Dr. Anton Dobart,

Director General of the Division for General Education, Educational Planning, and International Affairs, Federal Ministry for Education, the Arts and Culture, Vienna, Austria

Dr. Jan Hofmann,

Director of the State Institute for School and Media Berlin-Brandenburg (LISUM), Ludwigsfelde, Germany

Development and Implementation of Educational Innovations as a Consequence of Educational Monitoring, Educational Reporting, and Comparative Studies of Student Performance – Opportunities and Limitations

Petra Stanat

1. Objectives and Structure of the Regional Seminar

In the German-speaking countries, international assessments of student performance have stimulated a number of innovations designed to improve and maintain the quality of education systems. These innovations include continuous monitoring of educational outcomes at the system level, regular educational reporting, and increased funding of educational research. It is expected that such measures can help to identify target points for interventions that will, in turn, help to optimize the system. Translating the findings of educational monitoring into interventions is by no means a trivial task, however. It is rarely possible to draw direct conclusions from the findings of comparative educational assessments as to where reforms should be introduced. Likewise, it remains unclear under which conditions proposed innovations can be implemented, whether and in what way they will change educational practice, and to what extent they will actually produce the intended effects. In addition to their intended outcomes, policy and practice based on educational monitoring can also have negative side effects; for example, monitoring may result in a narrowed focus on those goals of an education system that are relatively easy to measure. Moreover, international assessments of student performance can encourage a culture of innovation that is largely reactive, thus undermining the capacity of education systems to initiate and implement innovations from within.

The OECD/CERI regional seminar for the German-speaking countries held in Potsdam from 25th to 28th September 2007 aimed at analyzing the relationship between educational monitoring, educational research, and innovations in education systems. The seminar involved presentations, panel discussions, working group discussions, and visits to various educational institutions. It was opened by Dr. Jan Hofmann, Director of the State Institute for School and Media Berlin-Brandenburg (LISUM), Ludwigsfelde, who was also responsible for organizing the Potsdam seminar. Opening words were then given by Burkhard Jungkamp (State Secretary, Ministry of Education, Youth, and Sports of the Federal State of Brandenburg, and member of the Commission of Deputy Ministers on “Quality Assurance in Schools”), Michael Thielen (State Secretary, Federal Ministry for Education and Research, Berlin), Tom Schuller (Head of the Centre for Educational Research and Innovation CERI, Paris), Prof. Dr. Erich Thies (Secretary General, Standing Conference of the Ministers of Education and Cultural Affairs of the Länder, Bonn), Dr. Anton Dobart (Federal Ministry for Education, the Arts and Culture, Vienna), and Hans Ambühl (Secretary General of the Swiss Conference of Cantonal Ministers of Education, Berne).

In four plenary presentations, Prof. Dr. Olaf Köller (Institute for Educational Progress, Berlin), Mag. Dr. Werner Specht (Federal Institute for Educational Research, Innova-

tion, and Development of the Education System, Graz), Oberschulrat Norbert Maritzen (Department of Education and Sports, Hamburg), and Prof. Dr. Kurt Reusser (Institute of Education, University of Zurich) analyzed the theme of the seminar from different perspectives. Their presentations and the ensuing discussions form the basis of this general report. The report also draws on accounts of educational reporting in Switzerland (Prof. Dr. Stefan C. Wolter, University of Berne), Germany (Prof. Dr. Hans Döbert, German Institute for International Educational Research, Frankfurt am Main), and Austria (Mag. Dr. Werner Specht, Federal Institute for Educational Research, Innovation, and Development of the Education System, Graz). In workshops, the seminar participants had the opportunity to engage further with the implications of educational assessments and educational monitoring and their implementation at different levels (education systems, individual schools/classrooms, transitions, pre-service and in-service teacher training). Visits to scientific institutes and educational institutions in the Berlin/Brandenburg region offered participants different perspectives on the implementation of educational standards. The institutions visited include the Institute for Educational Progress (IQB), the Berlin-Brandenburg Institute for School Quality (ISQ), the State Institute for School and Media Berlin-Brandenburg (LISUM), and the Max Planck Institute for Human Development. The seminar concluded with a panel discussion moderated by Dr. Jeanne Rubner *(Süddeutsche Zeitung)* and featuring the following representatives of educational research, education policy, and administration: Prof. Dr. Ingrid Gogolin (University of Hamburg), Dr. Hans-Gerhard Husung (State Secretary, Senate Department for Education, Science, and Research, Berlin), Burkhard Jungkamp (State Secretary, Ministry of Education, Youth, and Sports of the Federal State of Brandenburg; member of the Commission of Deputy Ministers on “Quality Assurance in Schools”), Dr. Stefan Luther (Federal Ministry of Education and Research, Berlin), DDr. Erwin Niederwieser (member of the National Assembly; spokesman on education for the Social Democratic Party SPÖ, Vienna), Hans Ulrich Stöckling (Education Minister, Canton of St. Gallen, former President of the Swiss Conference of Cantonal Ministers of Education, EDK, Berne).

This general report uses a model of interrelationships between accountability systems and educational outcomes developed by Joan Herman (2005), which was slightly modified for the present purposes (see Fig. 1), as a structure for summarizing the plenary presentations and the discussions they generated. The system of interrelationships was considered from various perspectives during the seminar, with attention focusing on the paths highlighted with bold type in Figure 1. In the following, the discussions pertaining to these paths are summarized in three parts. At the end of each section, the issues raised are illustrated by reference to the example of the situation of immigrant students with limited proficiency in the language of instruction (in the following referred to as language minority students). Improving the educational provision for these students has been identified by all three of the participating countries as a particularly pressing challenge.

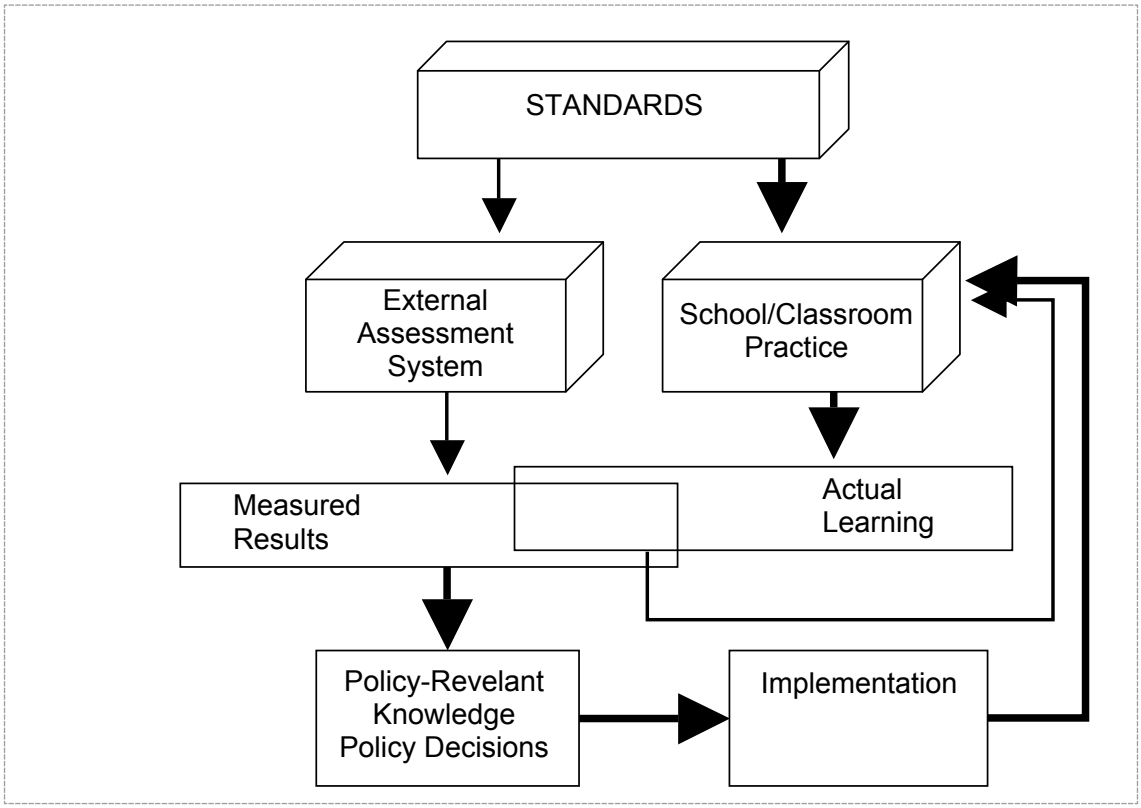

Fig. 1: Interrelationships between accountability and outcomes in educational systems (based on Herman, 2005)

2. From Measured Results to Policy-Relevant Knowledge and Policy Decisions

One of the key issues discussed at the seminar was how measured results can be translated into policy-relevant knowledge and policy decisions. This pertains to the results of educational monitoring and educational reporting as well as to findings from educational research that may inform policy decisions. The seminar participants discussed the requirements that must be met by data to be used in evidence-based policy, opportunities for and constraints on drawing lessons from the findings, and how the relationship between educational research and education policy should be configured to enhance the quality of evidence-based decisions and governance processes.

In the discussion on the quality of data used in evidence-based policy making – an issue addressed in the presentations of Olaf Köller and Werner Specht – attention was drawn to the quality assurance procedures that are standard practice in science, particularly the peer review procedure that forms the basis for decisions on project funding and publications. This procedure has proved highly effective and is largely undisputed in empirical educational research. It is now becoming more widely accepted in education policy as well, with project funding increasingly being allocated through open calls for tender, rather than on a discretionary basis. This development should help to improve the quality of thematic research intended to inform policy decisions.

There was, however, some disagreement on whether and how facilitating the scientific community's access to educational research data can enhance the quality of data-driven policy-relevant knowledge. In Germany, research data centers are currently being established to make social and economic datasets readily available for reanalysis. It is hoped that the available information will be analyzed more thoroughly and that the evidence

base will be optimized in an ongoing process of scientific exchange. However, there is also a risk of inexperienced individuals using inappropriate procedures to analyze the highly complex datasets and publishing misinterpretations that are subsequently difficult to correct. Such actions might damage the reputation of educational research as a whole. At the same time, however, it should be possible to mitigate this risk by, for example, providing detailed instructions on how to handle the data, ensuring the consistent application of scientific standards for publications, and putting in place support services to help officials and media representatives interpret the findings of empirical educational research.

Even findings from analyses that were conducted according to strict scientific criteria are frequently mis- or over-interpreted. This is essentially always the case when direct implications for policy decisions or classroom practice are drawn from such results. Olaf Köller described one such disputed interpretation in his presentation on the school structure debate in Germany, which has been reignited by the PISA results. The finding that social disparities in educational attainment are especially pronounced in Germany is often attributed to the early tracking of students to different secondary school types. As a result, widespread calls have been made for reform of the school structure. The PISA data do not provide evidence in support of such reforms, however. In fact, other studies, such as BIJU and COACTIV (e.g., Köller & Baumert, 2008; Kunter et al., 2005), have identified the differing instructional cultures of the secondary tracks as the main factor increasing social disparities in students' achievement gains. More generally, it is important to consider and integrate the results of all relevant studies when interpreting the findings of empirical educational research.

Given the marked increase in the amount of data being produced by empirical educational research, however, taking all relevant sources into account is anything but trivial. Among other things, it remains unclear what kind of evidence should be considered relevant. Correlational studies generally involve a high degree of uncertainty, and there is no indication whether findings obtained within one school system can be generalized to others (e.g., the findings described by Olaf Köller on the relative effects of schools and non-school environments for the development of disparities in the United States). For some years now, policy decisions on the introduction of educational programs in the United States have increasingly drawn on field experiments with random assignment to groups (randomized field trials); such research designs are known to yield relatively well-founded results on the effectiveness of interventions (Slavin, 2002). This kind of approach remains rare in the German-speaking countries, however, and its role in the context of educational research is not undisputed in the United States either (e.g., Olson, 2002).

It is not just the growing amount of data that makes it difficult for policy makers and practitioners to develop an informed understanding of research findings and their implications, but also the complexity of those data and the methods of analysis used. All three of the countries participating in the seminar see room for improvement in communication between the scientific community and educational policy makers. Werner Specht identified a shortage of "professional interpreters" of the data produced by empirical educational research. In Great Britain, brokerage agencies have been set up in response to this need. The seminar participants agreed that communication between science and policy makers must be improved. Whether this can only be achieved by setting up a new institution, such as a brokerage agency, was a matter of discussion, however.

Appropriate interpretation of the data generated by monitoring systems, educational reporting, and educational research depends on all those involved recognizing the role that these findings can play in policy decisions. The evidence generated by educational research can help ensure that policy decisions are increasingly well-founded and make their implications more transparent. Concrete, clear-cut conclusions can rarely be derived from the data obtained, however. Decisions in education have always been and will continue to be decisions involving a high degree of uncertainty. They have to allow for normative considerations and issues of acceptance. In other words, educational research can generate information that is useful for policy decisions, but it cannot relieve policy makers from the responsibility of making those decisions.

With respect to the illustrative example of language minority students, international studies of student performance have identified an urgent need for intervention in all three of the German-speaking OECD countries. Some findings, such as the pronounced effect of the home language on achievement, suggest that interventions should target students' command of the language of instruction. The international studies cannot provide insights into how best to achieve this objective, however. Rather, systematic analyses of the effectiveness of various approaches to develop students' language skills are needed; at present, there is a dearth of such research (Limbird & Stanat, 2006). Moreover, a weak command of the German language, low SES, and low levels of parental education do not seem to be the only factors contributing to the underperformance of some minority groups, such as young people of Turkish origin in Germany (Müller & Stanat, 2006). Which additional factors play a role remains to be determined.

The findings of international studies of student performance can thus identify possible *points of intervention* for policy decisions and reforms intended to support the development of young people from language minority groups. The data do not allow the nature of this intervention to be further specified, however. Additional studies, especially intervention studies, are therefore required. Given that language minority students are often severely disadvantaged, however, policy makers can hardly sit back and wait for the findings of these studies before taking action. As a result, various measures have already been introduced, the effectiveness of which remains uncertain. When these programs are implemented across the board, as has been the case for many preschool language intervention programs, it is difficult to gauge their effectiveness after the fact, as there are no suitable comparison groups. And even if there is evidence showing that a certain approach to improving students' command of German as a Second Language works, it is still necessary not only to determine the most effective and efficient means of implementing this approach, but also to provide an appropriate response to the widespread calls for home language support in schools. Given that there is currently no empirical evidence for the efficacy of such measures, the decision remains a normative one that reflects theoretical and policy-related considerations – specifically, the role of school in a mixed-culture society and the importance of the host society acknowledging minority group cultures (Stanat, in press). Likewise, the questions of whether and how schools should recognize and acknowledge the religious background of students from immigrant families – questions that have produced very different responses in different countries – can only be answered from a normative perspective.

3. From Policy-Relevant Knowledge and Policy Decisions to Implementation

The implementation of policy decisions is another challenge that was discussed extensively at the OECD/CERI seminar. The complexity of this transfer was addressed in the presentations by Norbert Maritzen and Kurt Reusser, which are documented in this volume, as well as in various workshops. Because the findings of educational monitoring almost never have clear-cut implications for measures to be implemented in areas identified as problematic, these data alone cannot justify policy decisions. Rather, analytic studies are needed to examine the efficacy of measures and ways to optimize their effects. When innovations are to be implemented across the board, knowledge of the conditions under which transfer processes are likely to succeed is also vital. Precisely this kind of process knowledge is often lacking, however, with negative consequences for evidence-based governance and its justification. Reforms tend to be implemented on the basis of plausibility assumptions rather than on the strength of evidence for their effectiveness and transferability.

Norbert Maritzen used the example of the introduction of educational monitoring systems to illustrate the challenges involved in the implementation of educational reforms. He drew attention to a fundamental problem that is well known from research on implementation and transfer, namely the widespread and largely implicit assumption that change can be strategically implemented from the top down. As a result, stakeholders and structural context factors – both of which are decisive for the successful implementation of innovations – are often ignored. According to Kurt Reusser, the danger of this kind of top-down approach is that acceptance of and compliance with reforms remains superficial (see also deLeon & deLeon, 2002).

To ensure successful implementation of innovations, it is important to involve stakeholders at various levels of the system in an iterative reform process, such that they make the necessary adjustments, engage in the necessary learning processes, and take responsibility for the reform process as a sense of ownership emerges. Kurt Reusser emphasized that processes of change in systems, institutions, and individuals tend to be slow-moving because they challenge many deeply anchored beliefs and habits. It is therefore important to allow enough time for the reforms to take effect and to put appropriate support systems in place.

The example of implementing measures to improve the educational outcomes of students from language minority backgrounds also presents numerous challenges. For example, staff need to be trained to provide targeted language support. The challenges of implementing these measures in preschools, in particular, but also in schools, should not be underestimated.

Preschool institutions see the new demands they are facing in the area of language support as an additional responsibility that requires increased resources to be made available. This aspect must be taken into account in the implementation of language support measures. Moreover, the new language support measures reflect an increased emphasis on the educational role of preschool (as opposed to the care role), which some perceive as an attempt to redefine the function of these institutions. Reforms of this kind can only succeed if all stakeholders are involved in negotiating a new understanding of

what preschool institutions are expected to achieve. Without such a process of negotiation and involvement, it seems unlikely that the intended reforms can be successfully implemented on the long term.

4. From Implementation to School/Classroom Practice and to Learning Outcomes

The introduction of systems of educational monitoring and educational reporting has resulted in greater transparency of decision-making processes and governance in education. The data generated by these systems allow for possible points of intervention to be identified and provide a basis for the measures implemented to be evaluated. However, it remains unclear to what extent the changes implemented by educational policy makers really take effect in school and classroom practice and which supporting measures are needed to achieve the intended effects. For example, little is known about how schools and teachers respond to feedback on their students' performance in educational assessments or to the reports of school inspectors. Further, it remains unclear in what form this feedback should best be delivered to stimulate the intended processes of instructional development. Feedback provided within the context of accountability systems can have a whole range of negative side effects, as shown by the literature on "teaching to the test" and "test score inflation" (Koretz, 2002).

One outcome of the increased transparency afforded by educational monitoring and educational reporting is that an increasingly informed public is now demanding improvements in the education system and evidence for the effects of measures introduced. Werner Specht sees this development as one of the most important effects of output-driven governance systems. Change-sensitive measurement tools capable of assessing the effects of the measures introduced are therefore necessary for accountability purposes. Given the potential problems entailed in repeated measurements (e.g., test repetition and training effects, regression to the mean), these are among the most difficult methodological challenges facing empirical social research. In particular, little is known about the effects that the implementation of policy decisions can be expected to have at the system level, and experiences with assessing such changes are still very limited. It thus remains unclear to what extent the current reform efforts will actually show up in the results of future educational assessments.

With respect to the example of improved language support, the effects of the measures introduced should become apparent in the improved educational attainment of language minority students. It remains to be seen whether and to what extent this goal can be achieved. One potential problem in this context is that various measures are currently being put in place to improve the quality of education for all students. It is conceivable that these measures might raise the overall level of student performance without closing the performance gap between language majority and language minority students, and that the specific effects of measures targeting language minority students might not be discernable. The question also remains of how much change must be achieved for outcomes to be considered satisfactory. The literature on the No Child Left Behind Policy in the United States shows how difficult it is to monitor the extent to which minimum standards are being achieved not only by the entire student population, but also

within subgroups (e.g., Linn, 2002). In the German-speaking countries, minimum standards are yet to be defined. As such, the performance gap between language majority and language minority students is generally used as an indicator for the effectiveness of measures targeting minority students. This approach implies that the situation can be considered satisfactory when the performance gap has closed completely. Because language minority students in the German-speaking countries often come from socially and educationally disadvantaged families, however, it is questionable that this goal is realistic. An alternative approach would be to take students' social background into account when determining and evaluating the performance gap. Inasmuch as the social situation of immigrant families depends in part – at least from the second generation onward – on the effectiveness of the educational provision for language minority students, however, such an approach would in effect promote the development of an underclass, which is hardly a desirable outcome.

As illustrated by this example, the outcomes of education systems must be evaluated from a normative point of view that are based on specified goals and targets. In addition to aspects of excellence and equity, it is important to include goals relating to the responsibility of our educational institutions to raise independent-thinking and conscientious citizens who are able to participate responsibly in the democratic processes of a pluralistic society. Such aspects are very difficult to measure in the context of educational monitoring and educational reporting, and it will be important to ensure that they are not lost from view as new models of governance are developed and implemented.

References

deLeon, P., & deLeon, L. (2002). Whatever happened to policy implementation? An alternative approach. *Journal of Public Administration Research and Theory*, 12, 467-492.

Herman, J. (2005). Making accountability work to improve student learning. *CSE Report 649*. Los Angeles, CA: CRESST.

Köller, O., & Baumert, J. (2008). Entwicklung von Schulleistungen. In Oerter, R. & Montada, L. (Eds.). *Entwicklungspsychologie* (6th ed., pp. 735-768). Weinheim: Beltz/PVU.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 38, 752-777.

Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., Jordan, A., & Neubrand, M. (2005). Der Mathematikunterricht der PISA-Schülerinnen und -Schüler. Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft, 8*, 502-520.

Limbird, C., & Stanat, P. (2006). Sprachförderung bei Schülerinnen und Schülern mit Migrationshintergrund: Ansätze und ihre Wirksamkeit. In Baumert, J., Stanat, P., & Watermann, R. (Eds.). *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 257-307). Wiesbaden: VS Verlag für Sozialwissenschaften.

Linn, R. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*, 3-16.

Müller, A. G., & Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei. In Baumert, J., Stanat, P., & Watermann, R. (Eds.). *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 223-255). Wiesbaden: VS Verlag für Sozialwissenschaften.

Olson, D. R. (2004). The triumph of hope over experience in the search for “what works”: A response to Slavin. *Educational Researcher, 33*, 24-26.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*, 15-21.

Stanat, P. (in press). Heranwachsende mit Migrationshintergrund im deutschen Bildungswesen. In K. S. Cortina, K. S., Baumert, J., Leschinsky, A., Mayer, K. U. & Trommer, T. (Eds.). Das *Bildungswesen in der Bundesrepublik Deutschland*. Reinbek: Rowohlt.

When Science Allows Itself To Be Instrumentalized – PISA and the German Debate on School Structure

Olaf Köller

Introduction

In Germany, the findings of PISA 2000 and PISA 2003 (Deutsches PISA-Konsortium, 2001, 2002, 2004, 2005) have re-ignited discussion on the systematic discrimination of students from socially disadvantaged and minority backgrounds. The relationship between social background and reading literacy in PISA 2000 is closer in Germany than in any other OECD country (Baumert & Schümer, 2001). In the political discussion, at least, the putative cause of this performance gap was soon diagnosed: Germany's tracked secondary school system offers children and adolescents from socially disadvantaged and minority backgrounds less favourable learning opportunities than those made available to their socially advantaged peers. The discussion culminated in the report submitted by Vernor Muñoz, Special Rapporteur on the right to education, to the United Nationals Commission on Human Rights in Geneva. On 31st March 2007, Germany's largest-circulation broadsheet, *Süddeutsche Zeitung*, headlined its online coverage of the report with "UN Inspector Condemns German School System."

Some scientific publications (e.g., Hanushek & Wößmann, 2006; Wößmann, 2007) based on reanalyses of data from international assessments of student performance also suggest that studies such as PISA or PIRLS (Bos et al., 2003) are capable of pinpointing the causes of the racial/ethnic and socio-economic disparities in student performance. In contrast, this article argues that large-scale international assessments and the reports based on them cannot provide scientifically founded insights into the causes of these disparities. The aforementioned studies were designed for the purposes of system monitoring, to provide education policy makers in the participating countries with indicators that can inform policy decisions. They were not designed to test complex causality hypotheses (e.g., on how the school structure affects disparities in student performance). As such, studies such as PISA and PIRLS can at most provide indications of where problems may lie, but they cannot determine causality. Conclusions drawn from the results of these studies thus tend to be political and not scientific.

This article presents findings from selected empirical studies that are far better suited to shed light on the relationship between school structure and disparities in student performance (for an overview, see Maaz et al., 2006). The underlying message is not that school structures are irrelevant to social disparities, but that closer inspection of research findings shows that the debate prompted by the PISA findings has failed to explore the issues in their full complexity.

Following a brief historical outline of the debate on school structure in Germany, relevant findings from PISA 2000 are summarized. Empirical findings are then presented to show that the role of school structures in racial/ethnic and socio-economic disparities cannot be determined without considering the educational decisions made by parents and elementary school teachers. Finally, findings are presented to show that measures such

as a later transition to secondary school have certainly not had the intended effects in the past.

1. The Structure of the German School System: A Recurrent Subject of Debate Since World War II

Four lines of argument in the debate on Germany's tracked secondary school system can be said to have characterized the political and scientific discussion of the 1960s and to have formed the basis for reforms to the tracked school system (Köller, 2003):

– A first line of argument deriving from educational psychology questioned the early selection of students to secondary tracks after 4th grade, the very limited means to correct these early tracking decisions, and the lack of scope to provide individualized learning opportunities within the three tiers of the secondary system.

– The second line of argument, again criticizing the processes of selection within the tracked system, was voiced by economists of education. Comparative international studies predicted a shortage of qualified labor that would threaten Germany's long-term economic competitiveness. All political parties were convinced of the need to avert this imminent "educational catastrophe" (Picht 1964) by tapping into reserves of talent and ability. The putative cause of the problems was soon diagnosed: the three-tier secondary school system was completely out of date; its inflexible and impermeable structure failed to meet the needs of a rapidly changing industrial and scientific society. Precisely this criticism has emerged again in recent years (e.g., OECD 2007).

– This diagnosis prompted theoretical debate on the advantages and disadvantages of opening up educational pathways, on the subject canon, and on the minimum duration of school attendance. The subsequent discussion of curricular and instructional reforms brought about the attempt in the 1970s to implement a scientific orientation as a common long-term aim of a tracked school system.

– A fourth and final line of argumentation drew on the previous criticism that students were selected into the outdated three-tier system at too early an age, but gave it a decidedly socio-political slant: social inequalities in educational participation were seen to be a direct result of the structure of the tracked system. An era of educational reform was motivated by the desire to reduce social, ethnic, and regional disparities in educational participation and outcomes.

In the years that followed, almost all measures of structural reform (e.g., the reform of rural schools, vocational-track *Hauptschule*, and comprehensive schooling, as well as the reintegration of students with special educational needs into mainstream schooling) were in part socially motivated. The opening and expansion of college-bound education was also expected to promote social equality (Dahrendorf 1965a, 1965b). Against this background, analyses of the microcensus data published in the mid-1980s proved all the more disappointing. Social disparities in educational participation were extremely persistent. Students of all social backgrounds seemed to have benefited to a similar degree from the expansion of the education system, but there had been no substantial change

in the structure of existing inequalities. In general, students' chances of attaining higher educational qualifications increased, but the relationship between social background and educational participation remained largely stable (Handl, 1985; Köhler, 1992).

Analyses covering a longer timeframe show that the relationship between social background and educational participation had weakened in certain countries, however. The effects were identified first and most clearly in Sweden and then in other countries (Leschinsky & Mayer, 1999). In Germany, the weakening of the link between social background and educational participation was most readily discernable immediately after World War II and in the 1950s – that is, before measures to reform the education system were implemented (Henz & Maas, 1995; Müller & Haun, 1994). Schimpl-Neimanns' (2000) reanalysis of microcensus and census data provides valuable insights into changes in the structure of social disparities in educational participation. Schimpl-Neimanns' results can be summarized as follows:

- The data do not support the hypothesis of unchanged social inequalities in educational participation. Despite high levels of stability in the overall pattern of social disparities, social inequality in some segments of the social structure has been reduced.

- Reductions in social disparities were seen primarily up to the end of the 1970s. In particular, the occupation of the head of the family was no longer as important; there was less change in the role of parents' educational levels.

- In the course of this development, it is particularly the socially discriminating effect of deciding between vocational-track *Hauptschule* and intermediate-track *Realschule* that has decreased. Children from less advantaged socio-economic backgrounds have benefited particularly from the expansion of lower secondary education.

- In contrast, there was little change in social disparities in *Gymnasium* attendance. In fact, the association between parents' educational levels and attendance of *Gymnasium* vs. *Realschule* seem to indicate increasing levels of inequality.

This summary of Schimpl-Neimanns' findings describes the situation in 1989. Because subsequent microcensus surveys did not assess the type of school attended by respondents' children, 1989 was the last year in which microcensus data could be used for such analyses. However, the data obtained from a nationally representative sample of 15-year-olds in the context of the PISA study (Deutsches PISA-Konsortium, 2001, 2002, 2004, 2007) allow the current situation to be examined. Baumert and Schümer (2001) analyzed the PISA 2000 data and found that Schimpl-Neimanns' conclusions remain valid. For example, over 50% of children of higher grade professionals were enrolled in a *Gymnasium*, compared with only about 10% of children of semiskilled and unskilled manual workers. Baumert and Schümer used multinomial logistic regression analyses to determine how social background influences educational participation, taking simultaneous account of the tracking options available in the German secondary system. These analyses revealed that children of higher grade professionals were around six times more likely to attend *Gymnasium* than were children from working class families.

PISA 2000 revealed particularly severe social disparities in students' reading literacy scores in Germany, with only marginal improvements being seen in PISA 2003 and PISA 2006 (Deutsches PISA-Konsortium, 2007).

2. Causes of Social Disparities in the Education System

Given the stability of social disparities in educational participation and educational outcomes in Germany, it might seem reasonable to attribute these inequalities to the institution of school itself and to conclude that working class children continue to be socially discriminated by the education system (Rolff, 1997). The empirical evidence for this hypothesis is extremely weak, however. Analyses of longitudinal data, without which it is impossible to test hypotheses of this kind, provide little support. Based on a longitudinal study of elementary schools in the United States, Entwisle and colleagues (Alexander & Entwisle, 1996; Entwisle & Alexander, 1992, 1994) found that – in comparison with learning in out-of-school contexts – schools serve to reduce disparities. More specifically, the analyses showed that the achievement trajectories of children from different social backgrounds are parallel during the school term, but that the achievement gap widens over the summer vacation: the attainment levels of children from less advantaged socio-economic backgrounds decrease, whereas children from more advantaged homes are able to maintain or even improve their levels of attainment. Over the course of the school career, the repeated alternation between homogeneous institutional learning opportunities and out-of-school environments holding very different opportunities for learning has a cumulative effect on the performance gap.

Reanalyses of data from a study by Trautwein, Köller, Schmitz, and Baumert (2002) (see also Köller & Baumert, 2008) support this argument. The authors examined achievement gains in 7th-grade mathematics using data from N = 1971 students in 125 classes. The mathematics achievement of the participating students was assessed at the beginning and end of the school year. An indicator of the prestige of their parents' occupations (SIOPS; Treiman, 1977) allowed inferences to be drawn about the students' social background. The main results of the multilevel analyses can be summarized as follows:

- The largest achievement gains were observed at *Gymnasium* schools, where knowledge gains over the period of observation were one standard deviation higher than in other school types.

- Students with higher cognitive abilities learned significantly more than students with lower cognitive abilities.

- When school type and cognitive abilities were controlled, social background no longer had a significant effect. In other words, within the same school type and given the same level of cognitive ability at the beginning of the school year, students from less advantaged socio-economic backgrounds learned exactly as much as their more socially advantaged peers.

These analyses thus identified, at least in part, the source of the social disparities in student achievement: the school types implemented in the German secondary system constitute differential developmental environments (Baumert, Trautwein, & Artelt, 2003), with students at *Gymnasium* schools learning significantly more. Because socially disadvantaged children are much less likely to be enrolled in *Gymnasium*, they cannot benefit from these positive effects and consequently fall behind their socially privileged peers, who are more likely to attend *Gymnasium*. If they succeed in gaining a place at a *Gymnasium*, however, socially disadvantaged students do just as well as their socially privileged peers.

Findings from investigations of the transfer recommendations made by teachers at the end of elementary schooling seem to provide the most compelling support for the hypothesis that children from lower social classes are systematically, though not necessarily intentionally, discriminated. Ditton (1992), Lehmann, Peek, and Gänsfuß (1997) and Bos and colleagues (2003) showed that children from less advantaged social backgrounds are less likely to be recommended for *Gymnasium* than are children with the same levels of academic achievement, but from more advantaged backgrounds. In other words, they had to meet higher standards to win admission to *Gymnasium*. It seems likely that elementary teachers' transfer recommendations are not guided by students' academic achievement alone, but that their predictions about students' future development are impacted by other variables that in turn covary with social background.

Drawing on longitudinal data from schools in Bavaria, Ditton and Krüsken (2006) reported some remarkable findings on the transition to secondary school. The Bavarian system is of particular interest in the present context: PISA 2000 showed Bavaria to be the German state with the greatest social disparities at the transition to *Gymnasium*, and Bavaria is one of the few German states in which the tracking decision is linked by law to students' grades at the end of elementary schooling. Three findings merit particular consideration.

- First, Ditton and Krüsken (2006) found that social disparities in learning gains increased from 3rd to 4th grade.

- Second, analyses of the transition to secondary education showed that level of educational attainment at the end of 4th grade is the main determining factor in the tracking decision.

- Third, family background does not play a role above and beyond educational attainment unless a student's grade point average at the end of elementary schooling is borderline. When students' grades do not make a clear case for enrolment in one school type rather than another, children from advantaged socio-economic backgrounds are indeed more likely to be enrolled in higher tracks. The analyses of Ditton and Krüsken cannot determine whether this is attributable to the higher educational aspirations of socially privileged parents or to the tracking recommendations of elementary teachers.

In summary, at least four conclusions can be derived from the empirical findings presented thus far:

1. There is strong empirical evidence for social disparities in educational attainment and educational participation.

2. In Germany, disparities in learning gains seem to be attributable primarily to the fact that the lower secondary tracks provide differential learning environments.

3. Disparities in educational participation emerge primarily when there is a degree of uncertainty about students' future development at the transition to secondary education.

4. Social disparities at the transition to secondary education seem to be largely attributable to the educational decisions made by parents and elementary school teachers.

3. Can Later Tracking or Abolition of Tracking Counter Social Disparities?

Two related changes to the structure of the German school system are being discussed in the wake of the PISA findings: first, deferment of tracking to prevent early allocation to the wrong track; second, the complete abolition of external tracking at the secondary level. The existence of empirical findings indicating that these models are not necessarily a blueprint for success has been ignored. These findings are outlined below.

Deferment of Tracking?

Based on the findings of PISA-E, the German national extension to the PISA study (Deutsches PISA-Konsortium, 2002, 2005), it has been argued (e.g., by Wößmann, 2007) that the results from states implementing six years of elementary schooling (i.e., Berlin and Brandenburg) demonstrate that later tracking serves to reduce social disparities in educational participation. Indeed, the PISA-E 2003 findings show that social disparities in *Gymnasium* attendance are lower in Berlin and Brandenburg than in any of the other 14 states, where students receive only four years of elementary schooling (Deutsches PISA-Konsortium 2005). In Brandenburg, the same applies to social disparities in educational attainment.

In both states, however, the reduction in social disparities seems to come at the expense of the overall level of educational attainment. In the domain of mathematics literacy, for example, Brandenburg and Berlin rank 12th and 13th of the 16 states in the PISA-E 2003 performance tables. *Gymnasium* students rank 13th (Brandenburg) and 15th (Berlin). This pattern of findings is by no means new. Roeder (1997) reanalyzed data from a study of *Gymnasium* schools conducted at the Max Planck Institute for Human Development at the end of the 1960s with a sample representative for all West German states prior to reunification. The study was initiated to examine whether deferred tracking (after 6th grade) had unfavorable effects on the learning outcomes of high-performing students, compared with tracking after four years of schooling. Roeder examined the performance of 7th-grade *Gymnasium* students in German, mathematics, and English, comparing a state in with six years of elementary schooling with states implementing four years of elementary schooling. Figure 1 shows the results of this study.

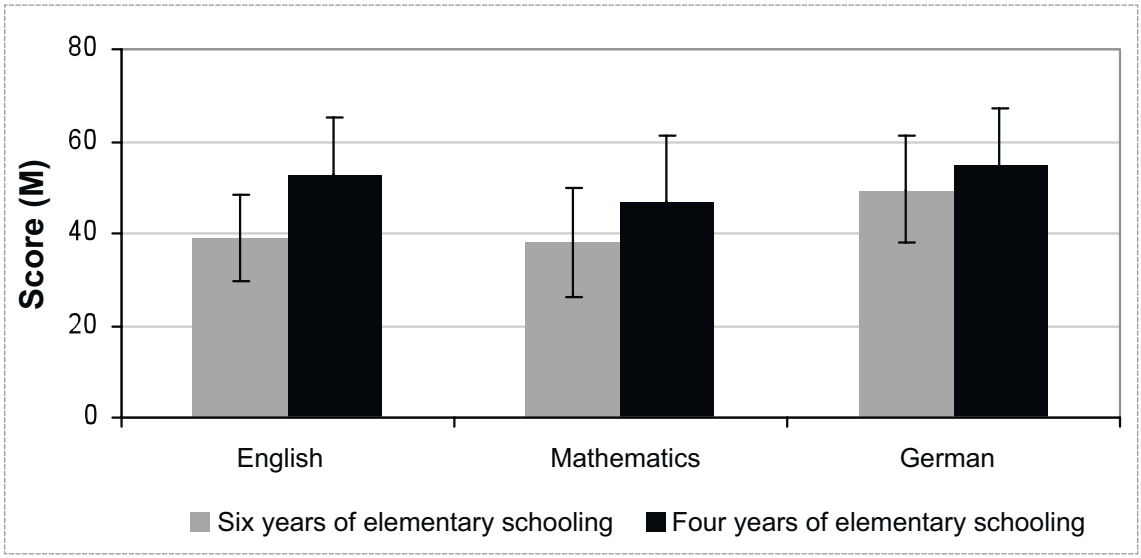

Fig. 1: Performance of 7th-grade *Gymnasium* students by years of elementary schooling (based on Roeder 1997)

The findings clearly show that *Gymnasium* students in states with four years of elementary schooling outperformed their peers in states with six years of elementary schooling. In other words, early tracking seems to be more conducive to the performance of higher-achieving students in all three subjects. The reanalyses of the data from the Trautwein et al. study (2002) outlined above told the same story.

In summary, later tracking does seem to reduce social disparities in student performance in the German school system, but at the expense of overall levels of educational attainment. These effects may well be attributable to the specific instructional culture of *Gymnasium* schools and to the fact that teacher candidates in Germany receive track-specific training, as discussed below.

Abolition of Track-Specific Teacher Training?

In view of the success of countries such as Finland in international educational assessments, calls have been made for all forms of tracking at the lower secondary level to be abolished. Indeed, Finland has succeeded in attaining high levels of student performance with very little external tracking at the lower secondary level. However, there is little point in discussing the model of the unitary school system in Germany without considering the system of pre- and in-service teacher training that is currently in place. In all 16 German states, pre-service (and often in-service) training for secondary teachers has traditionally been track-specific. For *Gymnasium* teacher candidates, there is a much stronger focus on content knowledge, often at the expense of pedagogical content knowledge and teaching methodology. The underlying idea is to ensure a good fit between the teacher and the students they will teach. Considerable differences are thus to be expected in the professional knowledge of teachers at different school types. The COACTIV project (Krauss et al., in press), which was embedded in PISA 2003, systematically examined the relationships between mathematics teachers' content knowledge and pedagogical content knowledge and their students' achievement. Only a small selection of the project's many findings can be reported here:

- Content knowledge is the key precondition for pedagogical content knowledge (correlation r = .79),

- *Gymnasium* teachers have much higher levels of content knowledge than teachers in other school types (mean differences above one standard deviation),

- Likewise, *Gymnasium* teachers have much higher levels of pedagogical content knowledge than teachers in other school types (mean differences above half a standard deviation).

Disregarding for a moment the fact that the teachers were trained to cater for students of different ability groups, the findings of Baumert and colleagues imply (a) that teachers trained for a tracked school system differ markedly in their levels of professional knowledge and (b) that these differences are likely to cause considerable disturbance to a unitary school system. It is not hard to imagine a teacher whose content knowledge lags behind that of the highest-performing students in an untracked system, and who is likely to have difficulty responding intelligently to student errors on demanding tasks.

The arguments presented demonstrate that any serious debate on changes to the structure of the school system must take account of their potential side-effects and, at the same time, consider the findings of broad empirical studies.

4. Conclusions

Social disparities that have developed in the German education system – possibly as a result of the reintroduction of tracking in secondary education after World War II – have been systematically researched since the 1960s, but they have by no means been overcome. The arguments presented in this article make it clear that the PISA findings cannot be used as an instrument of causal analysis. Taking a wider perspective by drawing on a broader basis of empirical findings, it is clear that:

- there is no firm scientific basis for deciding in favor of or against a tracked school system,

- the PISA findings certainly do not allow any clear conclusions to be drawn,

- changing to a unitary or less strictly tracked system can be expected to have side-effects (e.g., in consequence of track-specific teacher training), and

- we would be well advised to take account of these potential side-effects when formulating recommendations for educational policy.

References

Alexander, K. L., & Entwisle, D. R. (1996). Schools and children at risk. In A. Booth & J. F. Dunn (Eds.). *Family-school links. How do they affect educational outcomes?* (pp. 67-88). Mahwah: Erlbaum.

Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse. Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Ed.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323-407). Opladen: Leske + Budrich.

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In Deutsches PISA-Konsortium (Ed.). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 261-331). Opladen: Leske + Budrich.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G., & Valtin, R. (Eds.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.

Dahrendorf, R. (1965a): *Arbeiterkinder an deutschen Universitäten*. Tübingen: Mohr Siebeck.

Dahrendorf, R. (1965b). *Gesellschaft und Demokratie in Deutschland*. München: Piper.

Deutsches PISA-Konsortium. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Deutsches PISA-Konsortium. (2002). *PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.

Deutsches PISA-Konsortium. (2004). *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.

Deutsches PISA-Konsortium (Ed.). (2005). *PISA 2003 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Münster: Waxmann.

Deutsches PISA-Konsortium (Ed.). (2007). *PISA 2006: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des dritten internationalen Vergleichs*. Münster: Waxmann.

Ditton, H. (1992). *Ungleichheit und Mobilität durch Bildung: Theorie und empirische Untersuchung über sozialräumliche Aspekte von Bildungsentscheidungen*. Weinheim: Juventa.

Ditton, H., & Krüsken, J. (2006). *Der Übergang von der Grundschule in die Sekundarstufe I. Zeitschrift für Erziehungswissenschaft,9,* 348-372.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, 57, 72–84.

Entwisle, D. R., & Alexander, K. L. (1994). Winter setback: School racial composition and learning to read. *American Sociological Review*, 59, 446-460.

Handl, J. (1985). Mehr Chancengleichheit im Bildungssystem. Erfolg der Bildungsreform oder statistisches Artefakt? *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 37*, 698-722.

Hanushek, E. A., & Wößmann, L. (2006) Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal, 116*, C63-C76.

Henz, U., & Maas, I. (1995). Chancengleichheit durch die Bildungsexpansion. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 47,* 605-633.

Köhler, H. (1992). *Bildungsbeteiligung und Sozialstruktur in der Bundesrepublik: Zu Stabilität und Wandel der Ungleichheit von Bildungschancen*. Berlin: Max-Planck-Institut für Bildungsforschung.

Köller, O. (2003). Gesamtschule - Erweiterung statt Alternative. In K. S. Cortina et al. (Eds.). *Das Bildungswesen in der Bundesrepublik Deutschland* (pp. 458-494). Reinbek: Rowohlt.

Köller, O., & Baumert, J. (2008). Entwicklung von Schulleistungen. In R. Oerter & L. Montada (Eds.). *Entwicklungspsychologie* (6th ed., pp. 735-768). Weinheim: Beltz/PVU.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (in press). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*.

Lehmann, R. H., Peek, R., & Gänsfuß, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen – Bericht über die Untersuchung im September 1996*. Hamburg: Behörde für Schule, Jugend und Berufsbildung, Amt für Schule.

Leschinsky, A., & Mayer, K. U. (Eds.). (1999). *The comprehensive school experiment revisited: Evidence from Western Europe*. Frankfurt a. M.: Lang

Maaz, K., Hausen, C., McElvany, N., & Baumert, J. (2006). Stichwort: Übergänge im Bildungssystem. Theoretische Konzepte und ihre Anwendung in der empirischen Forschung beim Übergang in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft,9,* 299-327.

Müller, W., & Haun, D. (1994). Bildungsungleichheit im sozialen Wandel. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 46,* 1-42.

OECD. (2007). *Bildung auf einen Blick 2007*. Paris: OECD.

Picht, G. (1964). *Die deutsche Bildungskatastrophe*. Olten: Walter.

Roeder, P. M. (1997). Entwicklung vor, während und nach der Grundschulzeit: Literaturüberblick über den Einfluss der Grundschulzeit auf die Entwicklung in der Sekundarstufe. In F. E. Weinert & A. Helmke (Eds.). *Entwicklung im Grundschulalter* (pp. 405-421). Weinheim: Beltz/PVU.

Rolff, H.-G. (1997). *Sozialisation und Auslese durch die Schule*. Weinheim: Juventa.

Schimpl-Neimanns, B. (2000). Soziale Herkunft und Bildungsbeteiligung. Empirische Analysen zu herkunftsspezifischen Bildungsungleichheiten zwischen 1950 und 1989. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 52,* 636-669.

Trautwein, U., Köller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis of 7th grade mathematics. *Contemporary Educational Psychology, 27,* 26-50.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective.* New York: Academic Press.

Wößmann, L. (2007). *Letzte Chance für gute Schulen. Die 12 großen Irrtümer und was wir wirklich ändern müssen*. München: Zabert-Sandmann.

Purpose and Limits of a National Monitoring of the Education System Through Indicators

Stefan C. Wolter

can be found in:

Wolter, S. C. (2007). *Purpose and Limits of a National Monitoring of the Education System Through Indicators.* In: Soguel, N. C., Jaccard, P. (Hrsg.). Governance and Performance of Education Systems. Dordrecht: Springer, 57-84.

and online under:

http://www.skbf-csre.ch/fileadmin/files/pdf/bildungsmonitoring/IDEHAP_Springer_4.2007.pdf

Educational Reporting in Germany
– Or: How Can Indicators Inform Innovation in the Education System? –

Hans Döbert

The currently prevailing theory of the "new output-driven model of governance" suggests that education systems can be "steered" or directed (Fend, 2006). To what extent this is really possible, the role played in the process by national educational reporting, and the scientific and theoretical demands to be made of indicator-based educational reporting, remain open questions from the theoretical, methodological, and empirical perspectives.

This article assumes familiarity with Germany's first national educational report. With reference to key conceptual aspects of educational reporting, it examines the theoretical basis for using "educational reporting" as an instrument of governance and its potential contribution to the development of the education system.

1. The Historical Development of Educational Reporting in Germany

There have been a number of attempts to introduce a comprehensive system of educational reporting to Germany. In 1975, the German Education Council published a report on developments in the education system (Deutscher Bildungsrat, 1975). In 1976, the Federal Ministry for Education and Science published an interim appraisal of education policy (Bundesministerium für Bildung und Wissenschaft, 1976) and, in 1978, the Federal Government brought out a report on the structural problems of the federal education system (Deutscher Bundestag, 1978). Both of these publications were one-offs, however.

Reports taking a more analytical approach have emerged from the research community. For instance, authors based at the Max Planck Institute for Human Development have published a report on Germany's education system at irregular intervals since 1979 (most recently: Cortina, Baumert, Leschinsy, & Mayer, 2003), the German Institute for International Educational Research published a book on the perspectives of the German education system in 1988 (Weishaupt, Weiß, Recum, & von Haug, 1988), and the Institute for School Development in Dortmund has published yearbooks of school development biennially since 1980, as well as compiling a statistical handbook on the German education system, a second, extended edition of which was published in 2001 (Böttcher, Klemm, & Rauschenbach, 2001). Further publications taking an analytic approach include Weißhuhn's (2001) review of education in Germany and the analyses on education and economic well-being carried out for the Federal Government's second report on poverty and wealth (Weißhuhn & Große Rövekamp 2004). In addition, there have been various historical analyses of educational developments in Germany since 1945 (see, for example, Führ, 1996, and Führ & Furck, 1998). The comparative international report on educational statistics published by the Bund-Länder Commission for Educational Planning and Research Promotion (BLK) in 2002 provides a relatively comprehensive and elaborate account of the assessment and reporting of statistical data in different coun-

tries, as well as proposals for its improvement. Finally, a recent publication by the Federal Statistical Office publication gives an overview of important subdomains of the German education system (Statistisches Bundesamt, 2003).

In addition, numerous other reports deal with subdomains or subaspects of the education system. And, of course, educational issues are also addressed in the context of general economic and social reporting – in the Federal Government's regular report on poverty and wealth *(Armuts- und Reichtumsbericht)* and in the Federal Statistical Office's data report *(Datenreport).*

Whereas comprehensive analyses of states and trends in other societal domains are published on a regular basis (e.g., the German Council of Economic Experts' annual report on economic development, the Federal Government's family and youth report, and the report on poverty and wealth), there was previously no counterpart providing a complete overview of the state of the German education system. International and supranational reporting systems did not suffice to fill this gap. [1]

A qualitatively new phase of educational reporting in Germany began in 2002/03, with the educational report commissioned by the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK) as well as the concepts for educational reporting that were developed for the Federal Ministry of Education and Research (BMBF) in the context of expert reports on non-formal and informal education in childhood and adolescence, on the one hand, and initial and continuing job training/lifelong learning, on the other (Avenarius et al., 2003; Baethge, Buss, & Lanfer, 2003; Rauschenbach et al., 2003). As a result, three conceptual approaches to educational reporting developed by three separate expert groups focusing on different domains of education were now available in Germany. Not only did these publications lay out important standards for educational reporting in Germany, they also identified salient desiderata.

In sum, national educational reporting has become established within just a few years in Germany. Moreover, increasing numbers of state-specific reports are being produced. Schleswig-Holstein, Bavaria, and Baden-Württemberg have already published reports, and reports for Berlin/Brandenburg, Hamburg, and Saxony are set to follow in 2008. Several municipalities now also produce regional educational reports (e.g., Offenbach, Munich, Dortmund, Tübingen). Further reports at state and regional level are in preparation, and work on the next two national reports is in progress (2008 and 2010).

2. Educational Reporting as Part of a Comprehensive System of Educational Monitoring

Whereas the main focus of educational governance was previously on the provision of resources ("input-driven model"), increasing attention is now being paid to measures

1 The OECD's efforts to facilitate international comparison of education systems through the continuous development of educational indicators provide the best known example of international educational reporting. The OECD publishes two annual reports: *Education at a Glance* and *Education Policy Analysis*. The European Commission publishes its *Key Data on Education in Europe* at regular intervals (last published 2005). For an account of educational reporting in other European countries, see Döbert, Hörner et al. (2004).

of quality assurance (process management) and to the systematic assessment of the products of education in terms of output (short-term measures of results) and outcomes (long-term measures of impacts).

As a logical consequence of this understanding of governance, there is a need for the provision or generation of policy-relevant knowledge that can inform and direct policy decisions. A widespread international approach is to obtain this policy-relevant knowledge by means of system monitoring. System monitoring has three main functions: first, to observe, analyze, and report on the key aspects of a system; second, to put in place measures of system control including benchmarking; third, to generate or refine policy-relevant knowledge that can inform, direct, and justify policy decisions. The implementation of benchmarking relative to other countries gives system monitoring a *comparative international component.*

Educational monitoring promotes transparency in education and thus provides a basis for discussions of educational objectives and for political decisions. The focus of educational monitoring is on the work of educational institutions ranging from daycare centers to continued professional training and development in adulthood.

In early 2006, the KMK adopted a comprehensive strategy for educational monitoring in Germany. Its major components are:

• participation in international assessments of student performance,

• central cross-state assessments to evaluate student performance in Germany by reference to educational standards (in the 4th, 9th, and 10th grades),

• comparative assessments of student performance by reference to educational standards to evaluate of the performance of individual schools on a national basis,

• joint educational reporting at the federal and state level.

Whereas all other components of the KMK strategy relate directly to the work of educational institutions and address their various stakeholders (teachers and students, parents and the community), educational reporting is concerned with the transparency of education from the system perspective. Its main outcomes are a regular educational report and a public website providing in-depth supplementary information. The core of any educational reporting service is a manageable and systematic set of indicators that can be regularly revised and updated.

The objective of routine educational reporting is to facilitate ongoing monitoring of the education system on the basis of reliable data that allow current states to be assessed from the system perspective and developments to be tracked over time and described empirically. Educational reporting has three essential characteristics:

- Educational reporting is based on an understanding of education that distinguishes three *goal dimensions of education:* “individual self-regulation,” “human resources,” and “social participation and equal opportunities.” These goal dimensions reflect more than just a traditional, individually oriented understanding of education. The primary aims of education are to advance individual growth, personality development, and the acquisi-

tion and development of culture. As such, it is generally considered from the individual perspective. The three goal dimensions addressed in educational reporting, in contrast, take a system perspective. This is consistent with the main purpose of educational reporting, which is to describe the societal, and especially institutional, conditions of education. The three goal dimensions represent distinct aspects of this system perspective.

- Under the guiding principle of *education across the life course,* educational reporting assesses the scope and quality of educational provision, as well as its uptake by individuals, across the various levels of the education system. The perspective of education across the life course is as yet rather limited, because the available data allow individual educational trajectories to be reconstructed only to a very limited extent, if at all.

- Educational reporting provides *indicator-based* coverage of all domains of the education system.

These essential characteristics are, at the same time, key criteria for the development, selection, and reporting of indicators.

Describing a complex system by abstraction to a small number of descriptors inevitably presents certain problems. Although this approach makes a complex situation more manageable, the resulting description of the system is necessarily less detailed and accurate. Nevertheless, given that the primary objectives of educational reporting are to provide education policy and administration with data that can inform policy decisions and activities, on the one hand, and to meet the scientific and public need for information in concentrated form, on the other, educational reporting based on quantitative indicators is, despite its inherent limitations, the best method for presenting systematic, verifiable, and well-founded data. Of course, this does not mean to deny that aspects that are not directly assessable and/or quantifiable are also important for the education system.

Context and input variables, educational processes, and the main products of these processes are all relevant to policy decisions. Selection of topics for analysis and of relevant data is thus based on a system that has been widely introduced for the purposes of international educational monitoring, namely the context-input-process-product evaluation model. This model currently seems the most useful heuristic for structuring information in the context of educational reporting (Scheerens, 2002).

In view of the purposes of educational reporting, the criteria outlined, and the heuristic model mentioned above, the following *topic areas* were identified as central to educational reporting in Germany in 2005 [2]:

– context level: demography,

– input level: expenditure on education, human resources, educational provision/educational institutions, educational participation/participants in education,

– process level: use of time in education, transitions, quality assurance/evaluation,

2 See also the indicator model presented by the consortium on educational reporting in March 2005, which presents and discusses the topics selected (Konsortium Bildungsberichterstattung, 2005, pp. 14-20).

– product level: competencies, qualifications, educational outcomes.

This list of eleven topics is intended to remain constant for the first years of educational reporting, but it can and must be modifiable on the long term. For example, additions are planned at the context level (e.g., development of an indicator of socio-economic background, inclusion of available economic and social indicators) and the process level (e.g., the quality of teaching and learning processes in the various domains of education).

3. The Understanding of the Goals of Education Underlying Educational Reporting

The German educational report is based on an understanding that distinguishes three goal dimensions of education: *individual self-regulation, human resources,* and *social participation and equal opportunities.*

Individual self-regulation reflects an individual's ability to independently plan and shape his or her behavior, interactions with the environment, biography, and life in the community. In the context of the knowledge society, one focus of this broad and general goal dimension for the education system as a whole as well as for each of its parts is on continuously developing the capacity to learn from early childhood on into old age. The report thus makes conscious reference to the concept of *Bildung*, which is difficult to translate into other languages; it goes beyond the acquisition of competencies and qualifications that can be exploited on the labor market to encompass the idea of self-development, including the acquisition and development of culture.

The education system's contribution to human resources lies – from the economic perspective – in securing and developing the workforce in quantitative and qualitative terms and – from the individual perspective – in helping people develop competencies that enable them to earn a living in accordance with their personal abilities, preferences, and attributes.

To the extent that educational institutions foster societal participation and equal opportunities, they facilitate social integration and the acquisition of culture for all and, in so doing, counteract systematic discrimination on account of social background, gender, nationality, or ethnicity. In this way, education contributes to social cohesion and demographic participation.

4. Educational Reporting in Other Countries

To ensure that educational reporting in Germany is aligned with international approaches, there is a need for systematic consideration of international developments in the field of educational monitoring. This implies in-depth analyses of international and supranational report systems and, at the level of individual indicators, consideration of the consequences of using the ISCED classification system to describe the national peculiarities of the German education system.

As points of orientation and systematic frames of reference, international reports – especially those published by the OECD, the European Commission, and UNESCO – set important standards for national educational reporting. In particular, the OECD's annual publications *Education at a Glance* and *Education Policy Analysis* and the European Commission's *Key Data on Education in Europe,* with their combination of indicators in time series and in-depth analyses of specific topics, provide a benchmark for national educational reporting.

The findings of analyses of the available national and international educational reports can be summarized as follows:[3]

- There is evidently broad international agreement on the need for data- or indicator-based educational reporting.

- However, there is a great deal of variation in the form and structure of national educational reports as well as in the institutions responsible for them. In principle, there are three types of educational reports: a) reports based primarily on educational statistics (Canada, France, Japan), b) reports based primarily on the results of inspections (England, the Netherlands, Sweden), and c) educational reports written by scientists based primarily on commented data and research findings (Germany, Switzerland, and – in part – the United States).

- The emphasis given to specific domains of the education system differs, but the focus tends to be on school education.

- All reports are based on the context-input-process-product evaluation model.

- Not all countries present a "complete overview" of policy-relevant information on the education system; some publish very detailed reports on specific domains or aspects of the system.

- Coverage of out-of-school education differs. There is a declared political will to take a whole-system or life-course perspective, despite the data problems involved.

- It is relatively widespread practice for commentaries of the core database to be supplemented by current research findings and the results of comparative international studies of student performance; the trend is clearly toward targeted cooperation between statistics and science.

5. Implementation of the "Education Across the Life Course" Perspective – Data Problems

Educational reporting in Germany is intended to take the perspective of "Education Across the Life Course" and to cover all educational domains – from preschool care and

3 The institutions involved in educational in Germany joined forces to analyze the educational reports of the following countries: Canada, Czech Republic, Denmark, England, France, Japan, the Netherlands, Russia, Spain, Sweden, Switzerland, the United States. The full results of these investigations and a comparative analysis will be presented in a separate, forthcoming study.

education to higher education and continued professional training and development – including non-formal and informal learning environments. The main focus is on whether and to what extent educational institutions succeed in preparing learners to meet the demands of training programs, work, family, and society. A particular focus is placed on points of transition in education, which generally determine the future course of "educational careers," their success or failure. Most transitions occurring during school careers are recorded in the official statistics. However, additional data are required, especially on school entry itself. For example, information is needed on the proportion of children entering elementary schooling who did or did not attend kindergarten/preschool, on students' socio-economic and ethnic background (to date, such data are only available, if at all, from surveys), and the available data must be better standardized. The KMK took an important step in this direction in 2003, when it specified a core dataset for individual data on schooling in the German states. Individual states can opt to extend this minimum dataset (especially with respect to instructional data). Ideally, it covers the following linkable data segments: organizational data on schools, individual data on students, individual data on school leavers and graduates, individual data on teachers, data on school classes/programs, organizational data on instructional units.

Particular efforts should be made to improve the following aspects of the database for use in educational reporting in the coming years:

- Switching from school-based statistics to individual data and creating data pools that allow individual educational transitions to be tracked. This restructuring measure is particularly important for the analysis of educational transitions, especially in specific subpopulations (e.g., grade repeaters) and for cross-regional analyses. From the perspective of educational reporting, it is imperative that individual data on the students of privately funded schools are also obtained and made available for analysis.

- Continued development of competency measures: competency measures obtained at school entry, in elementary school, at lower secondary level, at graduation from upper secondary level, and in adulthood would be ideal for the purposes of educational reporting.

- Introduction of a coherent database on the providers of further education, a domain in which the data situation is still far from adequate. At present, it looks likely that data on participation in further education will be assessed through a household survey in the context of the Adult Education Survey. A comprehensive assessment of the providers of further education is still required, however. Only this kind of database can provide information on the (regional) provision of further education, the human and financial resources of the institutions, and their development, and allow the economic significance of the domain to be evaluated.

6. Coverage of Non-Formal Education and Informal Learning

In the recent international discussion, increasing attention has been paid to forms of learning that take place outside formal learning environments and that play a major role in promoting individual (learning) competencies – the skills of self-regulation and self-

organization, in particular – as important individual prerequisites for lifelong learning. The categories of non-formal education and informal learning have become established for the description of these forms of learning; they reflect the increasing de-segmentation of learning processes and can often, but by no means always, be regarded as complementary to formal learning.

A particular empirical challenge currently facing research is to analyze the patterns of relationships between non-formal education and informal learning, on the one hand, and school-based competency acquisition, on the other, at the level of individual data. How are the individual levels of reading, science, and mathematics proficiency diagnosed in assessments of student performance associated with specific patterns of communication within the family, with voluntary and political engagement in non-formal contexts, with certain forms of self-directed learning in peer contexts, or with patterns of computer and internet use? How can these relations be extended to incorporate social-communicative and individual self-regulation skills? Which effects are attributable to specific forms of informal learning remains an open question, as does how formal and informal learning processes interrelate at the institutional and individual levels – whether their effects tend to be compensatory/substitutive or complementary. Realistically, research will have to hypothesize systemic relationships and examine whether these hold (Beathge & Baethge-Kinsky, 2004).

7. Conceptualizations and Roles of Indicators

There are various approaches to the conceptualization of indicators. One widespread approach is based on a rather narrow understanding of indicators, which regards constructs with a clearly defined measurement model as indicators. This is the understanding that underlies well-known publications such as the OECD's *Education at a Glance*. National and international educational reporting (Germany, Canada, United States), in contrast, is based on a broader conceptualization of indicators (e.g., Bottani & Tuijnman, 1994; Fitz-Gibbon, 1996; Fitz-Gibbon & Tymms, 2002), in which indicators are regarded as more complex constructs composed of various statistical parameters. Each approach has its advantages and disadvantages. Which is preferable in a specific situation depends on the purpose for which they are used, the governance intentions, the contextual conditions, the preferences of the commissioning body, etc.

Educational reporting in Germany and the indicator research accompanying it are based on the broader conceptualization, which allows less "central" indicators to be reported with high validity.

According to accepted definitions (e.g., Fitz-Gibbon, 1996; Oakes, 1989; Ogawa & Collom, 1998), *indicators* are quantitatively measurable variables that – used as proxies for complex, usually multidimensional structures – provide clear and coherent descriptions of the conditions and performance of a system, in its entirety or in part. They generally have a conceptual foundation, are highly differentiated, and draw on an empirically sound basis – generally a specific combination of statistical parameters. Beyond their conceptual basis, indicators are usually expected to be of relevance to actions and applications, in that they provide insights into existing or potential problems.

Indicators can be aggregated at different levels, to enable profiling of the system as a whole, of specific domains of education, educational institutions, and levels of instruction, or of individuals (each at the international level, the country level, or the regional level). Educational reporting is primarily concerned with the development of the education system and its individual domains as well as with processes of interaction between those domains. At the same time, insights are provided into institutional conditions and individual educational trajectories within individual domains.

Indicators are composed of one or more *statistical parameters*. The precise definition of these parameters involves a number of technical decisions: specification of the datasets and the variables used to determine them, the mathematical formulae used to calculate them, specification of the population to be covered or sample selection, and questions of statistical detail, such as how to deal with missing values.

If regular educational reporting is to be used in Germany to advance policy-related knowledge on the performance of the education system, there is a need not only to establish a permanent infrastructure for national reports, but also for targeted research and development, as well as for scientific answers to fundamental questions. These relate primarily to the conceptual foundation of the indicators, empirical evidence for hypothesized patterns of effects, and interdependencies between indicators.

8. Non-Indicator-Based Policy-Related Information – The Concept of Thematic Focus

Every educational report focuses on an issue that is of particular relevance to policy-related decisions, but that can generally not yet be described by indicators, and describes this issue in depth in a separate chapter. Unlike the other sections of an educational report, this chapter does not have to be indicator-based, but may include survey data, findings from scientific studies drawing on other databases, descriptions of measures and methodological steps, etc. It can thus take a problem-oriented and analytic in-depth approach, examining key areas of development in the education system and discussing options for quality-oriented development.

Based on the fundamental idea that integration through education and integration in the education system are two closely related issues in childhood and adolescence, the 2006 educational report in Germany (Bundesministerium für Bildung und Forschung, 2006) had a special focus on “Education and Immigration.” Based on an international definition of labor market integration (OECD, 2005), the goal of integration in the education system can be understood to the effect that, over the course of time, children and adolescents from minority families succeed in attaining knowledge, skills, and qualifications comparable with those of their majority peers. Although there have been marked improvements in the educational and occupational achievements of students from minority families in the last 30 years, there is still a significant gap in the competencies acquired by minority and majority students. PISA, PIRLS, and other studies have compared the two groups and found considerable disparities in their cognitive abilities, their teachers’ transfer recommendations, and the secondary school type they attend that can be explained partly – but by no means fully – by the lower average socio-economic resources of minority families. Furthermore, a good command of the German language is a prerequisite for equal par-

ticipation in the German education system. The educational report 2006 elucidates these issues by addressing them from four main perspectives: the macro-societal perspective (size and structure of the immigrant population), the individual perspective (educational participation, educational trajectories and careers, and background characteristics of children, adolescents, and adults from minority backgrounds), the institutional perspective (the education system's approach to immigration), and the comparative international perspective (differences in competency acquisition in different minority populations and contrasting institutional approaches to the phenomenon of immigration).

The next report (2008) will focus on transitions from school to vocational and higher education and the labor market.

9. Limitations of Educational Reporting

The analytic potential of educational reporting derives principally from connectivity in terms of the linking of statistical data. In the various national and international educational reports, this is achieved by linking basic data, background variables, and reference data in various ways. From report to report, each individual indicator can be set in relation to different reference data and analyzed in terms of different background variables. All these derived parameters and comparative values can be reported across time, over a period of years. These differentiations within individual indicators provide a starting point for interpretation, analysis, and finally political evaluation. It is thus possible to modularize the indicator system and the entire educational reporting system, and to select from and report on a set of mutually independent units, depending on the data available and current political requirements.

Education at a Glance and other educational reports capitalize on these opportunities to differentiate within indicators in various ways. Different indicators are rarely combined, however. Likewise, statistical parameters that quantify relationships (e.g., correlation coefficients or multivariate analyses) are rarely used. The only "interactions" considered are those between the basis data of an indicator and various background variables; they are documented in purely descriptive terms. Indicator-based approaches can rarely provide answers to questions of cause and effect. Likewise, an indicator-based report cannot take adequate account of ongoing developments, for which representative data are not (yet) available.

Unlike scientific studies, the purpose of educational reporting is descriptive and evaluative. Educational reporting is "analytic" in the sense that it offers great scope for comparison and evaluation, but it is *not causal analytic*. Educational reporting therefore has an important role to play in scientifically founded educational monitoring, but it cannot – and is not intended to – answer all policy-related questions.

References

Ackeren, I. van, & Hovestadt, G. (2003). *Indikatorisierung der Empfehlungen des Forum Bildung. Reihe Bildungsreform. Vol. 4.* Berlin: Bundesministerium für Bildung und Forschung.

Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., & Rürup, M. (2003). *Bildungsbericht für Deutschland. Erste Befunde.* Opladen: Leske + Budrich.

Baethge, M., & Baethge-Kinsky, V. (2004). *Der ungleiche Kampf um das lebenslange Lernen.* Münster: Waxmann.

Baethge, M., Buss, K.-P., & Lanfer, C. (2003) *Konzeptionelle Grundlagen für einen Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen. Reihe Bildungsreform*. Vol. 7. Bonn: Bundesministerium für Bildung und Forschung.

Bartelheimer, P. (2006). Wozu Sozioökonomische Berichterstattung? – eine Kontroverse. *SOFI-Mitteilungen, 34,* 67-90.

Bartelheimer, P., Boes, A., Fuchs, T., Grimm, N., Hacket, A., Land, R., Mayer-Ahuja, N., & Weber, C. (2006) *Berichterstattung zur sozioökonomischen Entwicklung Deutschlands. Zweiter Bericht, Zwischenbericht, Teil I, Auswertung der Werkstattgespräche zur sozioökonomischen Berichterstattung im ersten Halbjahr 2006.* Göttingen.

Bos, W., Holtappels, H.-G., Pfeiffer, H., Rolff, H.-G., & Schulz-Zander, R. (Eds.). (2006). *Jahrbuch der Schulentwicklung. Vol. 14. Daten, Beispiele und Perspektiven.* Weinheim: Juventa.

Böttcher, W., Klemm, K., & Rauschenbach, T. (Eds.). (2001). *Bildung und Soziales in Zahlen. Statistisches Handbuch zu Daten und Zahlen im Bildungsbereich.* Munich: Juventa.

Bottani, N., & Tuijnman, A. (1994). International educational indicators: Framework, development and interpretation. In Centre for Educational Research an Innovation (Ed.). *Making Education Count. Developing and Using International Indicators.* Paris.

Bundesministerium für Arbeit und Soziales. (BMAS). (2005). *Lebenslagen in Deutschland. 2. Armut- und Reichtumsbericht der Bundesregierung.* Bonn: BMAS.

Bundesministerium für Bildung und Forschung. (BMBF). (2006). *Berichtssystem Weiterbildung IX. Integrierter Gesamtbericht zur Weiterbildungssituation in Deutschland.* Bonn/ Berlin: BMBF.

Bundesministerium für Bildung und Forschung (BMBF). (2007). *Berufsbildungsbericht 2007.* Berlin: BMBF.

Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ). (2005a). *Bericht über die Lebenssituation junger Menschen und die Leistungen der Kinder- und Jugendhilfe in Deutschland. Zwölfter Kinder- und Jugendbericht. Bildung, Betreuung und Erziehung vor und neben der Schule.* Berlin: BMFSFJ.

Bundesministerium für Familie, Senioren, Frauen und Jugend. (BMFSFJ). (2005b). *Familie zwischen Flexibilität und Verlässlichkeit. Perspektiven für eine lebenslaufbezogene Familienpolitik. Siebter Familienbericht.* Berlin: BMFSFJ.

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (BLK). (Ed.). (2002). Vergleichende internationale Bildungsstatistik. Sachstand und Vorschläge zur Verbesserung. In *Materialien zur Bildungsplanung und Forschungsförderung, 103.*

Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (BLK). (2006). *Heft 137-I. BLK-Bildungsfinanzbericht 2004/2005. Die aktuelle Entwicklung. Materialien zur Bildungsplanung und zur Forschungsförderung.* Bonn: BLK.

Cortina, K. S., Baumert, J., Leschinsky, A., & Mayer, K.-U. (Eds.). (2003). *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick.* Reinbek: Rowohlt.

Deutscher Bildungsrat. (Ed.). (1975). *Bericht '75. Entwicklungen im Bildungswesen.* Stuttgart: Klett.

Deutscher Bundestag. (Ed.). (1978). *Bericht der Bundesregierung über die strukturellen Probleme des föderativen Bildungssystems. Bundestagsdrucksache 8/1551.* Bonn.

Deutsches Jugendinstitut. (Ed.). (2003). *Zahlenspiegel. Daten zu Tageseinrichtungen für Kinder.* Munich: DJI.

Die internationalen Bildungsindikatoren der OECD - ein Analyserahmen. (1992) Ein OECD/CERI-Bericht. Frankfurt am Main.

Egeln, J., Eckert, T., Heine, C., et al. (2003). *Indikatoren zur Ausbildung im Hochschulbereich. Studie zum Innovationssystem Deutschlands Nr. 10-2003. ZEW Dokumentation 03-03.*

European Commission. (Ed.). (2007). *Key data on higher education in Europe – 2007 Edition.* Brussels: Eurydice.

European Commission. (Ed.). (2000). *Memorandum über lebenslanges Lernen.* Luxemburg: Office for Official Publications of the European Communities.

Fitz-Gibbon, C. (1996a). *Monitoring education. Indicators, quality and effectiveness.* London: Cassell.

Fitz-Gibbon, C. (1996b). Monitoring school effectiveness: Simplicity and complexity. In J. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson (Eds.) *Merging traditions: The future of research on school effectiveness and school improvement.* London: Cassell.

Führ, C. (1996). *Deutsches Bildungswesen seit 1945.* Neuwied & Kriftel: Luchterhand.

Führ, C., & Furck, C.-L. (1998). *Handbuch der deutschen Bildungsgeschichte, Vols. VI/1 and VI/2.* Munich: Beck.

Holtappels, H. G., Klieme, E., Rauschenbach, T., & Stecher, L. (2007). *Ganztagsschule in Deutschland. Ergebnisse der Ausgangserhebung der „Studie zur Entwicklung von Ganztagsschulen".* Weinheim: Juventa.

KMK. (2002). *Wichtige Beschlüsse der Kultusministerkonferenz „Qualitätssicherung in Schulen im Rahmen von nationalen und internationalen Leistungsvergleichen – Entwicklung Bildungsstandards? Beschluss vom 17./18.10.2002.*

Konsortium Bildungsberichterstattung. (2005). *Gesamtkonzeption der Bildungsberichterstattung.* Retrieved November 27, 2007, from http://bildungsbericht.de/daten/gesamtkonzeption.pdf

Konsortium Bildungsberichterstattung. (Ed.). (2006). *Bildung in Deutschland. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration.* Bielefeld: Bertelsmann and www.bildungsbericht.de

Noll, H.-H. (1998). Sozialberichterstattung: Funktionen, Merkmale und Grenzen. In Institut für Sozialwissenschaftliche Forschung (IFS) et al. (Eds.). *Jahrbuch Sozialwissenschaftliche Technikberichterstattung, Sonderband Beobachtungsfeld Arbeit* (pp. 25-35).

Noll, H.-H. (2002). Social indicators and quality of life research: Background, achievements and current trends. In N. Genov (Ed.). *Advances in sociological knowledge over half a century.* Paris: International Social Science Council.

Noll, H.-H. (2003). Sozialberichterstattung, amtliche Statistik und die Beobachtung sozialstaatlicher Entwicklungen. In Meulemann, H. (Ed.). *Der Sozialstaat in der amtlichen Statistik. Angebote und Nachfragen. Sozialwissenschaftliche Tagungsberichte Vol. 6.* IZ: Bonn, 65-84.

Oakes, J. (1986). *Educational indicators: A guide for policymakers.* Santa Monica, CA: RAND.

Oakes, J. (1989). What educational indicators? The case for assessing school context. *Educational Evaluation and Policy Analysis, 1*(2), 181-199.

OECD. (2006). *Education policy analysis. Focus on higher education. 2005-2006 edition.* Paris: OECD.

OECD. (2007). *Education at a Glance 2007. OECD Indicators.* Paris: OECD.

Ogawa, R., & Collom, E. (1998). *Educational indicators: What are they?* Riverside, CA: ERIC.

Rauschenbach, T., Leu, H.-R., Lingenauber, S., Mack, W., Schilling, M., Schneider, K., & Züchner, I. (2004). *Konzeptionelle Grundlagen für einen Nationalen Bildungsbericht – Non-formale und informelle Bildung im Kindes- und Jugendalter. Reihe Bildungsreform. Vol. 6.* Bonn: Bundesministerium für Bildung und Forschung.

Scheerens, J. (1992). Prozessindikatoren der Arbeitsweise von Schulen. In *Die internationalen Bildungsindikatoren der OECD – ein Analyserahmen. (1992) Ein OECD/CERI-Bericht.* Frankfurt am Main u. a., 62 -91. (Translation of: OECD/CERI (Ed.) (1991). *The OECD International Education Indicators. A Framework for Analysis.* Paris.).

Scheerens, J. (2004). *The conceptual basis of indicator systems.* Frankfurt: Lang.

Statistisches Bundesamt. (2003). *Im Blickpunkt: Bildung in Deutschland.* Wiesbaden: Statistisches Bundesamt.

Tietze, W. (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten.* Neuwied: Luchterhand.

Weishaupt, H., Weiss, M., Recum, H. von, & Haug, R. (1988). *Perspektiven des Bildungswesens der Bundesrepublik Deutschland. Rahmenbedingungen, Problemlagen, Lösungsstrategien.* Baden-Baden: Nomos.

Weißhuhn, G. (2001). *Gutachten zur Bildung in Deutschland.* Bonn: Bundesministerium für Bildung und Forschung.

Wissenschaftsrat (2002). *Eckdaten und Kennzahlen zur Lage der Hochschulen von 1980 bis 2000.* Cologne: Wissenschaftsrat.

The National Educational Report – A Step in the Direction of Evidence-Based Policy in Austria

Werner Specht

1. "Evidence-Based Policy" and National Educational Reports

In many countries today, increased efforts are being undertaken to create and implement evidence-based educational policy. This entails basing educational policy decisions more strongly than before on scientific research findings about the strengths and weaknesses of the education system in question. At the same time, it means that science and research contribute more to creating and analyzing the knowledge base for system governance. Both the OECD-CERI[1] and the European Union[2] have recently launched programs and events designed to improve the conditions for evidence-based educational policy in the member countries (Burns, 2007; Commission of the European Communities, 2007).

The following have been identified as crucial instruments for improving the knowledge-based governance of education systems:

- providing increased public funding to science and research in the field of education,
- establishing well-endowed research programs in developing fields of central importance for educational policy,
- strengthening the institutional framework for exchange between educational policy-making bodies and organizations in research and development, and
- setting up what are known as *brokerage agencies*. Their task consists, on the one hand, in compiling and presenting the scientific knowledge available on specific issues in a form that can enable informed educational policy decision-making. On the other, it consists in translating the research needs of educational decision-makers into viable research questions and thus in stimulating research

Educational reports are another important instrument of scientifically based educational policy making. Containing data and research findings on education, these reports provide governance knowledge to decision-makers at the levels of educational policy design and educational administration.

The production of reports on the situation of national education systems has been institutionalized in many western countries[3], but in the German-speaking countries, in particular, this instrument of policy support is still very young. The first educational report in Germany, published in 2003, was commissioned by the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK) and written by a consortium of authors (Avenarius et al., 2003). In 2006, a second volume followed, this time commissioned jointly by the federal government and the states (Avenarius et al., 2006).

1 http://www.oecd.org/document/29/0,3343,en_2649_37455_31237469_1_1_1_37455,00.html
2 http://interkoop.dipf.de/index.php?option=com_content&task=view&id=48&Itemid=64
3 For an overview, see Rürup (2003).

In Switzerland, the first national educational report was commissioned by the Swiss Conference of Cantonal Educational Directors (EDK), and compiled under the direction of the Swiss Coordination Center for Research in Education (Bildungsbericht Schweiz, 2006).

In Austria, planning papers for a national educational report have existed since 2003, but these did not initially meet with the support of political decision-makers. Since early 2007, concrete planning has been underway for the 2008 Austrian National Educational Report.

2. Report Form: Descriptive – Analytical – Normative?

2.1 International Variations

The educational reports published in the developed countries differ widely with regard to the type of text, authorship, and target audience:

- *The text types* range from purely descriptive texts presenting data on the education system, to analytical texts developing, describing, and interpreting quality indicators, to purely qualitative status reports or problem reports. The Swiss and German educational reports are essentially presentations and interpretations of carefully selected indicators that can be described – in some cases implicitly, in some cases explicitly – as quality features. The presentations themselves are of a mainly descriptive nature, but also contain evaluative comments on the findings derived.[4]

An entirely different kind of educational report is the report by the independent Dutch Inspectorate of Education, which presents the school inspectors' aggregated findings from the reporting period almost purely in the form of qualitative appraisals with suggestions for improving the situation (Inspectie van het Onderwijs, 2005).

- The *authors of the reports* may include: (a) the educational administrations themselves, whose statistical units compile the data (as is the case in France); (b) research institutes with close relationships to the school administrations, which political decision-makers commission to prepare educational reports (as is the case in almost all the German states); (c) independent institutions that prepare the reports on their own initiative or are commissioned to do so by the government (for example, the reports by the independent inspectorate of education in the Netherlands); (d) consortiums of researchers that work together to prepare reports commissioned by public authorities (the German educational report being a typical example). Probably the most important aspect of authorship is the authors' autonomy from, or dependence on, the public bodies that commissioned them, and the extent to which the contents of the report are subject to the commissioning bodies' influence and/or control. It seems desirable that the main questions to be answered in the report be formulated on the political and administrative level, and that the concrete reporting adheres exclu-

4 One example is the section of the German report dealing with the issue of repeating a school year (Indicator D2, pp. 54ff.), where the high average school career losses of students are assessed extremely critically.

sively to criteria of scientific relevance and quality.

- The *target readers* of the reports are, on the one hand, governments and administrative bodies, and on the other, the public at large. An educational report is thus equally a document that guarantees accountability to the public and that provides information to educational administrations. Differences exist in the extent to which countries' parliaments are formally required to consider and act on the findings of the educational reports.

2.2 The Conception for the Austrian Report

In Austria, a specific and relatively strong culture of scientific policy advice has emerged over the last few decades. University researchers have been commissioned to produce expert reports on central questions of the development, governance, and management of the school system. The reports have taken on programmatic status and have exercised a relatively strong influence on the discussion of educational policy and the decisions resulting from it.[5]

With this tradition, it was only logical that Austrian educational reports would not limit themselves to presenting data and indicators on the current state of the system, but would instead use these findings to take positions on important questions of how to develop and improve the school system, and recommend strategies. The concrete considerations on the structure of the report are outlined below (see Section 4.3).

3. Principles of Cooperation between Educational Policy, Government Administrations, and Scientific Research

A data- and output-oriented educational steering model relies completely on productive cooperation between policy-makers and public administrations, on the one hand, and the representatives and institutions of the scientific community on the other. It is thus crucial that both sides be brought together to work efficiently on developing and improving the education system.

This requires, however, that major prerequisites be met on both sides:

- *Prerequisites on the policy side:* Educational reports are highly sensitive matters because of the inherent tension between an administration's political interest in presenting itself in the best possible light and the actual intention of such reports, which is to present uncompromising evaluations from which priorities can be derived for educational planning. This tension gives rise to conflicting ideas about the distribution of roles between policy-makers and government administrations, on the one hand, and educational research bodies, on the other. Many politicians would prefer that these reports were entirely under their own control, that they were written by civil servants, and that the scientific community were solely responsible for data production. This,

5 Particularly important examples were Posch & Altrichter (1993, 1997), Eder (2002), and Haider et al. (2003, 2005).

however would severely reduce the informational function of educational reports.

A precondition for such reports to fulfill their function of providing *relevant* scientific findings for educational policy governance and management is thus the willingness of policy-makers to allow uncensored and unfiltered diagnostic findings to be published, and to enable the researchers involved to work largely autonomously.

- *Preconditions on the research side:* On the opposite side, the researchers involved are obligated to represent the reality of the education systems they study in a balanced and fair way, using only highly reliable data, and refraining from open critique of the agency commissioning the study. The latter point dos not mean that data and findings are not to be interpreted. It cannot be the sense of scientific reports, however, to intervene directly in the political process.

  The scientific probity and quality of the educational report depend above all on whether and to what extent conventional mechanisms of scientific quality control (colleagues' observations and critique, peer review procedures) actually function in a particular country and are applied in the specific case of educational reporting.

As emphasized in a previous paper (Specht, 2007), the kind of clear division of roles between policy and research that was formally developed by the Swiss EDK and that forms the basis for the Swiss educational report (Bildungsbericht Schweiz, 2006, p. 7) is a crucial precondition for success:

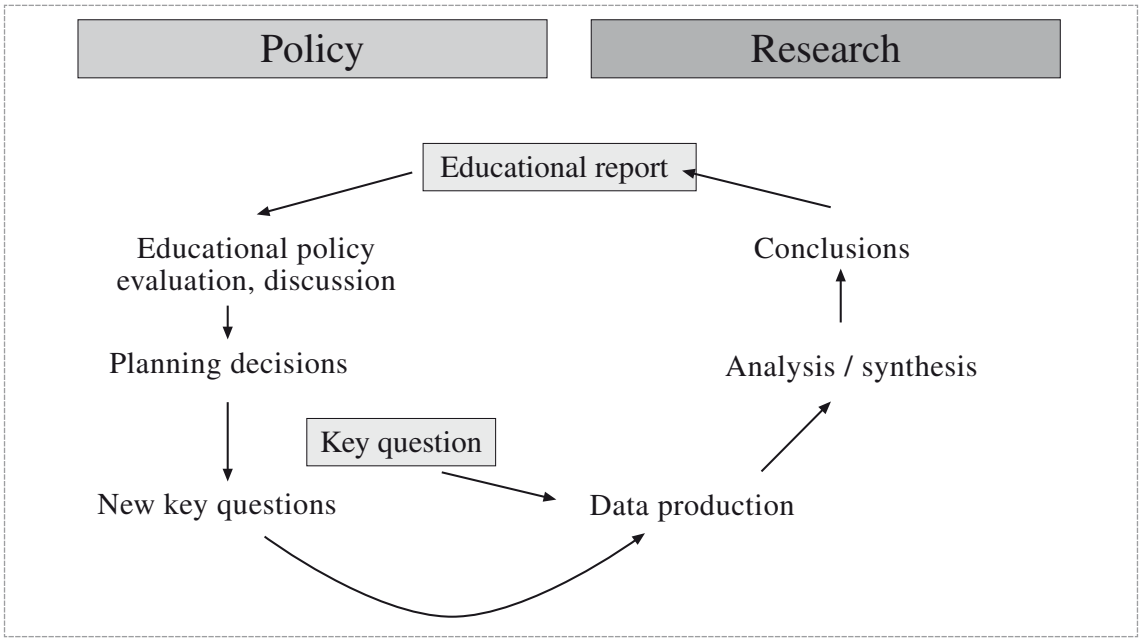

Fig. 1: Division of responsibilities and interaction between policy and research in educational reporting according to the Swiss Model

This concept contains a very clear division between the authority and responsibilities delegated to policy and research in educational reporting: the political decision-makers formulate the key questions to be analyzed and are solely responsible for the concrete planning decisions that will be made based on the reports. A consortium of experts produces and interprets data but also draws scientific conclusions from these findings and contributes these to the debate independently and on their own authority.

The model seems convincing because it assigns planning and decision-making authority to the responsible agencies of the democratic state but, at the same time, places these agencies under pressure to justify policy decisions based on the expertise of independent researchers.

This concept is also the model for the Austrian national educational report.

4. Austrian Educational Report 2008 – A Pilot Project

4.1 Functions

With its first national educational report, Austria aims to join in the international movement toward evidence-based educational policy. The report is designed to present the most comprehensive possible picture of the Austrian school system – from pre-school to general and vocational secondary schools – and to use this as a basis for an educational policy aimed at improving the quality of the education system.

The educational report slated to appear in 2008 will be a *pilot version* designed to fulfill the following functions in particular:

- *Developing quality indicators:* In the periodic reporting context, a first attempt will be made to develop a set of quality indicators for the Austrian school system and to provide figures on these indicators as far as possible using the available data. The indicators developed should be capable of linking to the indicator systems of the OECD and European Union and thus enable international comparison. They should also reflect the specifics of the Austrian situation that can be observed longitudinally in future reports.

- *Testing the indicators for their capacity to achieve educational policy consensus:* An important goal for indicators is to test them for their capacity to generate consensus at the national level and, in the long term, to achieve agreement on the most important goals for the education system that are reflected in those indicators. It is crucially important that quality in the education system not be defined exclusively by experts but that it attains its legitimacy through broad social discourse.

- *Ensuring the practicability and acceptance of the overall concept among educational policy-makers and the general public:* The pilot version will develop an overall concept for educational reporting. Whether this concept is meaningful and practicable for periodic national educational reporting should be discussed and assessed by the public and policy-makers; if necessary, modifications should be made. The pilot version is conceived explicitly as a learning opportunity to address the following questions:

  (a) organizational and resource issues such as costs, schedules, and forms of cooperation,
  (b) data availability and needs for improvement in the data situation,
  (c) how the report is to be divided into descriptive, analytical, and normative components,
  (d) periodicity of publishing future reports, and
  (e) how the results are to be used by policy authorities and government administrations.

4.2 Organizational Concept

In early 2007, the newly created Federal Institute for Educational Research, Innovation and Development of the Education system (bifie) was tasked with developing a concept for an educational report.

The federal institute worked together with the leading education expert in the country to develop the report concept. It used this opportunity to affirm and test its designated function as an interface and intermediary agency between the educational authorities and administrations, on the one hand, and the representatives and institutions of the research community, on the other.[6]

4.3 Concept for the Report's Content

The current concept for the report's content – the basic structure of which was developed and decided upon within the group of editors and has now also been approved by senior administrative officials – conceives of a basic binary division within the report:

Part 1: The Austrian school system in the light of data and indicators

The first section describes the Austrian school system as comprehensively as possible using data and indicators on the main aims and quality aspects of the school system.

The indicator section will present relevant, currently available data in an easily visualized form. It foresees developing those indicators that would be useful for longitudinal analysis (repeated measurements) in future educational reports.

Another important goal of this section is to reveal data gaps and needs for additional data and to make proposals for how data availability can be improved in future reports.

This section of the report adheres fairly closely to the educational reports from the German-speaking countries that served as its models, and are aimed primarily at representing and interpreting indicators on the education system.

Part 2: Focal points for quality development in the school system

The second section is more problem-oriented, analytical, and explicitly normative in nature. It addresses the main areas of the school system in which problems exist and improvements are needed, analyzes them based on the available data and research findings, and discusses options for improving these areas with the aim of quality development.

The main focal themes were identified by reference to different perspectives:

- The first group of themes chosen are *current focuses of educational policy discussion* in

6 Those involved in planning the report's content included: Prof. Dr. Herbert Altrichter (University of Linz), Prof. Dr. Ferdinand Eder (University of Salzburg), Dr. Günter Haider (University of Salzburg), Dr. Lorenz Lassnigg (IHS), Mag. Josef Lucyshyn (bifie), Prof. Dr. Georg Neuweg (University of Linz), Prof. Dr. Werner Specht (bifie), Prof. Dr. Christiane Spiel (University of Vienna).

Austria due to the strong pressure for change being felt in these areas. These themes include among others pre-school education and the transition from kindergarten to school, particularly the promotion of early linguistic development; the pressures resulting from problems at lower secondary level and their potential solutions; and the need to further improve the quality of special educational support for children with disabilities or impairments, particularly in integrative contexts.

- The second group of themes chosen as focal points focus on more narrowly *pedagogical questions* that appear to be of *high priority* in the current situation of increasing heterogeneity among students. They address, for example, possibilities for the provision of individualized learning opportunities and the framework conditions and teaching skills this requires, or the increased use and broader distribution of alternative, motivating, and non-selection-oriented forms of feedback on students' educational outcomes. In these thematic areas, the report does not attempt to enter the debate on fundamental pedagogical principles, but rather – in line with its policy-oriented character – addresses the conditions that either foster or impede more efficient teaching.

- The third group of themes relates to issues of *school system governance and management* that are of immediate educational policy importance, and where a significant reform deadlock can be identified. One is the question of quality control and quality improvement, which has given rise to wealth of expert reports and position papers in the last decade and a half, without any fundamental decisions being made on the political level. Another theme is the discussion of decision-making structures within the school system that has been ongoing since the beginning of the 1990s under the heading "school autonomy." In recent years, decision-making authority has been delegated, in numerous cases relatively arbitrarily, to individual schools or regions, without previous careful consideration of the best decision-making level for each particular question or problem.

- Fourth, this analytical section will address two themes that are particularly important in the context of efforts towards evidence-based governance of the school system. The first deals with the situation of educational research in Austria as the most important resource for better understanding the current situation and the connections between governance decisions (input), processes, and quality of results. The focus here is on how to stimulate educational research to make it more effective, while at the same time enabling it to be used more efficiently.

  Educational economics is a special case of educational research; it is a discipline lacking tradition or institutional foundations in Austria. The issues of funding and the possibilities for using educational economic research to create an informed system governance process will be the focus of a separate chapter.

This second section of the report diverges in its conception from the currently existing educational reports in Switzerland and Germany (including the German states), but it follows on Austrian traditions, as mentioned above, in basing key content areas for educational policy development on scientific analysis and expert studies. The chapters of the thematic section will also make very strong use of the empirical data. Starting from a comprehensive analysis of the education system in its current state, this section will

identify and describe options for making changes to improve each individual sphere of activity.

In *both* sections, the attempt will be made to establish links to international indicator systems (OECD, EU) and priority policy areas (e.g., the Lisbon goals of the European Union).

4.4 Quality Control

In order to ensure that the report is readable despite the heterogeneity of its content and its diverse authorship, and to keep the individual articles comparable in their structure, different forms of quality control will be used. One will be to provide external guidelines for text type and structure to ensure that the texts submitted are indeed comparable. Another will be to put all contributions through a two-stage review process. Here, in conclusion, we describe some of the guidelines for the texts.

Text type

The text type chosen for contributions to the report should be oriented toward the primary target readership. The report is aimed, on the one hand, at Austrian educational administrations and policy-makers. On the other hand, it serves to provide information on educational issues to the public at large. This leads to high demands regarding both text structure and the figures used to visualize results:

- On the one hand, the text should meet high scientific standards and represent the current state of scientific research on the issue.
- On the other hand, the text should be understandable for an educated layperson. It should not be overloaded with quotes or literature citations, but should be readable and written in a brisk, accessible style.

Policy analysis

Each thematic chapter contains two basic sections:

- The first section provides a *scientific* analysis of the issue at hand. It explains the relevance of the particular theme for the quality of the school system, and discusses findings from theory, research, and evaluation that characterize the current situation and its problems.
- The second section discusses current tendencies in *educational policy* for addressing the particular issue and identifies possibilities and options that can contribute to improving the situation. In the latter context, it is important not to take a purely pedagogical/psychological approach to the question, but to place the focus on how *political governance* can optimize the situation. This is perhaps an unaccustomed perspective for psy-

chologists or educational scientists. It requires not least of all broader ideas about how educational policy measures can have a positive impact on learning outcomes through the process of their implementation.

The conclusions drawn in the education policy section of the expert reports are to be fairly reserved in nature, and should not appear as lists of demands. They should take the form of strategy recommendations rather than detailed enumerations of actual measures.

Structure of the thematic reports

The reports should be structured roughly as follows:

1. *Problem analysis:* educational policy relevance of the thematic area

- What makes the theme relevant for the quality of the education system?
- Recent developments and/or findings that create pressure for change or adaptation.
- Treatment of the theme in the current political context: (a) priorities, (b) initiatives, (c) standpoints on the theme in a national framework.

2. *Situational analysis:* data, indicators, results of research and evaluation

- Connection back to the indicator section of the report (as far as possible): focus on data and indicators.
- Overview of the research and evaluation results on the theme and specific aspects thereof – international and national: established knowledge and disputed issues.
- Summary of the current state of scientific findings.

3. *Research questions*

- Knowledge gaps and needs for research.
- Suggestions for educational research on the theme in question.
- Suggestions for inclusion of indicators in the quantitative section.

4. *Political analysis* and options for development

- Different approaches and experiences in the international context: How is the theme seen and dealt with in other education systems?
- Possible conclusions to be drawn from research results, evaluations, and implementation studies.
- Alternative reform scenarios.
- Potential first/next steps to improve the situation.

4.5 Outlook

The overall concept of the Austrian educational report was presented in summer 2007 to the federal minister and senior officials of the ministry, who then allocated the necessary funds. Since this point in time, the responsibility for the quality of the report has lain entirely with the institutions carrying it out, the consortium of scientific advisors, and ultimately the authors themselves.

The report is slated for presentation to senior department officials at the end of 2008.

References

Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., & Rürup, M. (2003). *Bildungsbericht für Deutschland: Erste Befunde.* Opladen: Leske + Budrich.

Avenarius, H. et al. (2006). *Bildung in Deutschland: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration.* Bielefeld: Bertelsmann.

Bildungsbericht Schweiz 2006. (2006). Aarau: Schweizerische Koordinationsstelle für Bildungsforschung.

Burns, T. (2007). *Evidence in education: Linking research and policy. Knowledge management.* Paris: OECD.

Commission of the European Communities. (2007). *Towards more knowledge-based policy and practice in education and training.* Brussels: European Union (Commission Staff Working Document).

Deutsches Institut für Internationale Pädagogische Forschung (DIPF). (2007). *Wissen für Handeln – Forschungsstrategien für eine evidenzbasierte Bildungspolitik: EU-Konferenz im Rahmen der Deutschen Präsidentschaft,* Deutsches Institut für Internationale Pädagogische Forschung (DIPF). Retrieved March 6, 2008, from http://interkoop.dipf.de/index.php?option=com_content&task=view&id=48&Itemid=64

Eder, F. (Ed.). (2002). *Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen.* Bildungsforschung des Bundesministeriums für Bildung, Wissenschaft und Kultur, 17. Innsbruck et al.: StudienVerlag.

Haider, G., Eder, F., Specht, W., & Spiel, C. (2003). *Zukunft: Schule: Strategien und Maßnahmen zur Qualitätsentwicklung. Reformkonzept der österreichischen Zukunftskommission.* Vienna: Bundesministerium für Bildung, Wissenschaft und Kultur (Initial report of October 17, 2003).

Haider, G., Eder, F., Specht, W., Spiel, C., & Wimmer M. (2005). *Zukunft: Schule: Strategien und Maßnahmen zur Qualitätsentwicklung. Reformkonzept der österreichischen Zukunftskommission.* Vienna: Bundesministerium für Bildung, Wissenschaft und Kultur (Final report of April 6, 2005).

Inspectie van het Onderwijs. (2005). *The state of education in the Netherlands 2004/2005.* Utrecht: Education Inspectorate.

OECD. *Centre for Educational Research and Innovation (CERI) – Evidence-based Policy Research in Education.* Retrieved April 5, 2008, from http://www.oecd.org/document/29/0,3343,en_2649_37455_31237469_1_1_1_37455,00.html.

Posch, P., & Altrichter, H. (1993). *Schulautonomie in Österreich* (2nd updated and expanded ed.). Vienna: Bundesministerium für Unterricht und Kunst.

Posch, P., & Altrichter, H. (1997). *Möglichkeiten und Grenzen der Qualitätsevaluation und Qualitätsentwicklung im Schulwesen: Forschungsbericht des Bundesministeriums für Unterricht und kulturelle Angelegenheiten.* Bildungsforschung des Bundesministeriums für Unterricht und kulturelle Angelegenheiten, 12. Vienna: StudienVerlag.

Rhyn, H. (Ed.) (2007). *Heterogenität, Gerechtigkeit und Exzellenz: Lebenslanges Lernen in der Wissensgesellschaft; OECD/CERI-Regionalseminar für die deutschsprachigen Länder in Nottwil (Schweiz) vom 26. bis 29. September 2005.* Schulentwicklung, 38. Innsbruck: StudienVerlag.

Rürup, M. (2003). *Ausländische und internationale Bildungsberichte als Orientierung für die nationale Bildungsberichterstattung in Deutschland,* Deutsches Institut für Internationale Pädagogische Forschung (DIPF). TiBi: 27. Retrieved from http://www.dipf.de/publikationen/tibi/tibi7_ruerup_3.pdf

Specht, W. (2007). Die Qualität des Bildungssystems, Standards und Monitoring. In Rhyn, H. (Ed.). *Heterogenität, Gerechtigkeit und Exzellenz. Lebenslanges Lernen in der Wissensgesellschaft; OECD/CERI-Regionalseminar für die deutschsprachigen Länder in Nottwil (Schweiz) vom 26. bis 29. September 2005.* Schulentwicklung, 38, 195-226. Innsbruck: StudienVerlag.

Educational Monitoring in Germany – A System Innovation to Safeguard Quality Standards

Norbert Maritzen

In recent years, educational monitoring has become established in Germany as part of a comprehensive system of quality development and standards assurance in education. This innovation is the consequence of a whole series of developments that can only be outlined briefly here:

- Public confidence in the performance of individual schools and of entire school systems has been seriously shaken in recent years. Against this background, the traditional lack of transparency in matters of school quality has become increasingly unacceptable. When public confidence is lost, this kind of chronic information shortage can only aggravate problems of legitimacy at the school level and beyond.

- Far-reaching developments leading to increasing heterogeneity in the situations and demands facing today's schools threaten the governability of the system as a whole. The social and economic trends of the modern age have triggered centrifugal developments in the school system that have increased the objective difficulty of governance and quality assurance, as well as the risks involved.

- At the same time, confidence in the government's ability to guarantee school quality and educational standards through the classic means of state school supervision has plummeted. Schools and school systems cannot be properly regulated either by centralistic standard programs of rule-based governance or by custodial models of supervision. With its very existence under threat, the state governance apparatus is thus modernizing itself by introducing new methods.

The measures subsumed under the term "educational monitoring" thus reflect a critical and comprehensive process of transformation in the governance[1] of state schools in Germany that is still far from completion. In the course of this process, the relationships among the different levels of the system (class, school, region, state), among the stakeholders (parents, teachers, principals), and among the instruments and procedures (normative conditional programming, empirical monitoring of the processes and effects of education) are being thoroughly reconfigured.

In the next section of this article, I outline key conceptual characteristics of educational monitoring and critically address its character as a system innovation. I then turn to the innovative function of educational monitoring and highlight the transfer issue as a major challenge for system innovations. Finally, I consider the potentials and limitations of measures of educational monitoring to overcome the chronic information shortages that characterize governance in education as well as the problems of legitimacy that frequently arise.

1 This article makes implicit reference to the growing body of German-language research on governance in the school domain; see Altrichter, Brüsemeister, & Wissinger (2007), Brüsemeister & Eubel (2008), Heinrich (2007), and Kussau & Brüsemeister (2007).

1. Educational Monitoring as a Program of Innovation

What Is Educational Monitoring?

There has been substantial change in some of the governance-related tasks entailed in the domain of quality development and standards assurance in recent years; new tasks have been added and others are set to follow. The primary tasks in this domain have for some time been subsumed under the term "educational monitoring."[2] A system of educational monitoring serves three *principal functions:*

- *Accreditation/certification:* This does not mean quality control in the narrow sense (e.g., by reference to ISO standards), but all procedures of official quality certification on the basis of predetermined, formalized standards, either at the individual level (e.g., central examinations) or at the organizational level (e.g., external evaluation of schools, inspection, etc.).

- *Accountability:* One of the main objectives of educational monitoring is to produce indicators of educational quality based on transparent rules, at the level of individual institutions as well as at the level of the system.

- *Diagnostic information for systemic learning:* Educational monitoring provides diagnostic information that can be put to practical use at different levels of the system. The products of educational monitoring must therefore be generated with a view to the question of how those products can serve to optimize "states."

Educational monitoring is the systematic and routine collection, analysis, and dissemination of information about an education system and the context in which it is embedded. Its objective is to improve educational planning and governance by

- observing the system (school system, individual schools, contexts),

- providing description, comparison, and analysis,

- examining specific questions,

- identifying potential points of intervention, and

- presenting feedback on findings.

The instrument of educational monitoring can be applied at all levels of the education system: preschool, elementary education, general secondary education, vocational education, further education, and higher education (with the exception of research). Edu-

2 The German-language scientific literature has only recently begun to address educational monitoring. This reflects the situation on the ground; it is only recently that broad concepts of educational monitoring have begun to be implemented in the school domain. For accounts of the theoretical background, see Avenarius et al. (2003), Fitz-Gibbon (1996), Hendriks (2004), ISB (2005), Konsortium Bildungsberichterstattung (2005), KMK (2006), and Scheerens, Glas, & Thomas (2003).

cational monitoring cannot be separated from the system of public administration (i.e., administration subject to parliamentary control). It is situated at the often tense interface between providing knowledge relevant to the governance of institutions (principle of responsibility), on the one hand, and of ensuring public access to information (principle of transparency), on the other. It is not unusual for these two principles to come into conflict with each other.

Educational Monitoring: Key Areas of Activity

In recent discussion, the following interrelated and interdependent areas of activity have been identified as integral to a system of educational monitoring:

- setting, maintaining, and developing process and product standards,
- assessments to monitor the attainment of standards and for national and international benchmarking,
- central examinations at points of transition or at the end of educational programs,
- development of indicators to measure key input, process, product, and context variables in the education system,
- data processing for purposes of educational statistics,
- external evaluation of educational institutions by inspectors,
- focused summative system evaluations and analyses,
- reports on education systems and individual institutions,
- accreditation/certification of statistical data assessments, external evaluations, and scientific projects conducted by “third parties” in the educational domain.

Introducing and organizing educational monitoring from a systemic perspective means:

- that procedures are not used on a one-off basis (e.g., to evaluate specific measures and programs), but that they are implemented on an ongoing basis throughout the system as a fixed component of the system of governance,
- that evaluation and information initiate and sustain processes of learning and development that are consistent with the principle of feedback-driven learning,
- that different kinds of empirical assessment (tests, official data, etc.), each serving specific purposes, can be analyzed within an overarching framework, and that synergies can be exploited,
- and, finally, that a theoretical framework relating the elements of input, process, and product is used to systematize and interpret findings (see Scheerens et al., 2003, p. 15).

Specifying the key areas of activity listed above for the school context identifies the following main clusters of tasks:

- developing and implementing curricular and other normative guidelines for the improvement of schools and instruction (outcome-oriented educational standards and frames of reference for school quality);

- developing (pools of) items for use in educational standards, final examinations, and standardized assessments of student learning (e.g., in cooperation between states or with the Institute for Educational Progress, IQB);

- designing and conducting assessments of student performance and reporting on their findings;

- designing and conducting central examinations at points of transition in the educational biography (e.g., graduation from secondary school);

- designing and conducting evaluations of specific projects or measures and reporting on their findings;

- conducting thematic analyses of existing data sets, with a focus on the effects and implications of decisions made at the system level;

- making the results of empirical educational research accessible to schools, teacher training institutes, advisory bodies, and education authorities;

- compiling and analyzing the data available from the official statistics and targeted surveys concerning individual schools, school levels, school types, student subgroups, and specific aspects of the education system;

- regular (system-specific) educational reporting based on a system of indicators;

- preparing feedback for individual schools in the form of “school report cards”;

- implementing regular inspections as an external assessment of quality in individual schools.

The concept of educational monitoring, which goes far beyond the series of measures laid out in the educational monitoring strategy adopted by the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder (KMK, 2006), is based on a model of evidence-based school and system development (Klieme, Steinert, Ciompa, & Gerecht, 2005). What is characteristic of this model is that feedback is provided on the quality of both processes and products. Moreover, the model stipulates coverage of as many as possible of the process factors that school effectiveness research has identified as characteristic of successful schools. The decisive argument for considering both product *and* process components derives from the question of how quality *development* can be achieved. For 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.

The Drawbacks of Externally Induced System Reform

Measures of educational monitoring complex, externally induced reforms; they reflect intensified efforts on the part of policy makers and education authorities to address problems that have built up over time. These problems can be consolidated into the following questions:

- *Increased strain on the system:* To what extent might school, as an establishment within which the cultural inter-generational contract is institutionalized, reach an impasse in what Ulrich Beck has described as "reflexive modern" society? Do the injustices and social disparities that are, in part, generated and exacerbated by the school system itself not show just how difficult it is for schooling in its traditional, historically stable, form to honor normative expectations?

- *International competition:* To what extent has globalization made the educational sector one that – because it is no longer seen exclusively in terms of costs, but also increasingly in terms of investment – helps to determine the competitiveness of national economies? If so, what are the consequences of long-term international benchmarking (PISA, PIRLS, and TIMSS), especially when its results give reason to doubt a positive "return on investment"?

- *Need for political–administrative reconfiguration:* What transformations are being seen in the governability of the educational sector, on the one hand, and – a closely connected point – in the ability to govern within the multilevel division of responsibilities for the education system, on the other? Are we not seeing governance being ever more clearly revealed as metaphoric and essentially illusionary? How wide can the "explanatory gap" grow between political and government administrative practice, on the one hand, and classroom and instructional practice, on the other, to offset legitimacy deficits and to prevent new ones from emerging?

- *Technical development:* There has been tremendous progress in the technical procedures used to generate system knowledge in recent years. Empirical educational research is thus undergoing a development that has already taken place in other research disciplines, with large-scale technologies that can only be handled by international research consortia being introduced to the field of education for the purposes of politics and administration. To what extent does this development promise increased rationalization? How is the paradigm of empirical knowledge acquisition related to patterns and procedures of decision making in educational policy (opinion-based vs. evidence-based policy), on the one hand, and to the enhancement of educational practice (case studies vs. "data-driven development of schools and instruction"), on the other?

- *Professionalization of diagnostic methods:* Measures of educational monitoring are also intended to help optimize operative routines in the teaching profession; for example, by providing diagnostic knowledge on students' academic attainment and learning gains that cannot be generated by individual teachers. Given that they are externally induced, however, it is not all too easy for such reform measures to lose their scientific innocence? Does "You can know more" not soon become "You should and must know more, with direct implications for your classroom practice!"? Can the character of this approach as an "appeal" not also be reversed to address "those in high places," who also should ap-

ply the inherent principle of being able to know and having to know to their governance practice, with direct consequences for their own decisions and actions?

2. Transfer as a Core Problem of System Innovation

The introduction of a comprehensive educational monitoring system is a challenging, demanding, and complex project. In the following, I discuss the innovation character of educational monitoring and highlight the transfer issue as a major challenge for system innovation[3]

Myths and Mental Models

Examination of the German states' activities to introduce measures of educational monitoring reveals the (mainly implicit) strategic assumptions underlying this process. In rather drastic terms, implementation is seen as imposed from the top down. This understanding is based on a number of myths or mental models that can prove very persistent. Some illustrative examples are listed below:

- "One size fits all": Tests, certification procedures, quality models, report formats, etc., are – given the scale of the undertaking – available in very few variants, the aim being to provide full and uniform coverage of as many schools as possible. There is limited scope to adapt procedures to local, school-specific conditions; such adaptations would, in some cases, be incompatible with the function of educational monitoring. Measures of systemic innovation are necessarily offered "off the peg." When only a very limited number of sizes is offered, however, it is not possible to tailor products to "shorter arms," "longer legs," or "broader girths."

- "If you cut an elephant in half you get two small elephants": Innovations intended to apply to the system as a whole tend to be introduced with the aid of three combinable approaches: (i) extending the implementation process over time, (ii) starting with small-scale experiments that are then rolled out to wider areas, and (iii) starting with partial implementation of the project and then adding to it successively. There are numerous good pragmatic reasons for all three strategic options, not least reasons of financing and political enforcement. At the same time, all three approaches are based on the problematic premise that a part contains the whole, and that the temporary, regional, or partial implementation of innovations provides a sufficient understanding of the challenges involved in its full-scale introduction. That is not usually the case.

- The belief in the "self-transforming power" of political programs: System reforms such as the introduction of educational monitoring are, in the first instance, political programs. They are put on the agenda of political bodies with the aim of forming and informing public debate in competition with the non-state/non-administrative opinion-makers of

3 The German-language literature on innovation transfer and implementation strategies in education makes very little reference to international research, and even less to empirically based theories or empirical research on the implementation or efficacy of monitoring procedures (e.g., Borman, Hewes, Overman, & Brown, 2002; Hall & Hord, 2005; Matthews & Sammons, 2004; Rogers, 2003; Scheerens et al., 2003).

the media and the associations. Consequently, the texts produced are declaratory in nature, announcing the introduction of measures with an overstated degree of certainty. Despite all discursive efforts not to be left behind by the "empirical revolution," the empirical evidence for the certainty of these effects is inevitably still lacking.

- Copy-and-paste myths – the seductive power of best practice examples: Little is more popular in the field of education than reference to good examples. Copy-and-paste myths drive ministers and trade unionists alike on pilgrimages to the countries that come out at the top of the PISA tables; they motivate teachers to participate in workshops with representatives of model schools. There is very little consideration of how local experience and knowledge generated on the job have informed successful innovative solutions, of the extent to which this knowledge and experience are and remain specific to the situation in which they evolved, or of the extent to which historical and cultural circumstances limit their transfer to other contexts.

Framework for the Implementation of Innovations

If we wish to avoid falling prey to the "logic of failure" (Dörner, 2003) summarized above, we must be mindful of the general challenges involved in innovation transfer and aware of the findings and insights that have emerged from transfer research. A tentative selection of guidelines is outlined below:

- Take realistic account of the conditions on the ground: The reception and uptake of proposed innovations "at the chalkface" is selective, as is their integration into "tried and tested" and "time honored" structures; processes of uptake and acceptance always imply individual cost–benefit calculations and win–lose situations.

- At all levels of the system, factor in enough time for beliefs to be adjusted and new competencies acquired: All innovations to some extent make unreasonable expectations. Even when the measures proposed meet with general acceptance, they are very likely to cause a certain degree of frustration.

- Move away from blueprints and recipes – and toward the implementation of innovations as a co-constructive learning process: Innovative concepts and measures essentially have to be "reinvented" in practice by all those involved. Only then can they fulfill their promise with tangible and substantive effects.

- Think in terms of "full coverage" from the outset: This applies to diagnosing problematic areas as well as to planning the implementation of innovations, developing strategies for their execution, evaluating their effects, and not least calculating their costs. The question to be asked is, What will be the cost of doing nothing?

- Success is the product of will, ability, and necessity: It is not enough to stimulate and sustain the will to change. Professional, often scientifically based expertise in the fields of governance and practice is also essential, as is serious commitment to the implementation of innovations. Those who neglect or refuse to take the necessary steps must be brought under increased pressure of legitimation.

Finally, innovation transfer must be integrated within a multidimensional strategic concept. The matrix below provides a heuristic framework that makes it possible to identify potential points of intervention and instruments for a systematic strategic implementation concept. The matrix arranges potential points of intervention by target areas (horizontal categories) and dimensions of innovation (vertical categories). Aspects and instruments to be addressed can be located in each field of the matrix, thus informing strategic planning. For example, the matrix could be filled as follows:

	Contents	**Structures**	**Persons**
Normative Dimension	Embedding in curricular objectives Frameworks for the development of schools and instruction	Guidelines for structures of commitment Instructions for cross-level cooperation	Professional standards Task descriptions Specifications of duties
Strategic Dimension	Full-scale empirical analyses Standard programs that allow for variation Information offensives	Global resource concepts Institutionalization Networking System maintenance Evaluation systems	Multi-professional co-operation structures Consultancy and training/qualification
Operative Dimension	Materials, tools, procedures, instruments	Responsibility structures and frameworks for action Monitoring systems Technical infrastructure	Professional learning communities Evidence of learning gains in staff and students

3. Trust and the “Will to Know”: A Tense Relationship

Systematic educational monitoring is being introduced in the context of far-reaching reforms of the education system. The PISA findings shattered public confidence not only in the performance of German schools, but also in the institutional governance bodies, which evidently had very little idea of the true extent of the problems within the system. Their neglect is just as unforgivable as the low performance of the schools themselves. Many of the restructuring measures introduced have been designed to win back public confidence in these institutions’ ability to provide good governance.

Trying to create trust by means of transparency is a deeply ambivalent undertaking, however, in that trust between individuals and institutions relies on the former renouncing the need to know everything about the latter. Luhmann has described trust as a “mechanism for the reduction of social complexity” (Luhmann, 2000a). Educational monitoring places both individual schools and school systems under constant observation; it has the poten-

tial to provide a more complex understanding and “picture” of the obscure object of observational desire. It does this by means of an elaborate observation system, the validity of which is to be established by means of scientific expertise. The greater the investment in the rationality of the process, however, the more clearly it can signal a message of mistrust: “We have to do this because you either can’t or don’t want to.” The constitution of observer–object relations in the school context is thus highly ambivalent. Finally, I will specify some dilemmas that arise for the governance dimension at the interface of knowledge and trust:

- Educational monitoring reflects a change in forms of governance and management in the school sector. The organs of school policy and administration are integrating policy-relevant knowledge more systematically into their decision-making processes and disseminating that knowledge more consistently throughout the entire, loosely connected system (administration, individual schools, support systems, users). As a result, traditional forms of generating system knowledge (e.g., by the school supervisory authorities) will become obsolete unless efforts are made to redefine their functions. Yet a knowledge-based system of governance increases the complexity of relations between stakeholders, rather than making them more transparent. Although knowledge is potentially available to all, responsibilities remain unequally distributed. The capacity to commit and to the capacity to solve problems come into competition: knowledge that the different stakeholders can easily commit to reflecting on together rarely solves problems; at the same time, stakeholders with growing levels of autonomy find it increasingly difficult to commit by discursive, i.e., non-hierarchical, means to decisions on issues that are increasingly well grasped in all their complexities.

- The classic motive for applying science is to absorb uncertainty (Luhmann, 2000b, p. 183ff.). Uncertainty is the result of an actual or apparent lack of knowledge. People and social systems that systematically promote the acquisition of scientific knowledge do so in order to achieve higher levels of certainty. To the extent that their efforts succeed, however, the illusion that advancing our knowledge increases the certainty of our actions and decisions fades. On the contrary, the better informed we are about problematic issues, the more options are likely available to us, and thus the higher our level of uncertainty. The more we know, the more aware we become of the limits of our knowledge.

- Knowledge cannot be separated from the context in which it is generated. Within politics and administration as well as within schools, conditions are so complex that scientific knowledge cannot simply be extracted from one context and applied in another. Rather, there is a need for an investigative approach to questions arising in one’s own sphere of action, and for knowledge to be integrated within a broader body of explicit and implicit theories and linked up with solution algorithms that are often acquired on a case-by-case basis. Stakeholders may well to have to go through a painful process of abandoning axioms that have taken root over many years. The provision of scientific knowledge has the potential to facilitate system governance, but that knowledge remains inactive if stakeholders do not succeed in sublating competing or non-scientific knowledge in the sense of Hegel.

- The relationship between knowledge and lack of knowledge is also rather ambivalent from the perspective of legitimation. The more finely nuanced the information available in the public sphere and the more widespread its dissemination (e.g., specific knowledge of unfair mechanisms of social selection within the school system, or of failure to deliver on programmatic reform objectives), the more urgently it compels action. If the repertoire of policy instruments available and the networked efforts of political, administrative, and school-based stakeholders lag behind that knowledge, either there is a temptation to close the legitimation gap through intentional ignorance ("There's no need for everybody to know everything"), or a chronic crisis of legitimacy develops, with implications for the future of public sector institutions. The will to know is thus associated with a number of risks.

The last antinomy raises the question of responsibility. Max Weber distinguished between the ethic of conviction and the ethic of responsibility (Weber, 1992). The former measures the "moral value" of programs of action – as which pedagogical and education policy projects can be described – solely in terms of their intentions and contents. If the school system is to remain an area of public responsibility, it is vital that all stakeholders address the question of how to take responsibility, effectively and sustainably, for its future conditions and outcomes. Scientific insights into the conditions and effects of our actions do not make life any easier, no matter how thorough and exhaustive they are; on the contrary. An ethic of responsibility prescribes critical engagement with science. At the same time, it proscribes an easy escape through the concealed door of self-imposed ignorance.

References

Altrichter, H., Brüsemeister, T., & Wissinger, J. (Eds.). (2007). *Educational Governance. Handlungskoordination und Steuerung im Bildungs Evaluation, Wissen und Nichtwissen.* Wiesbaden.

Avenarius, H., Ditton, H., Döbert, H., Klemm, K., Klieme, E., Rürüp, M., Tenorth, H.-E., Weishaupt, H., & Weiß, M. (2003). *Bildungsberichterstattung für Deutschland. Konzeption.* Frankfurt.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis.* (Report No. 59). Center for Research on the Education of Students Placed at Risk.

Brüsemeister, T., & Eubel, K.-D. (Eds.). (2008). *Evaluation, Wissen und Nichtwissen*. Wiesbaden.

Dörner, D. (2003). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen.* Reinbek bei Hamburg.

Fitz-Gibbon, C. T. (1996). *Monitoring education. Indicators, quality and effectiveness.* London.

Hall, G. E., & Hord, S. M. (2005). *Implementing change: Patterns, principles and potholes* (2nd ed.) Boston.

Heinrich, M. (2007). *Governance in der Schulentwicklung. Von der Autonomie zur evaluationsbasierten Steuerung.* Wiesbaden.

Hendriks, M. A. (2004). Benchmarking the quality of education. *European Educational Research Journal, 3*(1).

ISB – Staatsinstitut für Schulqualität und Bildungsforschung. (2005). *Theoretisches Rahmenkonzept: Bildungsberichterstattung und Bildungsmonitoring.* Munich.

Klieme, E., Steinert, B., Ciompa, R., & Gerecht, M. (2005). Auf dem Weg zu einem integrierten System der Qualitätssicherung für Schulen. In H. Avenarius, K. Klemm, E. Klieme, & J. Roitsch, J. (Eds.). *Bildung gestalten – erforschen – erlesen* (pp. 68-91). Neuwied.

KMK. (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring.* Bonn.

Kussau, J., & Brüsemeister, T. (2007). *Governance, Schule und Politik. Zwischen Antagonismus und Kooperation.* Wiesbaden.

Luhmann, N. (2000a). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität.* Stuttgart.

Luhmann, N. (2000b). *Organisation und Entscheidung.* Opladen.

Matthews, P., & Sammons, P. (2004). *Improvement through inspection. An evaluation of the impact of Ofsted's work.* London.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.) New York.

Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational evaluation, assessment and monitoring. A systematic approach.* Lisse.

Weber, M. (1992). *Politik als Beruf.* Stuttgart.

Consequences of Student Assessment Studies and Educational Monitoring,– The Outlook for Educational Standards and their Implementation

Kurt Reusser & Ueli Halbheer

1. Introduction

The report on the results achieved by Swiss students in PISA 2000 was entitled "Equipped for Life?" (*Für das Leben gerüstet?;* Zahner et al., 2002). This question reflects a particular understanding of the purpose of schooling: its task is presumed to lie in preparing its attendees for life. According to this view, the school system should provide students with a basic education and competencies that will help them to productively meet the challenges of working life and of society. PISA is understood as a way of measuring the effectiveness of this basic education, although this understanding of effectiveness is problematic in that it does not give (adequate) consideration to the non-technological character of learning and instruction or to the complex system of interrelated effects involved. It does, however, reflect a change of perspective in educational policy; recently, "output" has become more important as a point of reference for the governance of the education system, instead of the more traditional focus on the structure and processes of educational provision.

As part of this new philosophy of results-based governance, a new terminology has developed, pointing to a new, broader understanding of how the quality of school-based educational processes can be secured. In particular, "educational standards," understood to mean binding standards linked to measurements of competency and performance, which are to be applied in all schools, have been the subject of considerable debate in the German-speaking countries in recent years. The important point is that the education system should be measured according to its outcomes, which means that ways of achieving accountability are needed at the level of the system (policy-makers), at the level of the individual school, and at the level of each individual teacher (in particular with regard to in-service teacher training). The goal is for feedback on performance to be used to trigger improvements in quality at the input and process levels. Figure 1 illustrates this idea; the arrows emanating from the output represent the intended effects on the other two levels.

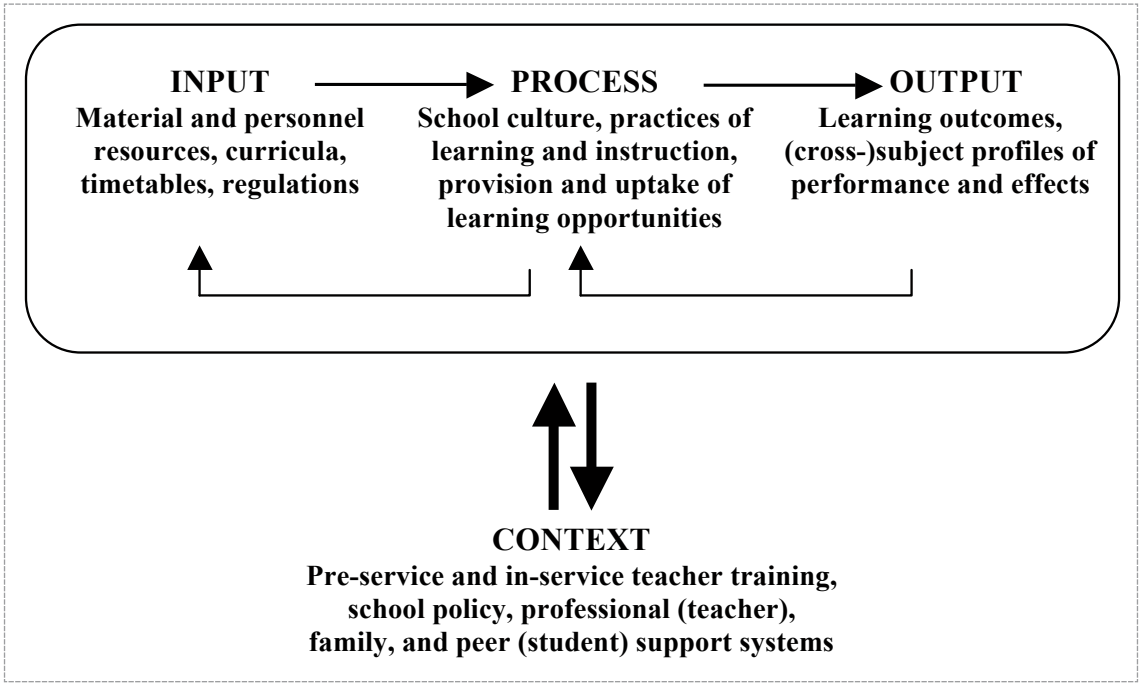

Fig. 1: A simple model of how education systems work

The account below is based on experiences gained with educational standards in a number of European countries and in the United States. Although there are many differences between the educational systems of the Netherlands, Sweden, England, and the United States, a number of shared elements are characteristic for the implementation of subject-based standards and the strategy of educational governance upon which that implementation is based. In particular, these include:

- The use of tests and examinations that focus on norms or ideal criteria,
- The presence of internal school evaluation programs and an external inspectorate,
- Attempts to nationalize curricula,
- The establishment of powerful feedback systems, which are linked to measures for improving teaching,
- Methods of implementation that are appropriate to the system and take the local level into account, and
- The reliability of a coherent policy.

Although we do not wish to attempt to provide a history of the concept of “educational standards”, we can state that standardization, according to the general meaning of the term, has always been a characteristic of the development of modern educational systems. Standards determine expectations in their area of validity and regulate practice. What is new is that for the first time, the academic performance of schools or the learning outcomes of students are to be empirically monitored across an entire geographical area.

In Germany and Switzerland, the need for new strategies to secure educational quality was discussed prior to PISA 2000, although the unsatisfactory results of that study dominated the debate about the introduction of educational standards, especially in Germany. In the wake of the review *The Development of National Educational Standards in Ger-*

many (Klieme et al., 2004), binding standards were introduced across all of the federal states. The newly created Institute for Educational Progress (IQB) was commissioned with developing test items for a national process of educational monitoring. In Switzerland, the Agreement on the Harmonization of Compulsory Schooling (Konkordats über die Harmonisierung der obligatorischen Schule, HarmoS) states that minimum standards for compulsory schooling should be introduced for the subjects of mathematics, the first (local) national language, the second national language, another foreign language, and science (Schweizerische Konferenz der kantonalen Erziehungsdirektoren, EDK, 2004).

2. "Implementation": An Interplay between Internal and External Impulses

Any new governance philosophy requires appropriate methods. In general, we can differentiate between the following approaches:

- *Outside-in perspective:* This perspective is characterized by a top-down view. A centralized, governmental form of management, often seen as technocratic, its leading concept is that reform packages will be developed from the outside and adopted by the system. It attempts to ensure that a system reform is coherent by defining administrative guidelines, objectives, rules, and instruments (e.g., teaching materials and curricula).

- *Inside-out perspective:* In accordance with a bottom-up understanding of innovation, this approach is based on the idea of participatory governance in the sense of a philosophy of school development. As a horizontal, contextual form of governance allowing partial autonomy, it permits external requirements to be adapted to regional or local contexts.

From a scientific point of view, any form of educational governance that is to promise success must combine the potential of these two perspectives; external objectives and stipulations must be adopted, but these must also be translated and adapted to the micro-context of the local situation. In other words, all of those involved must engage in a broad and ongoing process of learning, meaning that individuals, entire institutions, and systems must all undergo learning processes aimed at changing latent beliefs and habits that are very deeply anchored. Such processes are often very slow moving. In particular, complex ideas and reform intentions are integrated one step at a time via a number of different phases of acceptance: a phase during which the new ideas are employed in a non-binding fashion is followed by one in which they are integrated into institutional practice, and finally by a phase of renewal. In other words, quality can only develop through an interplay between external impulses and productive learning at the grass roots of the system. For this reason, it must be ensured that all of those involved accept the changes to the extent that they experience ownership of the innovation and that it makes sense to them. For an innovation to actually take effect in everyday school practice, what is needed is more than a superficial assimilation of new rules for securing quality; changes in attitude that concern the professional identity of teachers are required.

In this context, the fact that the educational system is made up of multiple levels presents an additional challenge. Figure 2 outlines a simple model that identifies the stakeholders at the different levels, and some of the activities and functions that are important in educational reforms.

National level: Federal government, states, and cantons (educational governance at a macro-level)
Regional support and monitoring systems at an intermediate level (in-service teacher training, quality improvement and evaluation)
Individual schools (meso-level) (institutional school work and improvement of teaching)
Teachers, students, and parents (teaching and learning at the micro-level of the individual lesson)

Fig. 2: System levels at which educational standards are implemented

Transfers designed for top-down implementation can encounter difficulties because the transitions between these levels are also transitions between different forms of knowledge and reasons for action. In other words, not all of the stakeholders involved interact with each other in the same way. In order for the intention behind a reform to be accepted by those responsible for implementing it, a variety of methods of translation and transformation are needed at the system interfaces. These translation and transformation processes are, ultimately, always based on learning – learning that takes place at both an individual and a collective level. Besides the fact that a comprehensive and long-term improvement in quality requires the involvement of all of the stakeholders at all levels, it must also refer to the three pillars of the quality system:

- The output dimension, which is managed through results and responses, i.e., by monitoring the results of performance tests,
- The process dimension, which is managed through the quality of teaching and learning processes, i.e., by enhancing the professionalism of the teaching staff,
- The input dimension, which is managed through objectives and resources, i.e., by designing curricula and standards.

An approach to implementation such as that presented here must reach every level (bottom-up and top-down) and all functional areas of the education system in order to meet its objective of ensuring quality. Introducing educational standards is a project that affects the entire school system; it must trigger the entire system to think about innovative ways of improving quality. As such, it offers an opportunity to use tools that may have a long-term effect on improving the quality of schools and of teaching.

3. Quality Development and Quality Assurance Tools at each Level [1]

Any profession needs technical, cognitive, social, and organizational instruments and methods for its development. These tools are the tangible form of processes of change. Even if education can only be technologized to a limited extent, it requires both tangible

1 This account is based on Oelkers and Reusser (2008).

and intangible tools and resources, as do educational reforms. Similarly, any changes to the system of teaching, the core of an education system, do not take place of their own accord; they require both triggers from and the support provided by a variety of tools. However, these are needed not only at the level of actual teaching, but also at every level of the system.

Below, we outline some tools that can be used at the different levels of the system; these are described in more detail in the review *Qualität entwickeln – Standards sichern – mit Differenz umgehen* (Oelkers & Reusser, 2008).

At the level of the system as a whole:

- At the level of the system, *educational standards* and the associated competency models and test instruments are the main tools for the overall reform process. If standards are to be accepted and internalized by the teaching profession as a framework of reference for interschool comparisons, they must be concise and comprehensible, must focus on the core areas and basic competencies of the subject in an understandable and justifiable way, and must meet high quality criteria in terms of didactics and methodology. It must be possible to illustrate and explain the standards using examples and tasks, and they must be seen as educationally valuable and capable of being put into practice. Equally, "teaching to the test" must not become the main experience of school-based learning.

- In addition, *educational monitoring* across the system, in which tests and other indicators of the present situation of an educational system are recorded regularly and over time, is indispensable for political accountability. It can by no means replace pedagogical processes or diagnoses of individual issues. Further information can be gained through regional surveys of learning progress and parallel studies, as are currently implemented in some parts of Germany and Switzerland. One vital issue is that any test results obtained must be statistically controlled for context or input parameters that affect educational performance; this ensures that fair comparisons can be made. Teachers will need assistance to interpret the results in a way that is useful to them in improving their own teaching.

- *Curricula* are another essential tool for the implementation of educational standards; they list the essential elements for each subject and define its inner structure and relationship with other subjects, whereas standards focus on the core competencies in each subject. Essentially, curricula are focused on inputs, whereas standards are designed to specify output in terms of measurable knowledge and skills. The complementary nature of this relationship is clear when core elements of the curriculum and the related competency descriptions are reflected in teaching practice and performance tests. Similarly, learning material that is used in some subjects as an instrument for managing instruction at a micro-level must cover learning contents relevant to competency standards without reducing the breadth of the curriculum.

- In addition, attention must be paid to the role played by the *school supervisory* bodies and *the school inspection system*. Any findings that allow well-founded and comparable statements to be made on the performance of students in a particular school must be dealt with very carefully. As the example of the Netherlands shows, the supervisory bodies must become an inherent part of a school's culture and must be capable of creating a productive culture of evaluation.

On the level of networked individual schools:

- Pilot programs such as the SINUS project ("Increasing the efficiency of mathematics and science teaching", Prenzel, 2000) initiated by the Bund-Länder Commission for Educational Planning and Research Promotion (BLK) have shown how *networks and learning communities* can be used to improve teaching at a local level, focusing on particular problems. Institutionalized cooperative networks can be useful, both for interpreting performance results and drawing conclusions from them, and on this basis for improving teaching in all schools.

- During the 1990s, there was a trend towards allowing schools more freedom to design their own activities; individual schools were increasingly seen as locations for innovation. This drew attention to a number of aspects of the ways in which schools are organized, which can play an important role in re-contextualizing reforms. Innovation research has shown that *head teachers* have a considerable influence on how the teachers in their school respond to reforms (Specht & Freudenthaler, 2004).

- Equally, *self-evaluation and external evaluation,* as part of the process of accountability, are core elements in the repertoire of school autonomy (or partial autonomy). They play an important role in providing information and feedback on educational processes, and thus complement the output information gained through performance tests. The evaluation of the SINUS pilot program has shown that incentives for ensuring that teachers cooperate fully can be valuable in the dissemination of new knowledge relating to both teaching in general and the teaching of specific subjects.

At the level of instruction:

Any process of school innovation that does not affect instruction and thus change students' learning processes is of little benefit. The reverse also applies: the outputs cannot be of high quality unless the processes are too, and learning outcomes will not meet the standards required unless appropriate learning opportunities are provided for all students. Teaching, the micro-level, is the litmus test for whether and how educational standards are implemented. As such, *improving schools* means *improving instruction*. High-quality forms of support and "sense-making" processes are particularly important at this level; teachers are being targeted at the very core of their professional behavior, and they want and need to be effectively supported in making these changes. Below, we list a number of instruments, some of which have been developed from recent research, for developing good instruction.

- Whether or not educational monitoring will have positive effects on instruction depends not least on whether teachers can make productive use of the *feedback* provided at the school and class level. In recent years, teachers have been able to make increasing use of tests developed to allow valid comparisons; teachers can compare the results achieved by their students with those of a sufficiently large and representative sample. In Switzerland, two new formats, "Klassencockpit" (Class Cockpit) (Klassencockpit, not dated) and "Check 5" (Kanton Aargau, Departement Bildung, Kultur, 2006) have been established and are being used by an increasing number of teachers to measure learning outcomes.

- When discussing the system as a whole above, we pointed to the importance of *learning material;* we should also note the importance of the *task culture* communicated by that material. In general, the tasks used in instruction should be consistent with the level and complexity of the relevant educational standards. The TIMSS and PISA studies have shown that students in Germany are relatively good at solving routine tasks, but have problems applying what they have learned. Indeed, the processing of elementary tasks takes up most of the time in mathematics classes. There is a clear need for more tasks that help to promote understanding and involve realistic applications of the knowledge gained, and/or that can be solved in several different ways; these tasks must be deployed intelligently to allow productive learning.

- *Pre-service and in-service teacher training* will be of major importance; training formats will have to be developed that allow teachers to analyze and reflect on their own teaching. To this end, a number of tools have been developed over recent years. These include subject-specific pedagogical coaching (e.g., Staub, 2004) or the use of video in improving teaching (e.g., Krammer & Reusser, 2004). Although these methods are labor-intensive, they allow teachers to examine their professional behavior closely and to attain a shared didactic understanding of learning and instruction, to discard previously unquestioned assumptions, and to begin the process of changing their professional behavior. An evaluation is needed of whether standards should also be developed for teacher training. One reason for considering this is that, at present, the provision of teacher training tends to be arbitrary, and this could be countered with specific opportunities for professional development.

The tools for quality assurance outlined here should allow the project of reform through educational standards to be contextualized in the existing structures of the educational system in a co-constructive and culture-sensitive way. Educational standards can thus become a central element in securing the quality of inputs, of processes, and of output, at every level of the school system.

4. Conclusions[2]

On the basis of these ideas, we can formulate some conclusions aimed both at the policy-makers concerned with creating and introducing educational standards and, in particular, at the schools and teachers on the ground:

2 This section is based on Oelkers and Reusser (2008).

- The introduction of educational standards requires a coherent quality policy for the educational system, in which quality development, the use of standards as a basic point of reference, and a culture of evaluation all come together.

- Educational standards that are monitored at regular intervals are innovative and useful because they measure the quality of the educational systems on the basis of output, and because they have the potential to visibly improve school culture and teaching.

- The implementation of educational standards must not focus on establishing a culture of testing as a method of governance; instead, the main objective should be to improve schools and teaching.

- The reform philosophy must be founded in a broad, multi-dimensional understanding of educational standards, which must include educational processes and input as well as goals and expectations characterized by multiple criteria.

- In order to create educational standards founded on competency models and to develop tests and pools of test items on that basis, three things are needed: the expertise of academic researchers, an appropriate infrastructure, and a time period of several years for development.

- Implementing educational standards will, it is hoped, achieve ambitious goals across a broad and complex system. This will only be possible if a long-term view is taken and effective ways are found of scaling the process up.

- Any educational monitoring system focused on output, and any external evaluation and monitoring of the degree to which goals are met at the level of the individual school, must involve school autonomy at the regional and local level, which must be meaningful and secured by law. Bureaucratic micro-management and top-down controls, which are often unsuitable for local conditions, must be abandoned in favor of horizontal, participatory, and contextual governance of the educational system.

- The results of tests must not be used on their own, without additional information, to determine students' future school careers; under no circumstances should they be linked to school funding or to promotion for teachers. Any such measures would undermine the usefulness of tests for improving quality. Incentives for improvement and the communication of professional best practice are more productive than a mentality of competition within or between schools that involves demotivating sanctions for teachers.

- Teachers must be recruited to support the reform: they are the most important stakeholders involved and those who can ensure its success. Along with a shared perception of current problems, this requires an adequately developed and realistic understanding of the reform's objectives and of the information that the educational standards and the tests linked to them can provide.

- A major paradigm shift such as changing the philosophy of educational governance requires a major input of resources and in-service training. Without additional investment, and without combining all of the resources available in order to realize this vision and to create the necessary preconditions, educational standards cannot be successfully implemented.

- Tools like those outlined above are needed not merely for the education system itself but also for the reform process.

- Support systems and providers of teacher education and training must be incorporated in the transformation process to ensure that the standards take effect at the targeted levels. Policy-makers must aim to ensure that the work of the contextual systems is aligned as closely as possible with the objectives of the reform initiatives, without focusing the support systems on that task to the exclusion of all others. In particular, innovative forms of compulsory and targeted continuing in-service teacher training in didactics and pedagogy are needed.

- As the introduction of educational standards is a complex design process aiming to integrate a new philosophy of governance into existing educational systems and their subsystems in a harmonious way, authentic models and examples are needed at every level and in every aspect of the project. Existing and new projects for improving quality should be coordinated with the implementation of educational standards or should be designed with that goal in mind from the outset.

- For the whole project of implementing educational standards, policy-makers must be aware both of their own limitations and of the limitations of the existing empirical research. The educational system cannot be governed by educational policy or calibrated by researchers. In terms of actual educational activity, school-based learning can only ever be managed to a very limited extent. Through professional teaching and positive conditions for learning, teachers provide educational opportunities for students; whether these opportunities are taken up and used productively depends not only on the teacher but also on the learner, and on a whole range of contextual and systemic variables.

- Implementing educational standards requires structurally resilient academic research on education that is independent of politics and policy. This research must not focus only on evaluation or on analysis and diagnosis, but also on developing pedagogy, and it must be capable of identifying and criticizing developments that are going astray.

References

Kanton Aargau, Departement Bildung, Kultur. (2006). 80 Prozent aller 5. Klassen beteiligten sich am Check 5. Retrieved March 20, 2006, from http://www.ag.ch/leistungstest/de/pub/check5/zweite_durchfuehrung.php

Klassencockpit. (not dated). *Ein Instrument zur Qualitätssicherung in der Volksschule.* St. Gallen: Kantonaler Lehrmittelverlag St. Gallen. Retrieved March 20, 2006, from http://www.klassencockpit.ch/informationen/unterstufe.aspx

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2004). *The development of national educational standards in Germany.* Bundesministerium für Bildung und Forschung. Bonn: BMBF.

Krammer, K., & Reusser, K. (2004). Unterrichtsvideos als Medium der Lehrerinnen- und Lehrerbildung. *Das Seminar, 10*(4), 80-101.

Oelkers, J., & Reusser, K. (2008). *Expertise: Qualität entwickeln – Standards sichern – mit Differenz umgehen.* Berlin: BMBF.

Prenzel, M. (2000). Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts. Ein Modellversuchsprogramm von Bund und Ländern. *Unterrichtswissenschaft, 28,* 103-126.

Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK). (2004). *HARMOS: Zielsetzungen und Konzeption.* Berne: EDK. Retrieved July 21, 2008, from http://www.edk.ch/dyn/11659.php

Specht, W., & Freudenthaler, H. (2004). Bildungsstandards. Bedingungen ihrer Wirksamkeit. *Erziehung und Unterricht, 154,* 618-629.

Staub, F. C. (2004). Fachspezifisch-Pädagogisches Coaching: Ein Beispiel zur Entwicklung von Lehrerfortbildung und Unterrichtskompetenz als Kooperation. In D. Lenzen, J. Baumert, R. Watermann, & U. Trautwein (Eds.). *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung. 3. Beiheft der Zeitschrift für Erziehungswissenschaft* (pp. 113-142). Wiesbaden: VS Verlag Sozialwissenschaften.

Zahner, C., Meyer, A. H., Moser, U., Brühwiler, C., Coradi Vellacott, M., Huber, M., Malti, T., Ramseier, E., Wolter, S., & Zutavern, M. (2002). *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen – Nationaler Bericht der Erhebung PISA 2000.* BFS/EDK: Neuchâtel.

Reflecting on Developmental Problems in Education Systems: Thirty Years of OCED/CERI Regional Seminars for Germany, Austria, and Switzerland, 1977-2007

Jan Hofmann

In 1977, during the first OECD/CERI regional seminar for the German-speaking countries in Dillingen on the Danube, the idea that European countries would one day be able to carry out research, to learn, and to study across national frontiers without restriction seemed a thing of fantasy. The existence of the two Germanies, the position of West and East Germany as frontier states on either side of the Iron Curtain, the unpopular combined state for the Czechs and Slovaks, and the forced multi-ethnic state of Yugoslavia under Marshall Tito all prevented political – and educational – developments that would later become a reality in a united Europe. Today, the face of Europe has changed.

Looking at educational policy in the German-speaking area of Germany, Austria, and (much of) Switzerland, we can see that in parallel with major social changes, the educational systems in each state have undergone a transformation. And the methodologies of, and basic questions for, educational research have changed utterly.

The OECD/CERI seminars, 16 in total, held every two years until 2007, threw light on these developments; indeed, in some cases they even sparked off a variety of changes themselves. The first 15 seminars were divided into groups of three, each group focusing on a common theme that was investigated by each country from a different angle: *Evaluating Educational Innovation* (1977/1979/1981), *Internal Reform of the School System* (1983/1985/1987), *New Developments and Challenges for Schools* (1989/1991/1993), *Learning in a Dynamic, Open Society* (1995/1997/1999), and *Learning in the Knowledge Society* (2001/2003/ 2005).

The seminar held in Potsdam in 2007 (on *Development and Implementation of Educational Innovations as a Consequence of Educational Monitoring, Educational Reporting, and Comparative Studies of Student Performance – Opportunities and Limitations*) was a stand-alone event. It represents a turning-point in many ways, one being the reform of German federalism, which threatened to put an end to German funding for the entire seminar series. It can be seen as a moment at which those involved took stock, and as a bridge between the work done in the past and that to be done in the future; both an end and a beginning.

Looking back, we can interpret the first seminar, in 1977, as a reaction to the crisis in the governance-oriented education policy that had dominated from 1965 to 1977. During the 1960s, the model of governance popularized by John F. Kennedy, with scientific advisors providing knowledge for political decision-makers, coincided with a sense of unease about social conditions. The student protest movement made its concerns heard loudly and clearly; the consequences, in Germany, Austria, and Switzerland, differed in intensity. In Germany, the movement escalated, driving some of those involved away from democracy and eventually underground to form groups such as the Red Army Faction and the Movement 2 June. From there, they presented the Federal Republic, its social-

liberal government led by Helmut Schmidt, and the population as a whole with an autumn of turmoil, murdering a number of representatives of the system in the very year in which our first seminar took place.

During this period, what Germans call “reform gridlock” was identified almost everywhere; empirical research, planning and governance would, it was believed, enable us to eliminate all the problems of the education system. And with an overly positive assessment of the countries’ economies and finances, there were no failings that would be left undiscovered and no problems that would not be solvable, if policymakers created the right conditions. This feeling is reflected in the topics of the seminars held in Neusiedl/See, Austria, in 1979, on the importance of pilot projects, and in Basle, Switzerland, in 1981, which investigated the concept of the “good school” and the “opportunities for work” it involved.

But a decade later, a sense of disillusionment had already become widespread. Participants made use of the opportunities presented by the tri-national *lingua franca* to take a look at how their neighbors were doing. Was the grass really greener on the other side of the fence? Or did the neighbors in fact have even more serious problems? Had they perhaps managed to move the discussion forward?

Flicking through the various publications documenting the individual seminars, we can clearly sense, between the lines, a feeling of relief that those “next door” were not overachievers either, that educational reform seemed to be a slow process everywhere, and that quick fixes or back-of-the-envelope solutions were useful only for overheated debates in talk shows or as saber-rattling in the run-up to general elections.

It was precisely the different interpretations of the idea of federalism that allowed us, mindful of the diversity between the neighboring states of Germany, Switzerland, and Austria, to identify many new opportunities – some almost utopian – to learn from each other, despite or perhaps even because of those differences, and to develop common goals and ambitions for education policy.

In that sense it is not surprising that the second set of seminars (between 1983 and 1987) was concerned with designing educational careers, reforming instructional content and methods, and quality of qualifications. In retrospect, coming up with new ideas was a very appropriate pursuit for a Europe through which a wind of change was sweeping, culminating in the fall of the Iron Curtain in 1989 and the reunification of Germany in 1990. It also went hand-in-hand with a change of political direction in (West) Germany, which was governed by a Christian Democratic/Liberal coalition for 16 years between October 1982 and September 1998 (becoming the government of the whole country after reunification).

The clear and logical consequence, evident even at the time, was that the third cycle of seminars (in 1989, 1991, and 1993) would be starkly affected by the rapidly changing political and social environment.

This was the “decade of self-help” in the educational sector, or, as Rainer Brockmeyer called it, that of “local school reforms” (in *Innovationen im Bildungswesen als überregionale Aufgabe: Die OECD/CERI-Seminare der BRD, Österreichs und der Schweiz 1977-1993,* Köllen Verlag, 1995, p. 16).

A universal development, one that became most obvious after the German Democratic Republic was amalgamated into the Federal Republic of Germany, was that children and young people now had completely different experiences behind them. Lobbyists, the economy, technology, and general education all made their claims on the education system. This set of seminars was concerned with re-orientation, with sorting and filtering. For that reason, the question of how schools could open themselves up to these new developments became the umbrella topic for the fourth cycle of seminars.

In 1989, in Bremen, the most important issues were clarifying the changed situation in which education found itself, discovering new methods and the time and space needed for these changes, and generating supports and in-service training for teaching staff.

Two years later, in Geras, Austria, in 1991, the seminar participants – with one eye, we can be sure, on the continuing geographical expansion of the European Union – discussed the demands of intercultural education and the problem of new qualifications given the increasing internationalization of education.

In 1993, in Einsiedeln in Switzerland, the main topic was the role played by central government, both in terms of guidelines and management. Given the developmental tasks that individual schools were supposed to fulfill, these central entities needed to be redefined.

The six seminars that followed, over 14 years between 1993 and 2007, examined topics such as *Learning in a Dynamic, Open Society – The Role of Schools* (Dresden, 1995), discussed *School Principals and Supervisory Authorities: New Roles and Challenges for the School System in a Dynamic, Open Society* (Innsbruck, 1997), demanded that schools *Orchestrate Diversity – The Management Role of Central Authorities Given Greater Independence for Schools* (Rheinfelden, 1999), debated *Learning In a Knowledge Society* (Esslingen, 2001) and *Lifelong Learning In a Knowledge Society* (Graz, 2003), and finally investigated the question of *Heterogeneity, Justice, and Excellence: Lifelong Learning in the Knowledge Society* (Nottwil, 2005).

And today?

Today, we find ourselves in crisis once again (or still in crisis, depending on your point of view). The Germans, in particular, were jolted into action by the “PISA shock” of 2000. Education has become the number one topic for everyone from manual workers to managers. One consequence is that increasing numbers of political decision makers have accepted that ongoing educational research – not project-based research, but scientific research designed to be carried on over decades – is a vital precondition for providing better educational services.

The question is no longer *whether* research into education should be carried out, but *how*, and *how much*. At the same time, we have developed certain reservations about black-and-white statistics that can be mis-interpreted and mis-used if the context in which they arose, or the specific national characteristics that affected them, are ignored. There is also considerable debate about the inherent limits to what research can tell us; effective ways of evaluating the research itself are also needed. In addition, research that is lo-

cated close to policymaking can be accused of giving up its independence and capacity for criticism. It seems that as one problem is solved, a new one is created; we must be permanently on the lookout for such issues. A system that can manage itself, continuously producing high-quality results without friction or tension, is an impossibility, at least in the area of school-based education.

Historically, educationalists in the German-speaking world believed in the “Nuremberg funnel” through which unlimited amounts of knowledge could be channeled into each and every child. Today, with our overburdened timetables, flexible classes, and 8-year college-bound tracks, one sometimes gets the impression that the funnel is being forced down the throat of the school itself; schools just have to swallow whatever they are fed. Still the most important and formative educational institution, schools are called on almost as a reflex whenever some societal task is to be fulfilled. From environmental awareness and a European identity through to anti-aggression training, schools, or rather, a highly diverse system of different types of school, are supposed to take care of it. Along the way, of course, they should be preparing the ground for future Nobel prize winners as well as integrating every child from an educationally deprived background.

In reality, these demands are being placed on institutions that, partly because of the parlous state of public finances, barely even manage to fulfill their key task of giving every child an opportunity to discover learning as an opportunity for development and growth and thus of fostering individual excellence. Whatever additional responsibilities are to be placed on schools, it is increasingly evident that the teaching staff must be involved more closely and that the standards to be applied should be as transparent as possible, as should the way they are implemented, monitored, and assessed.

The hunt for innovative educational concepts is a fragile hermeneutic circle of the possible. Researchers investigate selected aspects of school life and propose steps that might help to alleviate the problems identified. This brings us back to the idea I suggested above: a multiplicity of ideas may help us to develop utopian solutions, the ideal situation.

Politicians can build on this foundation to create the framework for educational change, although they will, of course, be limited by the arguments of interest groups and political realities. The way these inputs are realized depends on the extent to which individual institutions and teachers get involved. High-quality assessment can improve practice and trigger new questions for research. But if not all the players come aboard, or are taken on board, we will be left with missed opportunities.

The participants at the OECD/CERI regional seminars are all experienced practitioners in their fields, and yet they appreciate the opportunities that this forum provides. The seminars provide an opportunity for participants to discuss even their most fleeting thoughts, to develop ideas, and to speak aloud their hopes for the future.

The seminar series has proven its worth as a place for courteous, yet forthright debate. Or in other words: as good neighbors we have found that it is well worth getting together to exchange notes, and we look forward to doing so again in 2009.

Thirty Years of OECD/CERI Regional Seminars for the German-Speaking Countries

Year	Location	Topic	Publication
Evaluating Educational Innovation			
1977	Dillingen (Germany)	"Evaluating Educational Innovation"	*Evaluation schulischer Neuerungen,* OECD/CERI-Seminar, Dillingen (Germany), 1977, Verlag Klett-Cotta, Stuttgart, 249 pp.
1979	Neusiedl am See (Austria)	"Dimensions of and Limits on the Evaluation of Educational Innovations"	*Dimensionen und Grenzen der Evaluation schulischer Neuerungen,* OECD/CERI-Seminar, Neusiedl am See (Austria), 1979, Österreichischer Bundesverlag, Vienna, 262 pp.
1981	Muttenz bei Basel (Switzerland)	"Evaluating Innovations in the Elementary/Primary School"	*Evaluation von Innovationen im Bereich der Grundschule/Primarschule,* OECD/CERI-Seminar, Muttenz bei Basel (Switzerland), 1981, Verlag Paul Haupt, Berne, 465 pp.
Manner, Scope, and Methods of Internal Reform of the School System			
1983	Berlin (Germany)	"Curriculum Development and Instructional Practice in Germany, Austria, and Switzerland"	*Lehrplanentwicklung und Schulpraxis in Deutschland, Österreich und der Schweiz,* OECD/CERI-Seminar, Berlin (Germany), 1983, Köllen Verlag, Bonn-Oedekoven, 308 pp.
1985	Salzburg (Austria)	"Making the Transition from Compulsory School to the World of Work"	*Der Übergang von der Pflichtschule in das Berufsleben,* OECD/CERI-Seminar, Salzburg, 1985, Köllen Verlag, Bonn-Oedekoven, 181 pp.
1987	Lucerne (Switzerland)	"Willingness and Ability to Learn Between School and Working Life"	*Lernbereitschaft und Lernfähigkeit zwischen Schule und Beruf,* OECD/CERI-Seminar, Lucerne, 1987, Köllen Verlag, Bonn-Oedekoven, 186 pp.
New Developments and Challenges for Schools			
1989	Bremerhaven (Germany)	"How Can Schools Become Receptive to New Developments and Challenges?"	*Wie öffnet sich die Schule neuen Entwicklungen und Aufgaben?,* OECD/CERI-Seminar, Bremerhaven, 1989, Köllen Verlag, Bonn-Oedekoven, 234 pp.
1991	Geras (Austria)	"Learning for Europe: New Forms of Living and Learning at School"	*Lernen für Europa, Neue Lebens- und Lernformen in der Schule,* OECD/CERI-Seminar, Stift Geras (Austria), 1991, Köllen Verlag, Bonn-Oedekoven, 102 pp.
1993	Einsiedeln (Switzerland)	"How Can Schools Contribute to School Development?"	*Was können Schulen für die Schulentwicklung leisten?,* OECD/CERI-Seminar, Einsiedeln, 1993, Köllen Verlag, Bonn-Buschdorf, 347 pp.

Learning in a Dynamic, Open Society			
1995	Dresden (Germany)	"Learning in a Dynamic, Open Society – The Role of Schools"	*Lernen in einer dynamischen und offenen Gesellschaft - die Rolle der Schule,* OECD/CERI-Seminar, Dresden (Germany), 1995, Köllen Verlag, Bonn-Buschdorf, 303 pp.
1997	Innsbruck (Austria)	"School Principals and Supervisory Authorities: New Roles and Challenges for the School System in a Dynamic, Open Society"	*Schulleitung und Schulaufsicht - Neue Rollen und Aufgaben im Schulwesen einer dynamischen und offenen Gesellschaft,* OECD/CERI-Seminar 1997, Innsbruck (Austria), 1998, Studien Verlag, Innsbruck, 339 pp.
1999	Rheinfelden (Switzerland)	"Orchestrating Diversity – The Management Role of Central Authorities Given Greater Independence for Schools"	*Die Vielfalt orchestrieren - Steuerungsaufgaben der zentralen Instanz bei grösserer Selbständigkeit der Einzelschule,* OECD/CERI-Seminar 1999, Rheinfelden (Switzerland), 2000, Studien Verlag, Innsbruck, 254 pp.
Learning in a Knowledge Society			
2001	Esslingen (Germany)	"Learning in a Knowledge Society"	*Lernen in der Wissensgesellschaft,* OECD/CERI-Seminar Esslingen (Germany), 2001, Studien Verlag, Innsbruck, 400 pp.
2003	Graz (Austria)	"Lifelong Learning in the Knowledge Society: Preconditions and Environment"	*Lebenslanges Lernen in der Wissensgesellschaft. Voraussetzungen und Rahmenbedingungen,* OECD/CERI-Seminar, Graz (Austria) 2003, Studien Verlag, Innsbruck, 242 pp.
2005	Nottwil (Switzerland)	"Heterogeneity, Justice, and Excellence: Lifelong Learning in the Knowledge Society"	*Heterogenität, Gerechtigkeit und Exzellenz. Lebenslanges Lernen in der Wissensgesellschaft,* OECD/CERI-Seminar Nottwil (Switzerland), 2005, StudienVerlag Innsbruck, 260 pp.
Stand-Alone Seminar			
2007	Potsdam (Germany)	"Development and Implementation of Educational Innovations as a Consequence of Educational Monitoring, Educational Reporting, and Comparative Studies of Student Performance – Opportunities and Limitations."	*Bildungsmonitoring, Vergleichsstudien und Innovation,* OECD/CERI-Seminar Potsdam (Germany), 2007, Berliner Wissenschafts-Verlag Berlin, 206 pp.

Contributors

Hans Ambühl
Secretary General of the Swiss Conference of Cantonal Ministers of Education (EDK)

Haus der Kantone
Speichergasse
Postfach 660
3001 Berne 7
Switzerland
Tel.: +41 31 309 51 11
Fax: +41 31 309 51 50
www.edk.ch

Dr. Anton Dobart
Director General of the Division for General Education, Educational Planning, and International Affairs, Federal Ministry for Education, the Arts and Culture, Austria

Federal Ministry for Education, the Arts and Culture
Sektion I
Minoritenplatz 5
1014 Vienna
Austria
E-Mail: anton.dobart@bmukk.gv.at

Prof. Dr. Hans Döbert
German Institute for International Educational Research (DIPF), Center for Planning and Financing in Education, coordinator of national educational reporting in Germany, Head of the "Educational Reporting" project group at DIPF, Professor at the University of Erfurt.

German Institute for International Educational Research
Warschauer Str. 34-38
10243 Berlin
Germany
Tel.: +49 (0)30 29 33 60-14
E-Mail: doebert@bbf.dipf.de

Ueli Halbheer, lic. phil.
Research Fellow at the Department of Educational Psychology, Institute of Education, University of Zurich. Current research interests: educational standards and school development.

University of Zurich
Institute of Education
Freiestrasse 36
8032 Zurich
Switzerland
Tel.: +41 44 634 27 -53 or -61
www.didac.unizh.ch

Dr. Jan Hofmann
Director of the State Institute for School and Media Berlin-Brandenburg (LISUM), Brandenburg's representative in the steering group "Joint Projects of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder"; commissioned by the former project group "Innovations in the Education System" of the Bund-Länder Commission for Educational Planning and Research Promotion (BLK) with the organization of the 2007 OECD/CERI regional seminar of the German-speaking countries.

State Institute for School and Media Berlin-Brandenburg (LISUM)
14974 Ludwigsfelde-Struveshof
Germany
Tel.: +49 (0)3378 209-100
E-Mail: jan.hofmann@lisum.berlin-brandenburg.de

Prof. Dr. Olaf Köller
As Director of the Institute for Educational Progress, Prof. Dr. Olaf Köller is responsible for the development of item pools and their implementation. He was appointed Professor at the Humboldt University of Berlin in December 2004.

Institute for Educational Progress
Humboldt University of Berlin
Physical address: Luisenstraße 56, 10117 Berlin, Germany
Postal address: Unter den Linden 6, 10099 Berlin, Germany
Tel.: +49 (0)30 20 93-5335
Fax: +49 (0)30 20 93-5336
E-Mail: IQBoffice@iqb.hu-berlin.de
www.IQB.hu-berlin.de

Norbert Maritzen
Director of the Institute for Educational Monitoring (IfBM) of the Free and Hanseatic City of Hamburg

Institute for Educational Monitoring (IfBM)
Beltgens Garten 25
20537 Hamburg
Germany
Tel.: +49 (0)40 42 88 51-321
Fax: +49 (0)40 42 79 67-157
E-Mail: Norbert.Maritzen@bbs.hamburg.de
www.bildungsmonitoring.hamburg.de

Prof. Dr. Kurt Reusser
Professor of Education specializing in educational psychology and didactics at the University of Zurich. Main research interests: educational cognitive psychology and empirical educational research, teaching methods and video-based instructional research (mathematics, history), teacher training, teaching and learning with video.

University of Zurich
Institute of Education
Freiestrasse 36
8032 Zurich
Switzerland
Tel.: +41 44 634 27 -53 or -61
www.didac.unizh.ch

Mag. Dr. Werner Specht
Head of the Center for Educational Research and Evaluation at bifie (Graz), Adjunct Professor in School Quality Research at the University of Salzburg, previously a member of the Future Commission for the Reform of the School System.
Main research interests: quality development in education, governance of the school system, educational standards, integration and special education. Numerous publications in these areas.

bifie – Federal Institute for Educational Research,
Innovation and Development of the Education System
Center for Educational Research and Evaluation
Hans-Sachs-Gasse 3/II
8010 Graz
Austria
Tel.: +43 316 82 87 33-206
Fax: +43 316 82 87 33-6
E-Mail: w.specht@bifie.at

Prof. Dr. Petra Stanat
Professor of Empirical Research in Education at the Free University of Berlin, member of the Board of the Berlin-Brandenburg Institute for School Quality, member of the German PISA Consortium, member of the Editorial Board of Zeitschrift für Pädagogik, member of the German Council for Social and Economic Data.

Free University of Berlin
Department of Educational Science and Psychology
Empirical Research in Education
Habelschwerdter Allee 45
14195 Berlin
Germany
Tel.: +49 (0)30 838-55999 /-55981
E-Mail: petra.stanat@fu-berlin.de

Prof. Dr. Stefan C. Wolter
Director of the Swiss Coordination Centre for Research in Education (SKBF) in Aarau, Professor of the Economics of Education at the University of Berne, represents Switzerland in the Governing Board of CERI and the Education Policy Committee (OECD), head of the pilot project on educational reporting in Switzerland.

Swiss Coordination Centre for Research in Education (SKBF)
Entfelderstrasse 61
5000 Aarau
Switzerland
Tel.: +41 62 835 23 90
E-Mail: stefan.wolter@skbf-csre.ch