

No Child, No School, No State Left Behind: Schooling in the Age of Accountability¹

Stefan Thomas Hopmann²

Abstract: Why and under which conditions do international student assessment programs like PISA have success? How can the results of these assessments be useful for advocates of different, even contradictory, policies? What might explain different patterns of using assessment as a tool for school governance? Drawing on historical and comparative research and using PISA as an example, I provide a frame for discussing these and other questions around the international rise of accountability as a key tool of social change. The basic argument is that even though accountability is a global phenomenon, the ways and means of enacting and encountering accountability are not. How accountability is experienced depends on deeply engrained “constitutional mind-sets,” i.e., diverse cultures of conceptualizing the relation between the public and its institutions.

Keywords: accountability; curriculum; No Child Left Behind; school change; PISA; standards-based reform

Why is it that a comparative project like PISA can gain so much public attention in so many countries at the same time? What makes (some) governments tremble, parliaments discuss, journalists write, parents nervous, and teachers angry when PISA announces new results? Why are educational administrations and political committees eager to align their curriculum concepts to the one implicit in the PISA tests? Why is PISA in some places big news and in others news appropriate for a short notice on page 5 or in the education section? PISA is not the first project of this kind: What is different with PISA?

¹ This essay is a slightly revised version of an epilogue written for a book on PISA in comparative perspective (Hopmann et al. 2007). The response that the book received in a couple of countries—often aggressive in the case of the supporters of PISA, but otherwise very supportive—led me to believe that this issue might interest to readers who do not have much to do with PISA, but are concerned about the curricular impact of the current international trend toward “accountability.” I want to thank Gjert Langfeldt and Ian Westbury for the very helpful advice to both this version and the original chapter. Karin Lee Hansen of the University of Agder was a great assistance in completing the reference listing.

² In: Hopmann, S. T. (2008). No child, no school, no state left behind: schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417–456. Permission for re-publishing granted by Taylor & Francis and S.T. Hopmann.

Of course, there is no single explanation for PISA's mind-boggling success. PISA has obviously hit something in the public mind, or in the political mind, at least in Western European societies. This makes it "knowledge of most worth." It is unlikely that this success is a result only of the quality and scope of PISA itself. If one accepts at least some of the criticism concerning PISA's weak reliability and doubtful validity (see Hopmann et al., 2007), the opposite seems to be the case. What is of "most worth" in the eyes of the public is not the complicated and often overstretched research techniques or the specific design but the simple messages that front the public appearance of PISA: the league tables and the summaries that indicate what PISA sees as the weaknesses or strengths of different systems of schooling.

But even news of this kind has been around before—since the IEA began its comparative research in 1959—but it never had a similar impact. Thus, it is not enough to look at PISA itself to understand this story. It is necessary to understand, at the same time, how the social environment has changed: schooling, policies, and the public. The question is how PISA and its methodology fit into a larger frame of social transformation and thus could achieve the influence it now has. Moreover, one has to ask if there is one PISA achieving all this, or whether it is more appropriate to talk about the "multiple realities of PISA," i.e. the manifold ways in which PISA is enacted and experienced, which have been crucial for the success—and which the methodological mix utilized by PISA allows for.

In my view, the rise of PISA owes much to what I would call the emerging "age of accountability," i.e., a fundamental transformation ongoing in at least the Western world and which centers on how societies deal with welfare problems like security, health, resurrection, and education (see Hopmann, 2003, 2006, 2007). PISA fits this context in many different ways, depending on how accountability issues unfold in different societies. In some places the same fit is equally expressed by national policies like the "No Child Left Behind" legislation in the United States, or the development of national education standards, as has been the case in countries as diverse as Sweden, Germany, Switzerland, or New Zealand. What is important to note here is that the education sector is rather late in addressing such issues when compared to other areas such as health or security. In this perspective, PISA is but one example, a kind of collateral damage sparked off by the intrusion of accountability mechanisms into the social fabric of schooling.

To explain this observation, I will first outline how the transformation called “the age of accountability” can be understood. In the following section I will try to illuminate three basic modes of accountability, namely the strategies of “no child left behind,” “no school left behind,” and “no state left behind,” and how PISA fits into these different settings. In the last section, the implications of the multiple realities of accountability for current and future school development are discussed. In doing so, I rely on the results of the Norwegian research project, “Achieving School Accountability in Practice” (ASAP) (Langfeldt et al., 2008). Other sources are comparative projects that I have been involved with: “Organizing Curriculum Change” (OCC), including research projects in Finland, Norway, Switzerland, Germany, and the United States (see e.g., Künzli & Hopmann, 1998), and the dialogue project “Didaktik meets Curriculum,” which involved scholars from about 20 countries throughout the 1990s (see e.g., Gundem & Hopmann, 2002; Westbury et al., 2000), and, last but not least, the research done in connection to the international debate on “PISA According to PISA” (see Hopmann et al., 2007).

The Age of Accountability

Social scientists, economists, politicians, educators, and the public seem to agree that something fundamental is going on, something that changes, at least in principle, the social fabric of Western societies. However, they differ widely in what they see as the core of this transition. To name but a few more recent examples:

- The modern world system is, according to Immanuel Wallerstein (2004), characterized by the ever-expanding commodification (in Marx’s terms, “Verdinglichung”) of all natural resources, human relations, labor, knowledge, etc., forcing a lasting division of labor in and between nations and turning ordinary citizens into alienated tools of a globalized economy.
- Neoinstitutionalists like John W. Meyer also speak about globalization, but in organizational or structural terms, defining the current transition process as an outcome of the rapidly growing influx of international “institutions,” i.e., common ways of seeing and dealing with society, as provided by international organizations (such as the UN, the World Bank,

- and OECD) and the emerging “world polity” that supersedes national histories and policies alike (see e.g., Meyer, 2005).
- Theories of “reflexive modernity,” as provided by Anthony Giddens and others, would agree that the change is global, however, pinpoint the special implications for the members of society, e.g., the need to develop a reflexive stance toward the structures of society and the embedded risks for society as a whole as well as for the individual (see e.g., Beck, 2007; Beck et al., 2007).
 - Governmentality theories, drawing on Foucault’s famous 1977/78 Collège de France lectures (Foucault, 2006), point similarly to the impact the transition has for the state and its institutions as well as for the public and its members. But they see a growing transfer and diffusion of power relations into self-control mechanisms, making citizens internalize the (more or less alienated) mentality necessary to govern them(selves) (see e.g., Bröckling et al., 2000; Gottweis et al., 2004; Lange & Schimank, 2004).
 - New Public Management (NPM) theories do not disagree that there is a diffusion of power and a change of habits, but they rather see it as a positive force, making societies and their institutions and members more effective in a globalized world as customer-centered management and control techniques are introduced (see e.g., Buschor & Schedler, 1994; Pollit & Bouckaert, 2004).
 - Discussions on the welfare state explore similar issues, but as a question of how the modern “intervention state” is forced to dismantle its traditional comprehensive strategies governing resources, the law, and the social sphere in a more and more post-national world, and how welfare is re-modeled within this “unravelling” of the state and its institutions (see e.g., Esping-Andersen, 1996; Leibfried & Zürn, 2006; Scharpf & Schmidt, 2000a, b).
 - Finally, systems theories based on the work of Niklas Luhmann (see e.g., Luhmann, 1998) argue that the current transition grows from within, from the need of social systems to deal with an ever-growing complexity and contingency that forces a reflexive re-design of the ways and means of the social communication that constitutes the fabric of social systems (see e.g., Åkerstrøm Andersen, 2003; Rasmussen, 2004).

Of course, this is but a brief summary of the staggering number of transition theories that have flourished, despite the obviously prematurely proclaimed “end of history” (Fukuyama, 1992). Moreover, these approaches vary widely.

Some of them see the contemporary change as a late consequence of processes started with the invention of the modern state (e.g., Foucault; Wallerstein), whereas others point to more recent changes—to, for instance, the crisis of the welfare state or the rapid globalization process (e.g., Leibfried & Zürn; Meyer). Some see it as primarily a top-down process by which global developments overpower local traditions (e.g., Meyer; Wallerstein); others stress the role of intermediate levels such as the nation-state and its institutions (e.g., Giddens, Leibfried & Zürn; Pollit & Bouckaert); others see the main issue at the level of the impact of the transition process on those involved (e.g., Beck; Foucault). Some theories stress institutional patterns or social systems as the prime force (e.g., Meyer; Luhmann); others believe in actors and their policies as defining elements (e.g., Pollit & Bouckaert; Wallerstein); some try to sketch a third perspective in which actors and structures are seen as inextricably intertwined (e.g., Giddens; Foucault).

One should not complain about this amazing diversity of approaches. It is an expression of the difficulty in finding common ground at a time when the transition is still unfolding with growing but uneven speed in different places. Additionally, many of these authors and their followers use a similar pool of examples in spite of their differences. Even though they do not agree on all the why questions, they point to much of the same kind of evidence, as, for instance, in examples of:

- The redistribution of resources, risks, and responsibilities within and across societies;
- The destabilization, or at least the restructuring, of most public institutions and their relations to or competition with the private sector;
- The re-tooling of legitimation and control patterns within the public as well as the private sphere and its impact;
- The pressure on systems and actors toward taking a reflexive stance toward themselves and taking responsibility for their own “well-being.”

Accountability

When we look at the narrower issue of “accountability,” a similar wealth of models and approaches emerges. In addition to more or less implicit accountability concepts within the general transition theories mentioned

above (mostly constructed as “being made responsible” in one or another way, e.g., by Giddens, Foucault, or the NPM theories), different models of accountability have emerged based on the areas in which accountability is observed:

- In economic theories (see e.g., Laffont, 2003), where the concept originated, accountability is often constructed as a means by which a principal (the resource-giver) under conditions of limited information tries to multiply the ends by giving the agent (the resource-taker) incentives and/or forcing him by other means to account for the efficiency, quality, and results of his deliveries.
- In research on public administration (see e.g., Hood, 1991, 1995; Hood et al., 2004), accountability is often seen as a key tool of the New Public Management movement used to ensure that units and persons provide services according to the goals set for them or agreed with them. According to this approach, it unfolds as a combination of risk management and blame avoidance, by which those who are held accountable try to limit the scope of possible failure.
- In research on social policy, the same phenomenon is described as a “quasi-market revolution” (see e.g., Bartlett et al., 1998), i.e., as the intrusion of market-like mechanisms of distribution and control into the public sector: elements of competition, contractualization, and auditing are introduced into the service-rendering, often, as Hood (2004) has pointed out, in the form of a “double whammy,” i.e., as the co-existence of the traditional bureaucratic modes with the administrative tool kit fostered by NPM.
- Some educational and health-care researchers see the rise of ‘the age of accountability’ as a “revolutionary” move toward “evidence-based” practice, i.e., the growing expectation that professionals can present data to prove that they have performed professionally and efficiently (see e.g., Slavin, 2007; for health care, see e.g., Muir Grey, 1997).
- Generalized beyond the realms of public service, this leads to the concept of the emergence of an “audit society” (e.g., Power, 1997), the assumption that more and more areas of social life are being made “verifiable,” i.e., subjugated to regimes of counting what can be counted, and thus part of a measurable accountability.

- Interaction and transactions theories construe the personal costs of such transition, looking at accountability as an interpersonal relation in which we deal with “accounts,” “excuses,” and “apologies,” i.e., strategies to explain ourselves in ways that give a sustainable account of our efforts (e.g., Benoit, 1995).
- Finally, psychological approaches to accountability (e.g., Sedikides et al., 2002) look at the personal ways of dealing with accountability, how one develops mechanisms to attribute or to reject accountability embedded in roles and functions we have to perform.

Like the transition theories, accountability theories provide a wide array of possible causes and implications. Some see this process primarily as an effect of a growing “economization” of all parts of the society (e.g., social policy and audit theories); others see accountability as inevitably embedded in the social fabric of modern societies (e.g., the psychological explanations). Some see accountability primarily as a politically initiated restructuring effort (e.g., quasi-market theories); others see accountability simply as a legitimate means to ensure that customers or clients get what they have paid for (economic and NPM theories). Additionally, accountability is viewed on rather different levels. Utilizing a model developed by Dubnick (2005), we can discern:

- A first-order accountability, i.e., an accountability arising in face-to-face relations (as described by psychological models);
- A second-order accountability characterized by how well one follows the rules and standards set by a resource-giver (as described by public administration theories);
- A third-order accountability seen as “managerial accountability,” i.e., the use of accountability by a principal as a means to achieve better service and effectiveness of the agent. Finally,
- A fourth-order accountability based on the assumption that the one held accountable internalizes the norms, values, and expectations of the stakeholders and puts himself or herself into action (as pointed out by, e.g., theories of governmentality or of professionalism).

In practice, all of these can be intertwined. However, the dividing line is how it is assumed that the interaction comes into being and which levels rule

when compared to other levels (if they are not seen, as they are in economic theories, as an embedded rationale of social actors at all levels).

In addition, accountability concepts change over time and are different in different places. A good indicator of this is that (1) there is no common translation of the concept of accountability available in most non-English-speaking countries, and (2) public agencies or policymakers do not employ similar definitions of the elements and limits of accountability when “accounting for accountability” (see Birkeland, 2008; Dubnick, 2005). Nevertheless, most accountability analysts agree with the above-mentioned transition theories on some core issues, namely:

- That accountability procedures more and more permeate at least Western societies and thereby change the ways and means by which societies deal with themselves;
- That the rapid rise of accountability affects all areas of the public sector, from education to health, and their relation to the private sector;
- That this transition enforces a vast redistribution of resources and responsibilities and thereby a fundamental change in the interplay between resource providers and users, often described as a kind of implicit (values, norms) or explicit (standards, contracts) rearrangement of what is supposed to shape their relations.
- That this process unfolds at different speeds and with different patterns, depending on what kind of social setting they become a part of.

For my purpose herein, it is not necessary to decide which of these theories and models carries the most theoretical or empirical evidence. Rather, the common features shared by most of them offer enough of a starting point, even though this puts some of the “why questions” aside temporarily and moves the focus to the question of how the emergence of the age of accountability can be observed in action. In my view, its common core can be described as a slow but steady transition from what I call “management of placements” (*Verortung*) toward a “management of expectations” (*Vermessung*), by which the ways and means of dealing with “ill-defined” problems, such as health, education, security, and resurrection, are changed fundamentally (see Hopmann, 2000, 2003, 2006, 2007).

Managing Transition

Following in the footsteps of Max Weber (see Breuer, 1991), we can describe the rise of the modern state as the successive unfolding of a management of “placements” by which the risks of being born—how to get an education; who takes care of me when ill or old; who gives me security in my everyday life and my dealings; how to be at peace with myself and my neighbors—are taken care of by institutions run by professionals with a specific education on how to deal with such ill-defined problems. Such institutions (i.e., schools, hospitals, prisons, armies, bureaucracies, and churches) had a comprehensive mission in that their professionals needed leeway to define which of these problems required what kind of treatment. The institutionalized problem-sharing allowed for taking on more risks and moving beyond the care for immediate needs. Of course, which problems were considered ill- or well-defined changed over time, as did the resources available. But the internal distribution of resources and the evaluation of outcomes were mostly left to the professionals themselves, or the emerging professional communities, who defined and controlled the education, licensing, and practice of their members (see Abbott, 1988; Hopmann, 2003).

This comprehensive institutionalization had no fixed boundaries (which would have required the transformation from ill-defined into well-defined problems), thus opening a continuing process of the broadening of the scope and differentiation of means whenever new aspects of the problems seemed to become urgent. Thus, each and every field underwent a massive expansion, multiplying its tasks and treatments. In the past, for example, a couple of years in schools (and for a few, in universities) was all the public education available. Today, we spend 20 or more years of our lives in all kinds of professionalized educational settings, from child care to elder hostels. When once we met a doctor at the beginning and at the end of life, and maybe a few other times under extraordinary circumstances, today we spend a lot of time in each and every year with medical doctors, nurses, and other health professionals in waiting rooms, in treatment, or in emergency rooms. In short, the management of placement was extremely successful, so successful that Western societies now spend most of their public budgets on dealing with these problems. As long as the differentiation of the institutions did

not outspend the resources available, differentiation could go on and on, and with an ever-increasing speed.

This success story seems to come to an end in what social-policy theorists' term "the crisis of the welfare state," i.e., as resource limits and boundaries for further expansion become more and more visible (wherever they stem from). The legitimacy of the whole placement strategy relied on its ability to cover new, ill-defined problems by expansion and sophistication, but there is now mistrust and anxiety about whether this comprehensive help will be sustainable in the future. A very visible impact of this loss of trust is the rise of welfare patriotism on the right and the left in almost all Western societies, articulating, and maybe misusing, much of the unease citizens feel about the future and security of the inherited places and treatments (our welfare is said to be at risk because of immigrants, globalization, outsourcing, etc.).

One of the important responses to this is a stepwise transition from a management of placements toward a management of expectations. Instead of guaranteeing comprehensive institutions, there is an attempt to transform ill-defined problems into better-defined expectations as to what can be achieved with given resources. Standards, benchmarks, indicator-based budgets, etc. are examples of how this transition is managed. In that they do not necessarily imply long-term commitments, expectations can remain transient and volatile to changes in the social fabric of the expectations mix. This allows for more target-oriented management and accountability that, however, comes at a price: Whatever does not fit into an expectation regime becomes marginalized. Comprehensive coverage is replaced by a fragmented system of treatments available under certain conditions. The leftover, and not least the still ill-defined general issues—what does it mean to be well-educated, healthy, secure, feel well, etc.?—is either still connected to the former placements and/or transformed into temporary programs seemingly better equipped for addressing the remains immediately ("the patient in focus," "fighting crime," "strengthening social education," etc.).

Take the example of schooling: In earlier times public education was provided by 'a place called school' (Goodlad, 1983) run by professionals called teachers who decided within sketchy limits, based on professional and local traditions, how to teach and what achievement seemed to be sufficient. There was no external public evaluation of the quality of the services provided, except for extraordinary cases of failure, if the normal procedures seemed to

be professionally acceptable. Accordingly, good instruction was not defined primarily by its measurable outcomes, but rather by the professional judgment of the adequacy of what was done. Expectation management changes the picture dramatically. The core focus shifts to more or less well-defined expectations of what has to be achieved by whom. Good instruction is the one overlapping expectation and can be provided outside the traditional institutions and professions; in fact, everybody is welcome to provide it as long as the expectations are met. Of course, there are issues that are not (yet) covered by identifiable expectations; however, in case of conflicting goals, the balance will always tip toward those expectations that are well-defined enough to become part of the implied accountability of the treatment providers. The rest, that which is not addressed but seems to need to be taken into account (e.g., issues such as mobbing/bullying, gender, migration, etc.), is embedded into transient intervention programs of limited scope, sufficient to ensure the public that no ill-defined problem is left behind.

It is important to remember that this is expectation management and is not about outputs, outcomes, or “efficiency” as such—as, for instance, NPM theories see it. That is, only those results that can be “verified” according to the stakes given and do not meet expectations become problematic, and only those outcomes that meet the predefined criteria are considered a success. In fact, any caretaker of an ill-defined problem will always produce many more effects than any accountability system can observe and measure. Some of them may be simply by-products or minor collateral damage, but some impacts may be major contributions (e.g., inclusion into society, regulating biographies, etc.), which is beyond the short-sighted reach of the management of expectations. The line is drawn by the ever-changing fabric of expectations on the one hand and, on the other, by the simple fact that accountability needs something that can be counted, or where it is at least possible to measure the distance between expectations and results (see Slavin, 2007).

The emergence and spreading of accountability is a signifying hallmark of the whole transition process. PISA fits nicely into this transition, as we will see in the following sections. Seen as part of a management of placements PISA would be a disaster: it covers only a few aspects of the place, of schooling, and of the curriculum, and even these are covered in ways that account at best very indirectly. However, as a tool of expectation management, PISA fosters a transformation of what had been ill-defined issues (e.g., curriculum contents)

into seemingly well-defined attainment goals. It delivers, at the same time, a parameter for holding schooling accountable—for delivering according to the expectations embedded into its questionnaires. It contributes to the fragmentation of the field by transforming the conditions and constraints of this delivery into independent factors (e.g., social background, gender, migration, etc.) whose impact has to be minimized by way of teaching if expectations are to be fully met. The best representation of this is given by the “production functions” by which PISA-using economists calculate the transaction costs of schooling and the ways and means by which the principals (parents, the state) might maximize the effectiveness of the chosen agents (i.e., teachers, schools, or school systems; see Bishop & Woessmann, 2004; Micklewright & Schnepf, 2004, 2006; Sutherland & Price, 2007; Woessmann & Fuchs, 2004).

Constitutional Mindsets

When it comes to the public sphere, the transition from a management of placements toward a management of expectations meets different constitutional mindsets, i.e., deeply ingrained ways of understanding the relation between the public and its institutions.³ For example, the US Constitution is constructed as a protection of the individual against the misuse of power by governments and others. It sees the rights of the individuals as a given and the intervention of government as limited by these rights and protects citizens against any infringements of their constitutional freedoms. The First Amendment, for instance, states:

Congress shall make no law respecting the establishment of religion, prohibiting the free exercise thereof; or abridging the freedom of speech or the press; or the right of the people peaceably to assemble, and to petition the government for the redress of grievances.

Within the Prussian or the Austrian tradition, which comes from the opposite direction, civil rights are something constituted and limited by the law, i.e., it is the state and its (more or less enlightened) institutions that create and define the boundaries of social and individual life. Religious freedom, for instance,

³ For the basics of the following, see e.g., Haft and Hopmann (1990), Hopmann and Wulff (1993), and Zweigert and Kötz (1998).

may be granted, but the freedom is closely tied to state supervision of its organizations and institutions, and the state can set limits for the conditions for the “full” exercise of a religion (which creates problems for religions stemming from non-institutionalized traditions such as Islam). The Scandinavian constitutional tradition settled (at least in its beginnings) somewhere between these fundamentally opposed starting points: it acknowledges the right of the state to impose a constitution but limits its reach, making it subsidiary to local and regional legal traditions. That has changed gradually, but there is still no unified code of law, rather a pragmatic approach to regulating fields of interest based on practical experience and home-grown traditions. The local constituency is still seen as the core of the social fabric. Citizens are empowered to define a community life based on their own traditions within a broad constitutional setting. Thus, while there are state churches in Norway and Denmark, there has been plenty of leeway to establish new local traditions (e.g. as “free churches”); today any group of a certain size and permanence, and with a discernible creed, can establish itself as a “church,” with a right to receive state subsidies.

Of course, actual constitutional and legal structures are more mixed, the patterns much more blurred, than these different starting points indicate. But the mindsets on which they are founded seem to be alive and well and have a strong impact on how the public and its institutions conceptualize the legal and structural implications of social change. This is clearly the case when it comes to how accountability measures are embedded into the public system as a whole, and especially into the school system. There the main questions are: Is accountability about protecting the individual citizen (student) against poor service rendering? Or is the primary goal to strengthen the ability of local communities to run their institutions according to their own needs and aspirations? Or is it about holding the public system accountable for its contribution to the state’s welfare? Accordingly, put in the current educational context, one has to ask where the main focus of accountability is situated: at “no child left behind,” “no school left behind,” or “no state left behind”?

The Multiple Realities of Accountability

Seen in a historical perspective, PISA and the like are heavily indebted to the legacy of the assessment movement in the United States and its internationalization by the International Association for the Evaluation of Educational

Achievement (IEA), beginning in the late 1950s. Seminal works such as Caswell's *City School Surveys* (1929), the Eight-Year Study (1942), or Bloom's (1956) groundbreaking taxonomy of educational objectives, the national spread of the Scholastic Aptitude Test from the 1950s onwards, and the establishment of the National Assessment of Education Progress (NAEP) paved the way for an understanding in which student achievement was seen as the prime indicator of the quality of schooling. The rapid rise of assessment and evaluation as key tools of educational control was fueled by a constant flow of critical works on the poor state of the nation's schools. From Conant's (1959) report, *The American High School*, Rickover's (1963) *American Education: A National Failure*, Coleman's *Equality of Educational Opportunity* (Coleman et al. 1966), to *A Nation at Risk* (National Commission on Excellence in Education 1983) and the *Nation's Report Card* (Alexander et al., 1987), the basic tenor was the same: The US system is failing many of its students, as demonstrated by the results of local, state-wide, and national testing.

From the late 1980s onwards, this seemingly constant failing led to a more generalized approach to assessment, testing, and "reform"—"standards-based reform" (Ahearn, 2000; Fuhrman, 2001). State after state introduced state standards for the curriculum and—if it had not yet done so—statewide assessment of student achievement to ensure that these standards were met. It would be a wild exaggeration to pretend that this approach was an immediate success. In some cases, the introduction of state standards obviously spelled out disaster⁴; in others, modest gains at best could be reported, although their validity was, and is, heavily disputed.⁵ However, despite some 50 years of mixed experience with assessment and rather shallow results⁶ the next move was to introduce national legislation aiming at a unified approach to assessment and accountability, the "No Child Left Behind Act" of 2001, enacted under the Bush presidency and supported by an almost united Congress (Peterson & West, 2003).

⁴ See e.g., the Kentucky experience (Whitford and Jones 2000).

⁵ See e.g., Amrein and Berliner (2002), Braun (2004), Cannell (1988), Dorn (1998), Haney (2000, 2002), Ladd and Walsh (2002), Linn (2000), Saunders (1999), Swanson and Stevenson (2002), Watson and Suppovitz (2001).

⁶ See Cook (1997), Herman and Haertel (2005), McNeil (2000), and Mehrens (1998).

No Child Left Behind (NCLB)

It is worthwhile to look closely at the provisions of the NCLB in that they are paradigmatic for the way accountability is constructed within the US tradition. The comprehensive “statement of purpose” unfolds a wide array of issues:

The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments.

This purpose can be accomplished by:

- Ensuring that high-quality academic assessments, accountability systems, teacher preparation and training, curriculum, and instructional materials are aligned with challenging State academic standards so that students, teachers, parents, and administrators can measure progress against common expectations for student academic achievement; meeting the educational needs of low-achieving children in our nation’s highest-poverty schools, limited English proficient children, migratory children, children with disabilities, Indian children, neglected or delinquent children, and young children in need of reading assistance;
- Closing the achievement gap between high- and low-performing children, especially the achievement gaps between minority and non-minority students and between disadvantaged children and their more advantaged peers;
- Holding schools, local educational agencies, and states accountable for improving the academic achievement of all students, and identifying and turning around low-performing schools that have failed to provide a high-quality education to their students, while providing alternatives to students in such schools to enable the students to receive a high-quality education;
- Distributing and targeting resources sufficiently to make a difference to local educational agencies and schools where needs are greatest;
- Improving and strengthening accountability, teaching, and learning by using state assessment systems designed to ensure that students are meeting challenging state academic achievement and content standards and increasing achievement overall, but especially for the disadvantaged;

- Providing greater decision-making authority and flexibility to schools and teachers in exchange for greater responsibility for student performance;
- Providing children an enriched and accelerated educational program, including the use of school-wide programs or additional services that increase the amount and quality of instructional time;
- Promoting school-wide reform and ensuring the access of children to effective, scientifically based instructional strategies and challenging academic content;
- Significantly elevating the quality of instruction by providing staff in participating schools with substantial opportunities for professional development;
- Coordinating services under all parts of this title with each other, with other educational services, and, to the extent feasible, with other agencies providing services to youth, children, and families;
- Affording parents substantial and meaningful opportunities to participate in the education of their children (Section 1001).

However, when it comes to which goals are in focus and how accountability is supposed to foster these goals, this complexity is right away reduced to more specific expectations. Academic standards are, according to NCLB, the following:

Standards under this paragraph shall include:

- (i) challenging academic content standards in academic subjects that:
 - (I) specify what children are expected to know and be able to do;
 - (II) contain coherent and rigorous content; and
 - (III) encourage the teaching of advanced skills; and
- (ii) challenging student academic achievement standards that:
 - (I) are aligned with the state's academic content standards;
 - (II) describe two levels of high achievement (proficient and advanced) that determine how well children are mastering the material in the state academic content standards; and
 - (III) describe a third level of achievement (basic) to provide complete information about the progress of the lower-achieving children toward mastering the proficient and advanced levels of achievement (Section 1111).

Accountability is then based on these standards:

Each state plan shall demonstrate that the state has developed and is implementing a single, statewide state accountability system that will be effective in ensuring that all local educational agencies, public elementary schools, and public secondary schools make adequate yearly progress as defined under this paragraph. Each state accountability system shall:

- (i) be based on the academic standards and academic assessments ... and shall take into account the achievement of all public elementary school and secondary school students;
- (ii) be the same accountability system the State uses for all public elementary schools and secondary schools or all local educational agencies in the State...; and
- (iii) include sanctions and rewards, such as bonuses and recognition, the State will use to hold local educational agencies and public elementary schools and secondary schools accountable for student achievement and for ensuring that they make adequate yearly progress (Section 1111)

Finally, what is meant by “Adequate Yearly Progress” is defined in the subsequent paragraph:

- (B) Adequate yearly progress. Each state plan shall demonstrate, based on academic assessments described in paragraph (3) and in accordance with this paragraph, what constitutes adequate yearly progress of the state and of all public elementary schools, secondary schools, and local educational agencies in the state toward enabling all public elementary school and secondary school students to meet the state’s student academic achievement standards while working toward the goal of narrowing the achievement gaps in the state, local educational agencies, and schools.
- (C) Definition. “Adequate yearly progress” shall be defined by the state in a manner that:
 - (i) applies the same high standards of academic achievement to all public elementary school and secondary school students in the State;
 - (ii) is statistically valid and reliable;
 - (iii) results in continuous and substantial academic improvement for all students;

- (iv) measures the progress of public elementary schools, secondary schools, and local educational agencies and the state based primarily on the academic assessments described in paragraph (3);
- (v) includes separate measurable annual objectives for continuous and substantial improvement for each of the following:
 - (I) The achievement of all public elementary school and secondary school students.
 - (II) The achievement of:
 - (aa) economically disadvantaged students;
 - (bb) students from major racial and ethnic groups;
 - (cc) students with disabilities; and
 - (dd) students with limited English proficiency;
 except that disaggregation of data under subclause (II) shall not be required in a case in which the number of students in a category is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student (Section 1111).

I have quoted the text of the NCLB Act in such length because it provides a concise definition of what management of expectations in this perspective is about. The core of accountability is narrowly focused on student achievement measured by “academic standards.” Other functions of schooling (such as the role school plays for local communities or in shaping society) are hardly mentioned. At the same time, academic achievement is reduced to that which can be reported as “statistically valid and reliable,” leaving out any educational or social achievements that cannot be counted as required. Within this frame, responsibility is passed from the federal top, through intermediate levels such as state and district administrations, to teachers and local school leaders. They are expected to improve the test results by “evidence-based teaching” or even by “data-driven decision making” (see Education Commission of the States 2002, Marsh et al., 2006), which collapses the complexities of classroom work or school leadership into single-minded frame-sets of statistically significant achievement gains (see the comments by e.g., Berliner, 2005; Hargreaves, 2006; Ingersoll, 2003; Koretz, 2002). The starting idea of the “basic principles of curriculum and instruction” (Tyler, 1949), which was embedded in the methodologically

much broader approach of, e.g. the Eight-Year Study (Aikin, 1942) and which asked for a comprehensive understanding of schooling as a social and local institution, has, as it seems, dwindled to a concept of measurable yearly progress.

The response to NCLB within the education community has been almost evenly divided. While a majority of politicians and economists, and certain parts of the public, seem to support NCLB wholeheartedly, or at least its core concept of accountability, many educators are less enthusiastic. The public response seems to be fragmented based on social class, level of education, and political orientation (see Loveless, 2006). The professional reactions have much to do with it, or if not, one accepts the narrow focus of NCLB as reasonable. On the one side are those who see NCLB at least as a starting point for a possible school revolution, finally solving the school “crisis” in the United States (see e.g., Irons & Harris, 2007; Ladd & Walsh, 2002; Peterson & West, 2003). Some economists have even begun to calculate the economic spin-off of NCLB if modest gains can be sustained (Hanushek, 2002a, 2003, 2006; Hanushek & Raymond, 2003, 2005). Others are skeptical, pointing to the “impoverished” scope of the provisions (see e.g., Berliner, 2005) or the obvious implementation problems of the current approach.⁷ The construct of “adequate yearly progress” (AYP) in particular has created a tremendous challenge. The expectation of the progress that could be reached went far beyond the reality of slow and unstable school change—and adding minority criteria worsened the situation. Some critics fear that almost all US schools will end up on the watchlists of failing schools.⁸ And, as many researchers and practitioners have pointed out, NCLB will fail while schools and their leadership do not have the required “capacities,” i.e., the ability to identify their local mix of problems and to deal with these professionally (see e.g., Elmore, 2004). However, what is almost never challenged in this debate is the basic assumption of the whole enterprise, namely, that it is the student’s academic achievement that best reflects the quality of schooling and that it

⁷ For a variety of such problems, see e.g., Apple (2007), Birman et al. (2007), Chubb (2005), Deretchin and Craig (2007), Eberts et al. (2002), Gorard (2006), Le Floch et al. (2007), Martineau (2006), Mintrop (2003), and Zimmer et al. (2007).

⁸ See Herman and Haertel (2005), Linn (2005), Linn and Haug (2002), and Linn et al. (2002, 2005).

is the poor quality of instruction provided by poor teaching that is to blame for the fact that children are left behind. Holding states, school districts, and schools accountable is reduced to the requirement to do whatever is necessary to pass this accountability along to classrooms and teachers and, finally, to students themselves.

Similar criticism has been voiced about PISA's role in the United States (e.g., Bracey, 2005). However, even though it uses a similar approach to mapping school achievement, and even though the United States has been one of the driving forces behind it, PISA does not have much of a US audience in the shadow of NCLB and no significant impact on the wider public or the educational science community.⁹ Arguably, this reflects NCLB's status as a national law, whereas PISA is an international enterprise with no direct obligations for the states and the schools participating.

In general, PISA seems to have less impact where a national achievement control of some kind is already in place (as e.g., in Sweden), which bodes ill for the future of PISA as more and more countries introduce accountability systems of their own making. However, this would not explain why earlier studies like TIMSS have had more visibility in the United States. In my view, this shortcoming of PISA has much to do with the key fallacy of its design, namely to present itself mainly as a cross-national comparison, even though it shares with NCLB the same starting point, student achievement, which only to a very small degree (not least in the case of the United States with its manifold states) can be attributed to a specific "national" fabric of schooling (Hopmann et al., 2007). As a cross-national comparison, PISA is not much of an eye-opener for the US public; it only confirms what was known from earlier international studies, i.e., that US students don't do very well in such tests compared to the students of many other nations. In that the "winner" of PISA 2000 and 2003 was tiny Finland, and not an

⁹ For instance, a recent search of ERIC brought up fewer than 150 papers and books about PISA, most of them from outside the United States—nothing compared to the general issue of accountability (more than 18,000 hits) or NCLB (about 2000 hits). It does not even match the impact of TIMSS (with more than 400 hits) and is far below what the German equivalent of ERIC, the FIS Bildung, reports on PISA from Germany (more than 2,500 hits).

international competitor like Japan (which succeeded in TIMSS and did well on PISA too), there is, seemingly, nothing much for Americans to learn from successful PISA nations.

No School Left Behind

When the accountability wave hit the Nordic shores for the first time, the spontaneous reaction of the political and educational establishments was almost opposite to what had happened in the United States. While their accountability became a tool to centralize important elements of educational control, first at the state and later at the national level, the spontaneous reaction of the Scandinavians was decentralization. Although government offices and administrative departments (in Norway and Sweden) were created to satisfy the discourse of the New Public Management, and many national reports and white papers were commissioned on public service rendering and administration, none of this—except maybe for Finland, which was under much more economic strain following the breakdown of the Berlin Wall (Sinola, 2005; Uljens, 2007; Yrjölä, 2005)—led to a sustained and comprehensive accountability reform of the US kind (see Bogason, 1996; Pollit & Bouckaert, 2004; Prahl & Olsen, 1997). The concurrent debate about joining or not joining the EU may have had a share in this decision-making in that EU participation was often framed in terms of the risk of more centralization (Karlsen, 1994). However, tackling challenges by way of an issue-focused and pragmatic step-by-step approach with special regard for how lower levels of government, such as districts and municipalities, could deal with any emerging tool kit was consistent with what I am calling the Nordic constitutional mindset.

Until the 1990s, the main tool used to govern the school curriculum was curriculum guidelines, developed mainly by the state administrations by way of committees largely consisting of experienced teachers and subject-matter specialists (see Bachmann et al., 2004; Gundem, 1992, 1993, 1997; Sivesind et al., 2003). Local schools and teachers had considerable leeway to pick and choose within this curriculum frame in order to develop locally adapted teaching programs. There was no regular state-run evaluation of the outcomes of teaching, and, indeed, outside research even the concept was not familiar (Hopmann, 2003). In a Nordic perspective, schools were

seen as places run by highly educated and esteemed teachers who knew best how to do their jobs.¹⁰ Curriculum change was primarily seen as a matter of dialogue between local experience and national needs; changes were typically introduced by way of lengthy trials and with an often extraordinary involvement of all levels of schooling and administration. Of course, this was by no means a paradise of peaceful change: each and every curriculum reform has had its proponents and opponents, and the interplay between the school sector, research, politics, and the public was at times pretty contentious (Sivesind forthcoming). However, this played out within the context of school systems that enjoyed, for most of the time, broad support at all levels of society.

In this context it was no surprise that the first reaction to sharp national and international criticism of schooling was a redoing of what had been successful. In the case of Norway, for instance, the first contemporary criticism of the school system was voiced by an OECD panel in 1988 and by a national committee commissioned by the parliament (NOU 1988: 22). Reflecting the emerging NPM discourse, both concluded that the weaknesses of the national school system were, significantly, an outcome of an underperforming school governance structure that was not able to ensure that the goals of curriculum guidelines were being reached. Two conclusions were drawn: On the one side a sweeping reform of the whole school curriculum was launched, beginning with a new general curriculum frame, followed by new comprehensive guidelines for the upper secondary sector and the elementary and lower secondary schools. The frame stressed the double purpose of schooling as caretaker of the national and local heritage and as knowledge promoter (STM 29 1994/95). The subsequent curriculum guidelines received a new structure: they were to focus on the most important requirements and state these expectations in terms of goals (a kind of management-by-objectives approach) that could be reached by average schools. On the other side, a remake of the governance structure was inaugurated, constructing a double-faced reform combining a

¹⁰ As a background for understanding the Nordic education sector, one has to know that schools and their teachers played a pivotal role in the nation-building processes across the region and in the shaping of national identities (see e.g., Korsgaard 2004, Slagstad 1998, Telhaug 2005, Telhaug and Mediås 2003, Werler 2004). Moreover, schools are not only seen as places for the young but also as the cultural core of the local community, which turns the local and regional distribution of schooling into always-contested issues.

refocusing of national steering while stressing the importance of local autonomy and responsibility for reaching these goals (see Kirk-, Utdannings- og Forskningsdepartementet (KUF) 1997, STM 37 1990/91, STM 47 1995/96). The reform was supported by numerous in-service and research programs to help districts, municipalities, schools, and teachers identify the major obstacles and prepare for the enactment of the new guidelines. This new orientation was complemented by initiatives to develop school-based and peer-guided school improvement.¹¹ However, this first take-up of NPM-like measures infuriated many educators, politicians, and practitioners alike; these critics felt that the toolkit of accountability was an “instrumentalist mistake” that did not fit the national traditions of schooling and challenged the former strategy of placement, i.e., the compulsory comprehensive school (see e.g., Hovednak, 2000; Koritzinsky, 2000; Lindblad et al., 2003).

When a new liberal-conservative government felt that these first steps of reform were still not enough to ensure adequate school development, it commissioned a new national report to recommend additional measures. What emerged was a peculiar understanding of school development as the development of “quality,” in which “quality” represents a rather vague and all-encompassing understanding of whatever might affect the outcomes of schooling (see Birkeland, 2008, Sivesind forthcoming; STM 30, 2007). The then-secretary of education, Kristin Clemet, expressed the basic rationale of this approach as follows:

Society’s reasons for having schools, and the community tasks imposed on them, are still relevant today: Education is an institution that binds us together. We all share it. It has its roots in the past and is meant to equip us for the future. It transfers knowledge, culture and values from one generation to the next. It promotes social mobility and ensures the creation of values and welfare for all. For the individual, education is to contribute to cultural and moral growth, mastering social skills, and learning self-sufficiency. It passes on values and imparts knowledge and tools that allow everyone to make full use of their abilities and realize their talents. It is meant to cultivate and educate, so that individuals can accept personal responsibility for themselves and their fellows. Education must make it possible for pupils to develop so that they can make well-founded decisions and influence their own futures. At the same time, schools must change when society changes. New

¹¹ See e.g., Ålvik (1994), Granheim et al. (1990), Haug and Monsen (2002), Karlsen (1993), KUF (1994), and Nesje and Hopmann (2003).

knowledge and understanding, new surroundings and new challenges influence schools and the way they carry out the tasks they have been given. Schools must also prepare pupils for looking beyond the Norwegian frontiers and being part of a larger, international community.

We must nourish and further develop the best aspects of Norwegian schools and at the same time make them better equipped for meeting the challenges of the knowledge society. Our vision is to create a better culture for learning. If we are to succeed, we must be more able and willing to learn. Schools themselves must be learning organizations. Only then can they offer attractive jobs and stimulate pupils' curiosity and motivation for learning. We will equip schools to meet a greater diversity amongst pupils and parents/guardians. Schools are already ideals for the rest of society in the way they include everybody. However, in the future we must increasingly appreciate variety and deal with differences. Schools must have as their ambition to exploit and adapt to this diversity in a positive manner.

If schools are to be able to achieve this, it is necessary to change the system by which schools are administered. National authorities must allow greater diversity in the solutions and working methods chosen, so that these can be adapted and customized to the situation of each individual pupil, teacher, and school. The national authorities must define the objectives and contribute with good framework conditions, support, and guidance. At the same time, we must have confidence in schools and teachers as professionals. We wish to mobilize to greater creativity and commitment by allowing greater freedom to accept responsibility.

All plans for developing and improving schools will fail without competent, committed, and ambitious teachers and school administrators. They are a school system's most important assets. It is therefore an important task to strengthen and further develop the teachers' professional and pedagogical expertise and to motivate for improvements and changes. This Report held comprehensive efforts regarding competence development in schools. Education must be developed through a dialogue with those who have their daily work in and for schools. (STM 30 2004; own translation)

The difference to the accountability rhetoric of NCLB is striking. Whereas NCLB solely is focused on "academic standards" and on allocating responsibilities to states, districts, teachers, and students, the Norwegians talk about allowing for "greater diversity," about the core role of the teachers and—above all—about the "confidence in schools and teachers as professionals. The "comprehensive effort" announced is built around three dimensions—structure, process, and outcomes—and both the minister and the committee stress

time and again that one cannot expect better results without improving the structures and the processes and without considerable help from all sides (STM 30 2004). The committee tried to embed the new tools in a way that is less offensive to traditionalists by integrating the new into the familiar concepts of local monitoring and school autonomy. The proposals included the establishment of a national testing procedure to ensure that basic competencies are achieved, but stressing the “basic,” and seeing this first and foremost as a helping hand to assist schools in diagnosing where they may have a need for improvement (NOU 2003: 16).

The introduction of the national testing has been very difficult and is a not-yet-finished task, disputed by researchers and practitioners alike and still far from anything resembling NCLB (Langfeldt et al., 2008). Nobody speaks about “evidence-based teaching” or “data-driven decision-making” as prime tools to make school improvement work; the data are seen as a limited indicator, which has to be embedded in a wider understanding of a school’s program and needs. But even this limited aspiration has put a tremendous stress on both the national test developers and local communities and schools as they try to meet the new expectation regime. Because of their poor technical quality, the first wave of national tests was met by sharp criticism, even from their supporters. This forced the government to take a one-year break and completely redo the toolkit of assessment (see Langfeldt, 2008; Lie et al., 2005; Telhaug, 2006). As a result, many schools and municipalities felt more confused than controlled by the new measures. It seems that it will take some time before a more coherent pattern of working with national monitoring emerges and the different levels find sustainable strategies for dealing with the new toolkit of expectation management.¹² However, the prevailing attitude toward what might be expected can be illustrated by what a principal of a top-scoring school said at a national leaders conference: “One shouldn’t put too much into these results”; they reflected, he said, only a small part of his school’s programme and did not inform his school about the challenges they faced, not least in relation to special education. In all events, one should not expect his school to be on top next year; the next year’s class wasn’t close to the quality of this one. This was not just a

¹² See Elstad (2008), Elstad and Langfeldt (2008), Engeland et al. (2008), Isaksen (2008), Møller (2003), and Sivesind et al. (2006).

fine display of public Norwegian humility (“You shouldn’t believe you are someone”); he seemed genuinely concerned that the unexpected success would divert attention from the more pressing problems of his school and mislead parents and local politicians, with the implication of less support to tackle his school’s problems. This is a similar reaction to the one seen in Finland as they discuss the overwhelming PISA success of their country and its more or less unintended side effects (see e.g., Kivirauma et al., 2006; Simola, 2005; Uljens, 2007).

PISA and its predecessors like TIMSS played an important, but not a key, role in this development in Norway. The move toward a policy change had begun long before PISA came into being. The TIMSS and PISA data underlined that there were some substantial shortcomings to address, but PISA was not taken as a sufficient description of the challenges ahead in either the relevant committees or the parliament. Nor did PISA lead to a fundamental change in the course of action, with the one exception that the new generation of curriculum guidelines tries to adapt some of PISA’s competence conceptualizations. But this was not by chance. Most Nordic PISA researchers were scrupulous in outlining the reach of their results, pointing to the limited scope of PISA’s material, admonishing against any attempt to simplify the complexities, and warning against any expectation of comprehensive political solutions based on PISA (see e.g., Mejdning & Roe, 2006). The most substantial criticism of PISA’s reach came from within, from researchers with close connections to the project. They have analyzed particularly the match and mismatch of PISA constructs with their nation’s traditions of knowledge, culture, and schooling (see Dolin 2007; Olsen, 2005, 2007; Sjøberg, 2007). They have discussed whether and how PISA reflects the social and cultural diversity of student achievements (see e.g., Allerup, 2005, 2006, 2007). In addition, the Norwegian PISA project tried, from its beginning, to place a main focus on schools as the decisive units of action. This was not easy: PISA does not provide comprehensive, independently cross-checked school data but relies instead on the descriptions of school climate and classroom practice provided by the students and the teachers themselves, a weak source because of the well-known variance in the ways students and teachers describe the same experienced curriculum (Turmo & Lie, 2004).

It would exceed the scope of this essay to address the subtle differences between the Nordic countries with their different levels and shades of public

debate on PISA and national testing (see Langfeldt et al., 2008). What is important, however, is another fundamental difference to the NCLB approach that the Nordic countries have in common. As a recent survey of teacher education in the Nordic countries shows (Skågen, 2006), they share a fundamental trust in the quality of their teachers and the underlying teacher education. It is not that there is no room for improvement. Rather, they feel that teachers are well enough educated to do what is necessary if they are given the means and the challenge to do so. The core issue becomes then how to improve the local communities' "room to move," their ability to unleash teachers' energies, and to monitor progress in a supportive way (see Engeland et al., 2008).

No State Left Behind

What difference PISA can make was nowhere more visible than in Germany and—with a typical delay—in Austria. In Germany, PISA was from the beginning a major story, filling newspapers, forcing political responses, and engaging each and every one interested in school affairs (see Weigel, 2004). The Austrian reaction was somewhat slower; Austria seemed to have fared better in PISA 2000, at least better than Germany, which counts for quite a lot in Austria (Retzl et al., 2007). When it turned out that Austria scored worse in PISA 2003 and that the better results of 2000 might have been an artifact of flawed sampling (see Neuwirth, 2006), the climate of the discussion changed dramatically. Now both school systems were seen to be in a deep crisis, not least a crisis of their traditional school structures and their outworn forms of teaching. The response pattern as such was no surprise. Both countries have had, since the school reforms of the late 18th century (Melton, 1988), recurrent "big school debates" every 20–30 years. Every debate is a struggle about the national curriculum, and (about) every second debate is more specifically focused on the structures of schooling and their implications.¹³

The important role of school structural issues within this pattern results from the understanding in both countries that schools are state-owned

¹³ As was the case for Prussia/Germany in the early 19th century, in the 1850s, 1890s, and 1920s, and finally in the 1960s and early 1970s; see Hopmann (1988, 2000).

and state-run systems—at the national level in Austria and at the state level in Germany. Local municipalities have some responsibilities for “outer” school matters, such as buildings and equipment, but the curriculum, the hiring and firing of teachers, the licensing of school books, and the day-by-day control of all “internal” school matters, etc., are within the realm of the state’s school administration. Moreover, both countries have stratified school systems in which secondary schools are divided into different strands for “high” and “low” achievers, providing, e.g., different schools for “academic achievers” (Gymnasium), for more “practically oriented” youth (Realschule, Hauptschule, Berufsschule), and for children with “special needs” (Sonderschule).¹⁴ Both countries have systems of vocational education combining school with on-the-job training, sometimes beginning at the lower-secondary level, but more usually covering those who do not attend a Gymnasium or the like for upper-secondary education. However, in both countries, rates of attainment of the highest academic qualification (Abitur, Matura) and thereby access to universities are considered the key indicator of social equity.

In that it is the state, and the state alone, that regulates schools, school structures can be understood as institutionalized expressions of the state’s view on social class and stratification. The proverbial example of inequality in the school debates of the 1960s was the Catholic working-class girl from a rural area attending a Hauptschule; she is now replaced by the Moslem son of an immigrant family living in a poor inner-city district who also attends a Hauptschule or a Sonderschule. Within this frame, school-structure debates tend to become debates on social division; the stratified school system is regularly defended by conservatives and economists, whereas a move toward a comprehensive school system is an affair of the heart for social democrats and the labour movement—without regard to whether one system or the other has a better record in terms of social equity. In both countries, the core argument is the assumed but not proven effect stratification might have on human capital: does stratification lead to a structural underperformance of lower-class students, or does a comprehensive school limit the space and speed of development of high achievers, and vice versa?

¹⁴ The decision about which kind of school a student should attend is normally made following 4th or 6th grades.

How PISA fits in this frame is easily understood if one takes its official purpose as stated by its owner, OECD:

Quality education is the most valuable asset for present and future generations. Achieving it requires a strong commitment from everyone, including governments, teachers, parents and students themselves. The OECD is contributing to this goal through PISA, which monitors results in education within an agreed framework, allowing for valid international comparisons. By showing that some countries succeed in providing both high quality and equitable learning outcomes, PISA sets ambitious goals for others. (Angel Gurría, OECD Secretary-General n.d.)

According to the same source, PISA's "key features" have been, so far:

- Its policy orientation, with design and reporting methods determined by the need of governments to draw policy lessons.
- Its innovative "literacy" concept, which is concerned with the capacity of students to apply knowledge and skills in key subject areas and to analyze, reason, and communicate effectively as they pose, solve, and interpret problems in a variety of situations.
- Its relevance to lifelong learning, which does not limit PISA to assessing students' curricular and cross-curricular competencies but also asks them to report on their own motivation to learn, their beliefs about themselves, and their learning strategies
- Its regularity, which will enable countries to monitor their progress in meeting key learning objectives.
- Its contextualization within the system of OECD education indicators, which examine the quality of learning outcomes, the policy levers and contextual factors that shape these outcomes, and the broader private and social returns to investments in education.
- Its breadth of geographical coverage and collaborative nature, with more than 60 countries (covering roughly nine-tenths of the world economy) having participated in PISA assessments to date, including all 30 OECD countries (p. 7)

The policy orientation, with design and reporting methods determined by the need of governments to draw policy lessons, has led to a wealth of national and OECD reports using PISA data to assess the quality of school structures

and schooling, issues of social inequality, gender, migration, etc., and, not least, again and again comparisons of countries and their PISA performance in relation to other OECD indicators.¹⁵

Of course, this approach has a number of implicit assumptions, which are all but self-evident:

- The assumption that what PISA measures is somehow important knowledge for the future. There is no research available that proves this assertion beyond the point that knowing something is always good and knowing more is always better. There is not even research showing that PISA covers enough to be representative of the school subjects involved or the general school knowledge base. PISA items are based on the practical reasoning of its researchers and pre-tests of what works in all or most settings—and not on systematic research on current or future knowledge structures and needs (see Bodin, 2007; Dohn, 2007; Jahnke, 2007; Meyerhöfer, 2007; Sjøberg, 2007).
- The assumption that the economic future is dependent on the knowledge base monitored by PISA. The little research on this theme, which assumes that there is a direct relation between test scores and future economic development, relies on strong and unproven arguments that have no basis when, for instance, comparing success in PISA's predecessors and later economic development (see e.g., Fertig, 2004).
- The assumption that PISA measures what is learned in schools. This is not PISA's own starting point, which is not to use national curricula as a point of reference (as e.g., TIMSS does; see Sjøberg, 2007). The decision to focus on a small number of issues and topics that can be expected to be present in all involved countries leaves open the question of how these items represent the school curriculum as a whole¹⁶ beyond the fact that those who are successful in school do, on average, better on PISA, which is hardly a surprise inasmuch as PISA requires cognitive and, not least, language skills, which are helpful in schools as well. Some even argue that

¹⁵ Most of this is available at <<http://www.pisa.oecd.org>>.

¹⁶ See Benner (2002), Dolin (2007), Fuchs (2003), Hopmann (2001, 2006), Ladenthin (2004), Meyerhöfer (2007), and Sjøberg (2007).

PISA first and foremost monitors whatever intelligence testing monitors (Rindermann, 2006), which could lead to the somewhat irritating implication that, according to PISA, e.g., Finns are more “intelligent” than Germans or Austrians.

- The assumption that PISA measures the competitiveness of schooling: One has to keep in mind that at best 5–15 percent of the variance in the PISA results can be attributed to lasting qualities provided by the schools studied.¹⁷ Most of the variance can be attributed to factors from the outside, such as social background (see Baumert et al., 2006), that are mostly beyond the reach of schooling.
- The assumption that PISA measures and compares the quality of national school provision, not at least of school structures, teacher quality, the curriculum, etc. Although school effects as such have a very limited role in the results of PISA, one has to add that (a) PISA has had a considerable sampling and cultural-match problem, which reduces its trustworthiness as an indicator for national systems, at least for systems with the small differences seen between Western countries (see Hopmann et al., 2007); and (b) since Coleman’s (1967) seminal study, it has been well known that schools only have a very limited impact on social distributions of educational success when compared to factors such as the social fabric of the surrounding society (see Shavit & Blossfeldt, 1993). Furthermore, by virtue of its very design, PISA is forced to drop most of what might indeed indicate the specifics of national systems (see Dolin, 2007; Langfeldt, 2007).

In short, PISA relies on “strong assumptions” (Fertig, 2004) based on weak data that appeal to conventional wisdom (“Education does matter, doesn’t it?”; “School structures make a difference, don’t they?”), but offers almost no empirical and historical research supporting its implied causalities. But this has not kept both PISA researchers and the public from using PISA as if such causal relations are given. Otherwise, one would not be able to explain the two main impacts that PISA has had on school administration

¹⁷ See Watermann et al. (2003); for the principal problems of reconstructing schooling and teaching based on such data, see Rauin (2004).

and policy-making in Austria and Germany that directly refer to this frame of reference, albeit using it somewhat differently. Thus, PISA's approach to competency measuring has been a sweeping success in both countries, in part fueled by the "national expertise" (Klieme et al., 2003) produced by researchers close to the PISA efforts who argue that a national monitoring of student achievement based on an approach similar to PISA is both necessary and feasible (see Jahnke, 2007). Based on this, the German education ministers of the states have established a process that is moving toward national standards and given a helping hand to the mounting of a National Institute for Progress in Education (Institut zur Qualitätsentwicklung im Bildungswesen) with similar functions to the US NAEP. Both of these accomplishments are significant in Germany, given that curriculum matters are normally considered to be state, not federal, responsibilities and that there was no previous tradition of state-run outcome controls (except for some standardizations of final exams in a few states).¹⁸ Similarly, the Austrian government has initiated a not-yet-finished project to develop and implement national competency standards as an alternative to curriculum guidelines and to combine this with regular testing.¹⁹ All this in spite of the fact that the impact of the use of national or state assessment on what PISA and similar projects measure is at best weak in either direction²⁰ and that the overall importance of meeting the goals that PISA (and state) standards happen to measure is at best a good guess without a solid research foundation.

Whereas this approach seems to have support across the whole political spectrum, the second impact has proven to be rather divisive. Based on PISA and similar studies, researchers and politicians have—as mentioned above—reopened the debate on school structures. Interestingly, both sides—proponents and opponents of a comprehensive system, proponents and

¹⁸ Prior to this point, there have been more than 4000 different state curriculum guidelines that were the main road to defining expected results, without any regular control for whether they were achieved (as in the Nordic countries; see Hopmann 2003).

¹⁹ See the material collected at the official site, Gemeinsam Lernen, <<http://www.gemeinsamlernen.at>>.

²⁰ See Amrein and Berliner (2002), Bishop and Woessmann (2004), and Woessmann and Fuchs (2004).

opponents of early school start, proponents and opponents of an integrated teacher education, etc.—feel themselves encouraged by the very same PISA data that the other faction also uses. The most prominent example of this is a national report on schooling, written by a number of “leading experts” (i.e., researchers utilizing PISA and the like) on behalf of the Confederation of Bavarian Industry (VBW), which—focused on equity issues—argues that there is ample research evidence for re-organizing the whole school system as a two-tier organization (see Aktionsrat Bildung 2007). On the other hand, the leader of the PISA effort at OECD, Andreas Schleicher, is totally convinced that PISA proves the advantages of a comprehensive system. The leader of the national PISA effort in Austria, Günther Haider, managed first to support a continuation of the current stratified structures then a transition toward a comprehensive system, in both cases claiming PISA data as evidence for his recommendations (see Retzl et al., 2007).

Public criticism of the empirical evidence provided by PISA has been limited in both countries. The “devastating” results were too much in line with the political need to find good causes at the end of the economically painful reunification process in Germany and at a time when both countries felt themselves economically underperforming compared to other European countries, not least those seemingly more successful in PISA, such as Finland. Even the scientific discourse took the economic reasoning backing PISA for granted, arguing that PISA reduced “Bildung” to economic necessities and the needs of globalization, thereby acknowledging the unfounded premises of PISA’s ability to monitor and guide the school curriculum (see e.g., Huisken, 2005; Lohmann, 2007). Except for the obvious case of the Austrian sampling problems (Neuwirth, 2006), the few methodological objections that were voiced were either ignored or ridiculed by the PISA community and its supporters and has had—at least up to now—no substantial impact on the public standing of PISA in either Austria or Germany (Hopmann & Brinek, 2007).

The no-state-left-behind approach of the OECD and its German and Austrian consorts leads to the somewhat paradoxical effect that PISA has had the most impact by way of the by-products of the PISA research, which in design and methodology are most probably the weakest links of the whole enterprise. But even this is not without precedent. Re-reading Georg Picht’s (1964) volume on the “education catastrophe,” which initiated

the last “big school debate” in the mid-1960s, one is amazed how little of his evidence could stand the test of time and how much of it was simply speculative. However, the book was the single most important lever for the ensuing debate on how the school system should adapt to the social changes at the end of the post-war reconstruction period, a process that ended with the biggest expansion of the educational system and of public expenditure in the history of schooling. This process also included the temporary transfer of a toolkit, “scientific” curriculum development, from the United States—in spite of its then self-pronounced “moribund” state on its home turf (Hopmann, 1988). At least in Austria and Germany, PISA seems to have achieved something similar, helping politicians, educators, and the public get the educational field in touch with the transition processes going on in the whole public sector by providing them with a sense of what “manageable expectations” might be and with tools to monitor their success—or failure.

Comparative Accountability

The overall picture of the accountability approaches I have reviewed shows three very different basic philosophies of what this transition is about. The table below summarizes and pinpoints the main assumptions and entry points of each approach. In the nature of public schooling, each approach carries elements of the others.²¹ Each strategy has its own strengths and carries its own risks depending on the larger concept of expectation management it is a part of.

²¹ Additional analysis of more countries would show that there are mixed patterns, combining elements of different modes of accountability (e.g., the case of Switzerland would probably reveal a mixture of ‘no school’ and ‘no canton’ strategies; Canada a mixture of “no child” and “no school,” etc.; see. e.g., Klatt et al. (2003), Stack (2006). And there is a fourth pattern, where “no accountability has yet arrived,” and where the public sector is in the early stages of a transition toward accountability policies—and therefore not yet open for the influx of international accountability measures (as, for instance, in Italy, where PISA has not been a public issue, and even the government has treated it as almost non-existent).

Basics of the No Child, No School, No State Left Behind strategies

	No Child	No School	No State
Core Data	Student achievement	Aggregated school achievement data	Aggregated national student achievement
Main Tools	Standards controlled by testing	Testing with regard to opportunities to learn (OTL)	Competencies measured by random testing
Stakes	High stakes	Low stakes	No stakes (PISA) Low or high stakes (standards)
Driving Force	Blame	Community spirit	Competition
Main levels of attribution	Classroom and teaching	Local school management	School systems/society at large
Best Practice	Data-driven	Customized	Research-based
Accountability	Bottom-up	Bridging the gap	Top-down
The role of PISA	Almost none	Supporting act	Main act

The “no child” approach has the advantage of a clear focus. Everybody knows what counts and how it is measured. But the price for this is what Berliner (2005) calls “an impoverished view on educational reform,” a system of accountability checks that places “statistical significance” above all other ways of looking at individual and institutional achievements. The very narrow conceptualization lends itself to reduced remedial strategies: “evidence-based” or “best-practice” models, or “data-driven decision-making” only make sense if it is assumed that assessment data are all that counts and that local conditions do not play a significant role, or at least can be overcome if one does as those who are successful do. But while NCLB’s “blueprint” (US Department of Education 2007) honors a naïve empiricism, much of the NCLB-induced research provides more complex insights into the complexities of school life, using a whole range of mixed methods and avoiding the fallacies of an engineering approach to social transition (see e.g., Elmore, 2004). However, it seems unlikely that the insights produced by this research will have any lasting impact on the NCLB movement: the prime implication of this research, the importance of capacity-building in local schooling with special regard to the unique mix of challenges at hand, is contrary to any belief that the same high stakes are for everyone and distributing blame and shame in large portions is a reasonable approach to making accountability work.

Moreover, it is this one-sided focus that allows the transformation of the apparent problems with equity and equality, minorities, special needs, gender, etc., into individualized liabilities whose impact on achievement has to be minimized, if not eradicated. If high stakes as a sole approach to this fails (and research points to that it will; see Linn, 2007), there are a number of technical options to ease the burden, such as lowering the ceilings, adding opt-out clauses for the worst students and/or schools, inflating the number of stakes such that everybody can succeed in something, and (not least) creating more school choice and vouchers, which leaves the responsibility of choosing the right school with the parents. All of these options are under consideration in the current debate on the renewal of NCLB (US Department of Education 2007). “Choice” is, of course, the core of a strategy of passing the basket on to the next in line of the accountability chain, i.e., the ones who seemingly bring the liabilities to school: the parents, the minorities, the poor, those with special needs, etc. We can expect more accountability tools, e.g., contractual attainment goals and/or connecting welfare subsidies or other sanctions with them, to make these families directly responsible for the outcomes. The achievement gap will not disappear with these moves. Instead, what once was considered as a failure of the school system to cope with the diversity of society (see e.g., Coleman et al., 1967) will be turned, step by step, into a problem of individual customers failing to meet rising expectations.

The no-school-left-behind approach of Norway and most of the Nordic countries is far away from such reductionism but also pays a heavy price. The double task of embedding the new strategies in the traditional toolkit and doing so in close co-operation with the local level obviously leaves many wondering if there is a real change process going on and, if there is, what it consists of. No real sense of new obligations has emerged in schools and municipalities, and it seems as if they respond to the new accountability expectation with classic Nordic “muddling-through”: planning, co-ordinating, and reporting on a local level time and again, with no real stakes in the process and inconclusive outcomes (see Elstad 2008; Engeland et al., 2008). That the first national tests were a technical disaster (see Lie et al., 2005) and prompted a break in the whole process did not really help, nor did the new curriculum guidelines of 2006, which, in spite of much of the rhetoric, do not require more adjustments than earlier guidelines, i.e., most teaching

did not change significantly as a result of their adoption, and the prime concern of teachers remains with local adaptation, not national outcomes (see Bachmann & Sivesind, 2008).

But this will not be the end of the story! The key question is what will happen if the current effort, which even in a Norwegian perspective is quite expensive, fails to achieve significant and sustainable gains beyond those that come as the system becomes used to the new tools? Social and economic inequalities are rising rapidly, and, knowing how this can affect both schools and students, growing achievement disparities and gaps will be no surprise. Two response patterns seem likely: First, move even more rapidly toward more radical accountability strategies, i.e., raising stakes, adding more national testing, and, most importantly, adding sanctions for those who continue to fail. This would put tremendous pressure on the comprehensive school system: homogenous schools without too many non-achievers will succeed and tell the public that the time of an all-encompassing school has come to an end. Second, introduce more choice and private options into the system (as is already the case in Sweden and Denmark), thus allowing schools to remove their most challenging parents and the most challenged children, leaving the public school as the main route for “average” people (see Kvale, 2008). Both strategies would imply a definite end of the “one school for all” notion. However, this idea is deeply engrained in the social fabric of Nordic societies, and the move will be no easy task (and will lead in Norway to a continuing back and forth of who is allowed to opt out and why). But there are no other ways to reconcile the former management of places with the new needs of accountability, even if it takes time before the “muddling through” is forced to accept this consequence as inevitable.

None of this applies to the two leading examples of the no-state-left-behind strategy, Austria and Germany. Both have fragmented school systems in which comprehensive schools play no significant role. Moreover, for the moment, both have easy access to knowing if their school improvement is working. All they seemingly have to do is to wait for the next PISA wave; it will then be clear who has lost or won in the interstate competition (of course, only if one believes that PISA indeed is able to say something about that question). The main risk lies in the deeply engrained traditions of how to deal with “big school debates,” because these traditions transform the achievement problem into one of school structure and other institutional

change. The issues at stake are more or less the same in both countries (see Aktionsrat Bildung, 2007; Retzl et al., 2007): Comprehensive schools or different tracks? Free school choice and private ownership of schools, or rather the opposite? Compulsory pre-school education, and if so, for whom? Unified teacher education or different routes for different types of schooling? Special schools for special needs or inclusive education? Keeping the double structure of vocational education (school plus on-the-job training) or integrating vocational education in some general kind of upper-secondary schooling?

If the attempts to force a comprehensive restructuring fail (and there is no empirical or political evidence indicating that this could turn out otherwise), at least two possible outcomes are likely: The first outcome would be to move toward a more NCLB-like approach to accountability, i.e., adding more stakes and tests (e.g., unified entrance and exit exams), including all levels, and becoming more all-encompassing than is possible within PISA, i.e., by requiring more data on single schools, school districts, and the different federal states, eventually extending the screening beyond student achievement toward indicators on teaching patterns, teaching materials, teacher qualifications, student-career data, and the like. But, at least in Germany, such an approach faces the obstacle that schooling is constitutionally a matter for states, not the federal government, which means that there are no means for enforcing alignment beyond that which all states agree upon. In Austria, the federal government has the necessary constitutional backing for its involvement. However, since Austria's reconstruction after World War II, the federal government has been built on a compromise between the two largest political wings (the social democrats and the conservatives), each controlling about half of the states. It is unlikely that any lasting agreement on a comprehensive accountability approach is feasible. Which brings the second option to the fore, namely to dissolve, or embed, the national accountability measures in an internal competition between the different federal states. Those confident of their success would prove the advantages of their chosen solutions by their own data; those not meeting the standards would have to answer by their own explanations of why a mismatch was unavoidable. In the end, there would be an inextricable hodgepodge of testing, controlling, monitoring, etc., with each state having its own toolkit of accountability measures.

But at least two problems would be left out of either option: on the one hand, none of this addresses the problems of sustainable inner-school

development or capacity-building over the long run. A race to match changing expectations by restructuring the system will not leave energy and resources to address the tricky problems of “no teaching left behind.” Second, both approaches would lead to further marginalization of the special-needs students who are already invisible or turned into liabilities in the PISA approach (see Hörmann, 2007). They promise to become even more marginalized inasmuch as they do not assist in winning a competition that has individualized academic achievement as its basic rationale (Hopmann, 2007). Each and every new round of testing would only reaffirm their “lower” abilities and the “superiority” of the schools dealing with high achievers, thus petrifying the hierarchy of schools. In that this hierarchy has always been experienced as an expression of the social fabric of society and the state’s position toward it, one can only imagine how rapidly this will lead to inner tensions in a school system surrounded by a society with rapidly growing social inequalities.

PISA in Transition

Most of these emerging issues stem from inner tensions between the former management of placement and the new expectation regime. The data-driven NCLB disintegrates the ‘place called school’ (Goodlad, 1983) into concurrent, but not intertwined, individual challenges of meeting the standards. The no-school-left-behind approach has difficulties in embracing a coherent set of expectations as it dissolves the idea of accountability into its old routines of institutionalized muddling-through. The no-state-left-behind strategy is at risk of answering the new expectations by functionalizing them for a renewal of the conventional restructuring game, without really changing what is going on inside schools and classrooms.

The success of PISA within this transition is made possible by a certain fuzziness of design and self-presentation. It treats the links between student, school, and national achievement as self-evident, thus allowing for a black-box approach to schooling itself, where the coincidence of results and factors is transformed into correlations and causalities without establishing how this linearity comes into being. Within a management of placements, PISA and the national testing inspired by it would be dysfunctional in that it covers only a few aspects of schooling, and those in a way that does not allow for research-based decision-making concerning the whole school or

even teaching and learning under given conditions. However, as a tool of expectation management, PISA allows each setting to address problems as they are framed within the respective constitutional mindsets, using PISA as “evidence.” Thus, PISA refreshes the never-ending dispute in Germany and Austria on school structures and their relation to social class and diversity, reinigorates the co-dependency of national government and local community in the Nordic countries, and reaffirms the starting point of the NCLB discourse on failing schools and teachers as the main culprits for the uneven distribution of knowledge and cultural capital in Western societies.

The irony of this story is, of course, that PISA achieves this not in spite of, but because of, its shortcomings. Although it uses advanced statistical tools, PISA stays methodologically within the frame of a pre-Popperian positivism that takes item responses for the realities addressed by them. There is no theory of schooling or of the curriculum to allow for a non-affirmative stance toward the policy-driven expectations that, according to OECD, “determine” the design and reporting methods of PISA (OECD, 2007). There is no systematically embedded concept of how yet unheard voices and non-standardized needs could be recognized as equally valid expressions of what schooling is about. Accordingly, none of the newer developments in educational research addressing the situatedness, multi-perspectivity, non-linearity, or contingency of social action plays a significant role in PISA’s design (Hopmann, 2007)—even though the PISA data could be used within mixed methods or other more comprehensive research designs, which could address some of PISA’s inherent weaknesses as well (Olsen, 2007). But to incorporate such developments on a large scale would be close to impossible. They do not lend themselves to such generalized bottom lines as league tables; to include them in a large-scale study of the size of PISA would require resources far beyond that available to even that project.

As an entry to the commencing accountability transition, PISA has done a significant job in facilitating and illustrating the difficulties any approach to these issues will have to face. But it might be that the PISA frenzy has already reached its peak (the comparatively low-key responses to the last wave of December 2007 point in this direction). But if the PISA frenzy is drawing to a close, it will not be because of the technical mishaps and fallacies. Such details go unnoticed by politicians and the public. If PISA loses its unique position, it will occur because of its success, because of the multiplying of

PISA-like tools in national and state accountability programs. If the NCLB experience holds true, PISA will be reduced to being just one voice in the polyphonic concert of assessment results, and—having no sanctions other than statistical blame—will be overcome by accountability measures that carry more immediate risks for those involved.

The important question for the future of educational research is how much PISA then will be left behind and to what extent its methodological reductionism will prevail as the state of the art of comparative research. But the more pressing question centers on the long-term effect its conceptualization of student achievement will have on the public understanding of what schooling is about. What will happen to the school subjects left out, to the special needs that are marginalized, to school tasks that have nothing to do with higher-order academic achievement to the school functions that move beyond a one-dimensional kind of knowledge distribution? Perhaps there are new, not yet seen possibilities hidden in the multiple realities of the transition from the management of placement toward the management of expectations, even some that make research, policy, and schooling accountable for not leaving their social conscience behind on their march into the emerging age of accountability.

References

- Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. University of Chicago Press.
- Ahearn, E.M. (2000). *Educational Accountability: A Synthesis of Literature and Review of a Balanced Model of Accountability*. VA.
- Aikin, W.M. (1942). *The Story of the Eight-year Study, with Conclusions and Recommendations*. Harper & Brothers.
- Åkerstrøm Andersen, N. (2003). *Borgerens kontraktligg relse [The citizen's contractual journal]*. Hans Reitzels Forlag.
- Aktionsrat Bildung (2007). *Bildungsgerechtigkeit. Jahresgutachten 2007*. Verlag Sozialwissenschaften.
- Alexander, L., James, H.T., & Glaser, R. (1987). *The Nation's Report Card: Improving the Assessment of Student Achievement*. National Academy of Education.

- Allerup, P. (2005). PISA præstationer - målinger med skæve målestokke? [PISA performances - measurements with imprecise scales]. *Dansk Pædagogisk Tidsskrift*, 53(1), 68–81.
- Allerup, P. (2006). *PISA 2000's læseskala: vurdering af psykometriske egenskaber for elever med dansk og ikke-dansk sproglig baggrund* [PISA 2000's reading scale: Assessment of psychometric values for students with Danish and non-Danish backgrounds]. Rockwool Fondens Forskningsenhed og Syddansk Universitets forlag.
- Allerup, P. (2007). Identification of group differences using PISA scales: considering effects of inhomogeneous items. In S. T. Hopmann, G. Brinek and M. Retzl (Eds), *PISA zufolge PISA - PISA According to PISA: Hält Pisa, was es verspricht? [Does PISA Keep What It Promises]* (pp. 175–203). Lit Verlag.
- Ålvik, T. (Ed.) (1994). *Skolebasert vurdering: en artikkelsamling* [School-based assessment: A collection of articles]. Ad notam Gyldendal.
- Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18).
- Apple, M.W. (2007) Ideological success, educational failure? On the politics of No Child Left Behind. *Journal of Teacher Education*, 58(2), 108–116.
- Bachmann, K., & Sivesind, K. (2008) Regn med meg! Evaluering og ansvarliggøring i skolen [Count on me! Assessment and accountability in school]. In G. Langfeldt, E. Elstad & S. T. Hopmann (Eds), *Ansvarlighet i skolen* [Accountability in school]. Cappelen.
- Bachmann, K., Sivesind, K., & Hopmann, S.T. (2004). *Hvordan formidles læreplanen? [How in the national curriculum's content communicated?]*. Høyskoleforlag.
- Bartlett, W., Roberts, J.A., & Le Grand, J. (Eds) (1998). *A Revolution in Social Policy: Quasi-market Reforms in the 1990s*. Policy Press.
- Baumert, J., Stanat, P., & Watermann, R. (Eds) (2006). *Herkunftsbedingte Disparitäten im Bildungswesen: Vertiefende Analysen im Rahmen von PISA 2000*. Verlag für Sozialwissenschaften.
- Beck, U. (2007) *Weltrisikogesellschaft: Auf der Suche nach der verlorenen Sicherheit*. Suhrkamp.
- Beck, U., Giddens, A., & Lash, S. (2007). *Reflexive Modernisierung: Eine Kontroverse*. Suhrkamp.

- Benner, D. (2002). Die Struktur der Allgemeinbildung im Kerncurriculum moderner Bildungssysteme: Ein Vorschlag zur bildungstheoretischen Rahmung von PISA. *Zeitschrift für Pädagogik*, 48(1), 68–90.
- Benoit, W.L. (1995). *Accounts, Excuses and Apologies: A Theory of Image Restoration Strategies*. State University of New York Press.
- Berliner, D.C. (2005). Our impoverished view of educational reform. *Teachers College Record*, 106(6), 949–995.
- Birkeland, N. (2008). Ansvarlig, jeg? Accountability på norsk [What, am I accountable?]. In G. Langfeldt, E. Elstad & S. T. Hopmann (Eds), *Ansvarlighet i skolen [Accountability in school]*. Cappelen.
- Birman, B., Le Floch, K. C., Klekotka, A., Ludwig, M., Taylor, J. Walters, K. Wayne, A., Yoon, K.-S., Vernez, G., Garet, M., & O'Day, J. (2007). *State and Local Implementation of the 'No Child Left Behind Act': Vol. II: Teacher Quality under NCLB: Interim Report RP-1283*. Rand.
- Bishop, J.H., & Woessmann, L. (2004) Institutional effects in a simple model of educational production. *Education Economics*, 12(1), 17–38.
- Bloom, B.S. (Ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I: Cognitive Domain*. David McKay.
- Bodin, A. (2007). What does PISA really assess? What it does it not? A French view. In S. T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 21–56). Lit Verlag.
- Bogason, P. (Ed.) (1996). *New Modes of Local Political Organization: Local Government Fragmentation in Scandinavia*. Science Publishers.
- Bozkurt, D., Brinek G., & Retzl, M. (2007). PISA in Österreich: Mediale Reaktionen, öffentliche Bewertungen und politische Konsequenzen. In S. T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 321–362). Lit Verlag.
- Bracey, G.W. (2005) Research: put out over PISA. *Phi Delta Kappan*, 86(10), 797.
- Braun, H. (2004) Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives*, 12(1).
- Buschor, E., & Schedler, K. (Eds) (1994). *Perspectives on Performance Measurement and Public Sector Accounting*. Paul Haupt.

- Cannell, J.J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: Issues and Practice*, 7(1), 5–9.
- Caswell, H.L. (1929). *City School Surveys: An Interpretation and Analysis*. Columbia University.
- Chubb, J.E. (Ed.) (2005). *Within Our Reach: How America Can Educate Every Child*. Rowman & Littlefield.
- Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of Educational Opportunity*. US Department of Health, Education and Welfare, Office of Education.
- Conant, J.B. (1959). *The American High School Today: A First Report to Interested Citizens*. McGraw-Hill.
- Cook, T.D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky & W. R. Shadish (Eds), *Evaluation for the 21st Century: A Handbook* (pp. 30–52). Sage.
- Deretchin, L.F., & Craig, C.J. (Eds) (2007). *International Research on the Impact of Accountability Systems Teacher Education Yearbook XV*. Rowman & Littlefield.
- Dohn, N.B. (2007). Knowledge and skills for PISA: assessing the assessment. *Journal of Philosophy of Education*, 41(1), 1–16.
- Dolin, J. (2007). PISA: an example of the use and misuse of large-scale comparative tests. In S.T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 93–126). Lit Verlag.
- Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6(1).
- Dubnick, M.J. (2005). *Accountability and the promise of performance: in search of the mechanisms*. *Public Performance and Management Review*, 28(3), 376–417.
- Eberts, R., Hollenbeck L. & Stone, J. (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources*, 37(4), 913–927.
- Education Commission of the States (ECS) (2002). *No Child Left Behind Issue Brief: Data-Driven Decision-making*. Educational Commission of the States.
- Elmore, R.F. (2004). *School Reform From the Inside Out: Policy, Practice and Performance*. Harvard University Press.

- Elstad, E., & Langfeldt, G. (2008). Hvordan forholder skoler seg til målinger av kvalitetsaspekter ved lærernes undervisning og elevenes læringsprosesser? [What is the school's position regarding assessments of qualitative aspects of teachers' instruction and pupils' learning outcomes?]. In G. Langfeldt, E. Elstad & S.T. Hopmann (Eds), *Ansvarlighet i skolen [Accountability in school]*. Cappelen.
- Engeland, Ø., Langfeldt, G., & Roald, K. (2008). Kommunalt handlingsrom: hvordan møter norske kommuner ansvarsstyring i skolen? [Municipal freedom of movement: How do Norwegian municipalities deal with responsibility allocation in schools?]. In G. Langfeldt, E. Elstad and S.T. Hopmann (Eds), *Ansvarlighet i skolen [Accountability in school]*. Cappelen.
- Esping-Andersen, G. (Ed.) (1996). *Welfare States in Transition: National Adaptations to Global Economies*. Sage.
- Evers, A., Rauch, U., & Stitz, U. (Eds) (2002). *Von öffentlichen Einrichtungen zu sozialen Unternehmen. Hybride Organisationsformen im Bereich sozialer Dienstleistungen*. Edition Sigma.
- Fertig, M. (2004). What can we learn from international student performance studies? Some methodological remarks. *RWI Discussion Paper No. 23*. Rheinisch-Westfälisches Institut für Wirtschaftsforschung.
- Foucault, M. (2006). *Geschichte der Gouvernementalität 1: Sicherheit, Territorium, Bevölkerung: Vorlesung am Collège de France 1977/1978*, Ed. M. Sennelart. Suhrkamp.
- Fuchs, H.-W. (2003). Auf dem Wege zu einem neuen Weltcurriculum? Zum Grundbildungskonzept von PISA und der Aufgabenzuweisung an die Schule. *Zeitschrift für Pädagogik*, 49(2), 161–179.
- Fuhrman, S.H. (Ed.) (2001). *From Capitol to the Classroom: Standards-Based Reform in the States 100th Yearbook, Part 2, of the National Society for the Study of Education*. University of Chicago Press.
- Fukuyama, F. (1992). *The End of History and the Last Man*. Free Press.
- Goodlad, J.I. (1983). *A Place Called School: Prospect for the Future*. McGraw-Hill.
- Gorard, S. (2006). Value-added is little value. *Journal of Education Policy*, 21(2), 235–243.
- Gottweis, H., Hable, W., Prasack, B., & Wyndra, D. (2004). *Verwaltete Körper: Strategien der Gesundheitspolitik im internationalen Vergleich*. Böhlau.

- Granheim, M., Kogan, M., & Lundgren, U. P. (Eds) (1990). *Evaluation as Policymaking: Introducing Evaluation into a National Decentralized Educational System*. Jessica Kingsley Publishers.
- Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33(1), 69–86.
- Gundem, B.B. (1992). *Prosjektet læreplan og fag: Læreplanadministrering: fremvækst og utvikling i et sentraliserings-desentraliseringsperspektiv* [The national curriculum project: National curriculum administration: Growth and development from a centralization-decentralization perspective]. University of Oslo.
- Gundem, B.B. (1993). *Mot en ny skolevirkelighet? Læreplanen i et sentraliserings-desentraliseringsperspektiv* [Towards a new school reality? The national curriculum from a centralization-decentralization perspective]. Ad notam Gyldendal.
- Gundem, B.B. (1997). *Læreplanhistorie - historien om skolens innhold---som forskningsfelt: en innføring og noen eksempler* [The history of the national curriculum – the history of school content – as a research field: An introduction and some examples]. Universitetet i Oslo, Pedagogisk forskningsinstitutt.
- Gundem, B.B., & Hopmann, S. (Eds) (2002). *Didaktik and/or Curriculum: An International Dialogue*. Peter Lang.
- Haft, H. & Hopmann, S.T. (Eds) (1990). *Case Studies in Curriculum Administration History*. Falmer.
- Haney, W. (2000). The myth of the Texas miracle in education. *Educational Policy Analysis Archives*, 8(41).
- Haney, W. (2002). Lake Wobeguaranteed: misuse of test scores in Massachusetts, Part 1. *Educational Policy Analysis Archives*, 10(24).
- Hanushek, E.A. (2002). *The long-run importance of school quality*. NBER Working Paper No. 9071. National Bureau of Economic Research.
- Hanushek, E.A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485), 64–98.
- Hanushek, E.A. (2006). Alternative school policies and the benefits of general cognitive skills. *Economics of Education Review*, 25(4), 447–462.
- Hanushek, E.A., & Raymond, M.E. (2003). Lessons about the design of state accountability systems. In P. E. Peterson and M. R. West (Eds),

- No Child Left Behind: The Politics and Practice of School Accountability* (pp. 127–151). Brookings.
- Hanushek, E.A., & Raymond, M.E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297–327.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society: Education in the Age of Insecurity*. Teachers College Press.
- Haug, P., & Monsen, L. (Eds) (2002). *Skolebasert vurdering: erfaringer og utfordringer [School-based assessment: Experiences and challenges]*. Abstrakt.
- Herman, J.L., & Haertel, E.H. (Eds) (2005). *Uses and Misuses of Data for Educational Accountability and Improvement 104th Yearbook, Part 2, of National Society for the Study of Education*. Blackwell.
- Hood, C. (1991). A public management for all seasons. *Public Administration*, 69(1), 3–20.
- Hood, C. (1995). Contemporary public management: a new global paradigm? *Public Policy and Administration*, 10(2), 104–117.
- Hood, C. (2004). *Institutions, blame avoidance and negativity bias: Where public management reform meets the 'blame culture'*. Paper presented at the CMPO Conference on Public Organisation and the New Public Management, Bristol.
- Hood, C., Rothstein, H., & Baldwin, R. (2004). *The Government of Risk: Understanding Risk Regulation Regimes*. Oxford University Press.
- Hopmann, S.T. (1988). *Lehrplanarbeit als Verwaltungshandeln [Curriculum work as administrative action]*. IPN.
- Hopmann, S.T. (2000). Lehrplan des Abendlandes: am Ende seiner Geschichte? Geschichte der Lehrplanarbeit und des Lehrplans seit 1900 [Curriculum of the West: at the end of its history? History of curriculum work and the curriculum since 1900]. In R. W. Keck and Ritzi, C. (Eds), *Geschichte und Gegenwart des Lehrplans: Josef Dolchs 'Lehrplan des Abendlandes' als aktuelle Herausforderung* (pp. 377–400). Schneider Verlag Hohengehren.
- Hopmann, S.T. (2001). Von der gutbürgerlichen Küche zu McDonald's: Beabsichtigte und unbeabsichtigte Folgen der Internationalisierung der Erwartungen an Schule und Unterricht [From home-style cooking to McDonald's: Intended and unintended consequences of the

- internationalisation of expectations of schools and teaching]. In E. Keiner (Ed.), *Evaluation in den Erziehungswissenschaften* (pp. 207–224). Beltz.
- Hopmann, S.T. (2003). On the evaluation of curriculum reforms. *Journal of Curriculum Studies*, 35(4), 459–478.
- Hopmann S.T. (2006). Im Durchschnitt Pisa oder: Alles bleibt schlechter [On average Pisa or: Everything remains worse]. In L. Criblez (Ed.), *Lehrpläne und Bildungsstandards: was Schülerinn und Schüler lernen sollen: Festschrift zum 65. Geburtstag von Prof. Dr. Rudolf Künzli* (pp. 149–172). Bern.
- Hopmann, S.T. (2007). Keine Ausnahme für Hottentotten: Methoden der vergleichenden Bildungswissenschaften für die heilpädagogische Forschung [No exception for Hottentots: Methods of comparative educational science for curative education research]. In G. Biewerand and M. Schwinge (Eds), *Internationale Sonderpädagogik* (76–98). Klinkhardt.
- Hopmann, S.T., & Brinek, G. (2007). Introduction: PISA according to PISA: Does PISA keep what it promises. In S. T. Hopmann, G. Brinek and M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht?* [According to PISA - Does PISA Keep What It Promises] (pp. 9–20). Lit Verlag.
- Hopmann, S.T., Brinek, G., & Retzl, M. (Eds) (2007). *PISA zufolge: Hält PISA, was es verspricht?* [According to PISA -Does PISA Keep What It Promises]. Lit Verlag.
- Hörmann, B. (2007). Disappearing students: PISA and students with disabilities. In S. T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht?* [According to PISA -Does PISA Keep What It Promises] (pp. 157–174). Lit Verlag.
- Huisken, F. (2005). *Der ‘PISA-Schock’ und seine Bewältigung: Wieviel Dummheit braucht/verträgt die Republik?* [The ‘PISA shock’ and how to overcome it: How much stupidity does the republic need/tolerate?] VSA-Verlag.
- Ingersoll, R.M. (2003). *Who Controls Teachers’ Work? Power and Accountability in America’s Schools*. Harvard University Press.
- Irons, J.E., & Harris, S. (2007). *The Challenges of No Child Left Behind: Understanding the Issues of Excellence, Accountability, Choice*. Rowman and Littlefield.
- Isaksen, L. (2008). Skoler i gapestokken [Schools in the pillory]. In G. Langfeldt, E. Elstad & S.T. Hopmann (Eds), *Ansvarlighet i skolen* [Accountability in school]. Cappelen.

- Jahnke, T. (2007) Deutsche Pisa-Folgen [German Pisa consequences]. In S. T. Hopmann, G. Brinek and M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 305–320). Lit Verlag.
- Jahnke, T., & Meyerhöfer, W. (Eds) (2006). *PISA and Co: Kritik eines Programms [PISA and Co: Criticism of a programme]*. Franzbecker.
- Karlsen, G.E. (1993). *Desentralisering - løsning eller oppløsning: søkelys på norsk skoleutvikling og utdanningspolitikk [Decentralization - solution or dissolution: Spotlight on Norwegian school development and education policy]*. Ad notam.
- Karlsen, G.E. (1994). *EU, EØS og utdanning: en studie av bakgrunn og innhold, føringer og konsekvenser for norsk utdanning [The EU, EEA and education: A study of background and content, guidelines and consequences for Norwegian education]*. Tano.
- Kirk-, Utdannings- og Forskningsdepartementet (KUF) (1994). *Underveis: Håndbok i skolebasert vurdering: grunnskole og videregående skole [Underway: Handbook of school-based assessment: Primary and upper secondary school]*. KUF.
- Kirk-, Utdannings- og Forskningsdepartementet (KUF) (1997). *Rapport om nasjonalt vurderingssystem - Moe utvalget [Report on national assessment system – the Moe committee]*. KUF.
- Kivirauma, J., Klemala, K., & Rinne, R. (2006). Segregation, integration, inclusion: the ideology and reality in Finland. *European Journal of Special Needs Education*, 21(2), 117–133.
- Klatt, B., Murphy, S., & Irvine, D. (2007). *Accountability: Getting a Grip on Results*. Bow River.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J. Tenorth, H.E., & Vollmer, H.J. (2003). *Zur Entwicklung nationaler Bildungsstandards: Expertise [On the development of national educational standards: Expertise]*. Bundesministerium für Bildung und Forschung.
- Koretz, D.M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, 37(4), 752–777.
- Koritzinsky, T. (2000). *Pedagogikk og politikk i L97: Læreplanens innhold og beslutningsprosessene [Education and politics in L97: The national curriculum's content and decision-making processes]*. Universitetsforlaget.

- Korsgaard, O. (2004). *Kampen om folket: et dannelsesperspektiv på dansk historie gennem 500 år* [*The fight for the people: An educational perspective on Danish history through 500 years*]. Gyldendal.
- Künzli, R. & Hopmann, S.T. (Eds) (1998). *Lehrpläne: Wie sie entwickelt werden und was von ihnen erwartet wird. Forschungsstand, Zugänge und Ergebnisse aus der Schweiz und der Bundesrepublik Deutschland* [*Curricula: How they are developed and what is expected of them. State of research, approaches and results from Switzerland and the Federal Republic of Germany*]. Rügger.
- Kvale, G. (2008). 'Det er ditt val!' - om fritt skuleval i to norske kommunar. [It's your choice! On free school choice in two Norwegian municipalities]. In G. Langfeldt, E. Elstad, & S. T. Hopmann (Eds), *Ansvarlighet i skolen* [*Accountability in school*]. Cappelen.
- Ladd, H.E., & Walsh, R.P. (2002). Implementing value-added measures of school effectiveness. *Economics of Education Review*, 21(1), 1–17.
- Ladenthin, V. (2004). Bildung als Aufgabe der Gesellschaft [Education as a task for society]. *Studia Comenia et Historica*, 34(71/72), 305–319.
- Laffont, J.-J. (Ed.) (2003). *The Principal Agent Model: The Economic Theory of Incentives*. Edward Elgar Publishing.
- Lange, S. & Schimank, U. (Eds) (2004). *Governance und gesellschaftliche Integration* [*Governance and social integration*]. Leske & Budrich.
- Langfeldt, G. (2007). PISA - undressing the truth or dressing up a will to govern? In S.T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge PISA - PISA According to PISA: PISA zufolge: Hält PISA, was es verspricht?* [*According to PISA -Does PISA Keep What It Promises*] (pp. 225–240). Lit Verlag.
- Langfeldt, G. (2008). Resultatstyring som verktøy og ideologi. Statlige styringsstrategier i utdanningssektoren. [Result management as tool and ideology. Nationalized management strategies in the education sector]. In G. Langfeldt, E. Elstad, & S.T. Hopmann (Eds), *Ansvarlighet i skolen* [*Accountability in school*]. Cappelen.
- Langfeldt, G., Elstad, E. & Hopmann, S.T. (Eds) (2008). *Ansvarlighet i skolen* [*Accountability in school*]. Cappelen.
- Le Floch, K. C., Martinez, J. F., O'Day, J., Stecher, B. M., Cook, A., Vernez, G., Birman, B., & Garet, M. (2007). *State and Local Implementation of the 'No Child Left Behind Act': Vol. III: Accountability Under NCLB: Interim Report RP-1303*. Rand.

- Leibfried, S., & Zürn, M. (2006). *Transformation des Staates*. Suhrkamp.
- Lie, S., Hopfenbeck, T.N., Ibsen, E. & Turmo, A. (2005). *Nasjonale prøver på ny prøve: Rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005* [National tests being retested: Report from a committee study analysing and assessing the quality of tasks and results of national tests, spring 2005]. *Acta Didactica*, 1/2005. Universitetet i Oslo, Institutt for lærerutdanning og skoleutvikling.
- Lindblad, S., Johannesson, I. A., & Simola, H. (2003). Education governance in transition: an introduction. *Scandinavian Journal of Educational Research*, 46(1), 237–245.
- Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4–16.
- Linn, R.L. (2005) Issues in the design of accountability systems. In J.L. Herman, & E.H. Haertel (Eds), *Uses and Misuses of Data for Educational Accountability and Improvement 104th Yearbook, Part 2, of the National Society for the Study of Education* (pp. 78–98). Blackwell.
- Linn R.L., & Haug, C. (2002). Stability of school building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 24(1), 29–36.
- Linn, R.L., Baker, E.L. & Betebenner, D.W. (2002). Accountability systems: implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3–6.
- Lohmann, I. (2007). *Was bedeutet eigentlich 'Humankapital'? GEW Bezirksverband Lüneburg und Universität Lüneburg: Der brauchbare Mensch. Bildung statt Nützlichkeitswahn* [What does 'human capital' actually mean? GEW Bezirksverband Lüneburg and University of Lüneburg: The useful human being. Education instead of utilitarianism].
- Loveless, T. (2006). *The Peculiar Politics of No Child Left Behind*. Brookings.
- Luhmann, N. (1998) *Die Gesellschaft der Gesellschaft* [The society of the society]. Suhrkamp.
- Marsh, J.A., Pane, J.F., & Hamilton, L.S. (2006). *Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research*. Rand.
- Martineau, J.A. (2006). Distorting value-added: the use of longitudinal, vertically-scale student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31(1), 35–62.

- McNeil, L.M. (2000). *Contradictions of School Reform: Educational Costs of Standardized Testing*. Routledge.
- Mehrens, W.A. (1996). Consequences of assessment: what is the evidence? *Educational Policy Analysis Archives*, 6(13).
- Mejding, J., & Roe, A. (Eds) (2006). *Northern Lights on PISA 2003: A Reflection from the Nordic Countries*. Nordic Council of Ministers.
- Melton, J.V.H. (1988). *Absolutism and the Eighteenth-Century Origins of Compulsory Schooling in Prussia and Austria*. Cambridge University Press.
- Meyer, J.W. (2005). *Weltkultur: wie die westlichen Prinzipien die Welt durchdringen*, ed. G. Krücken [World culture: how Western principles permeate the world, ed. G. Krücken]. Suhrkamp.
- Meyerhöfer, W. (2005). *Tests im Test: Das Beispiel PISA [Tests in the test: The example of PISA]*. Barbara Budrich.
- Meyerhöfer, W. (2007). Testfähigkeit: Was ist das? [Test capability: What is it?] In S. T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 57–92). Lit Verlag.
- Micklewright, J., & Schnepf, S.S. (2004). *Educational Achievement in the English-speaking Countries: Do Different Surveys Tell the Same Story?* IZA.
- Micklewright, J., & Schnepf, S.S. (2006). *Inequality of Learning in Industrialised Countries*. IZA.
- Mintrop, H. (2003). The limit of sanctions in low-performing schools: a study of Maryland and Kentucky schools on probation. *Education Policy Analysis Archives*, 11(3).
- Møller, J. (2003) Coping with accountability: a tension between reason and emotion. In C. Sugrue (Ed.), *Passionate Principalship: Learning from Life Histories of School Leaders* (pp. 89–103). Routledge Falmer.
- Muir Grey, J.A. (1997). *Evidence-based Healthcare: How to Make Health Policy and Management Decisions*. Churchill Livingstone.
- National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC: US Government Printing Office). ERIC ED 226 006.
- Nesje, K., & Hopmann, S.T. (Eds) (2002). *En lærende skole: L97 i Skolepraksis [A learning school: L97 in school practice]*. Cappelen.

- Neuwirth, E. (2006). *PISA 2000: Sample Weight Problems in Austria*. OECD Education Working Paper No. 5. OECD.
- NOU 1988:22. *Med viten og vilje [With knowledge and willingness]*.
- NOU 2003:16. *I første rekke [First in line]*.
- Organisation for Economic Co-operation and Development (OECD) (2005). *Education at a Glance: OECD Indicators*. OECD.
- Organisation for Economic Co-operation and Development (OECD) (n.d). *PISA - The OECD Programme for International Student Assessment*. OECD.
- Olsen, R.V. (2005). *Achievement Tests From an Item Perspective: An Exploration of Single Item Data from the PISA and TIMSS Studies, and How Such Data Can Inform Us about Students' Knowledge and Thinking in Science*. Doctoral thesis, University of Oslo.
- Olsen, R.V. (2007). Large-scale international comparative achievement studies in education: their primary purposes and beyond. In S. T. Hopmann, G. Brinek and M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 265–294). Lit Verlag.
- Peterson, P. E., & West, M. R. (Eds) (2003). *No Child Left Behind: The Politics and Practice of School Accountability*. Brookings.
- Picht, G. (1964). *Die Deutsche Bildungskatastrophe: Analyse und Dokumentation [The German education catastrophe: Analysis and Documentation]*. Walter-Verlag.
- Pollitt, C., & Bouckaert, G. (2004). *Public Mangement Reform: A Comparative Analysis*. Oxford University Press.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford University Press.
- Programme for International Student Assessment (PISA) (2001). *Knowledge and Skills for Life: First Results from PISA 2000*. OECD.
- Programme for International Student Assessment (PISA) (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. ERIC ED 480 801.
- Rasmussen, J. (2004). *Undervisning i det refleksiivt moderne: politik, profession, pædagogik [Education in the reflexive modern: Politics, profession, pedagogy]*. Hans Reitzel.

- Rauin, U. (2004). Die Pädagogik im Bann empirischer Mythen: Wie aus empirischen Vermutungen scheinbare pädagogische Gewissheit wird [Pedagogy under the spell of empirical myths: How empirical assumptions become apparent pedagogical certainty]. *Pädagogische Korrespondenz*, 32(1), 39–49.
- Rickover, H.G. (1963). *American Education, A National Failure: The Problem of Our Schools and What We Can Learn from England*. Dutton.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? [What do international school performance studies measure? School performance, student ability, cognitive skills, knowledge or general intelligence?] *Psychologische Rundschau*, 57(1), 69–86.
- Scharpf, F.W., & Schmidt, V. (Eds) (2000a). *Welfare and Work in the Open Economy: Vol. 1: From Vulnerability to Competitiveness*. Oxford University Press.
- Scharpf, F. W., & Schmidt, V. (Eds) (2000b). *Welfare and Work in the Open Economy: Vol. 2: Diverse Responses to Common Challenges in Twelve Countries*. Oxford University Press.
- Sedikides, C., Herbst, K. C., Hardin, D. P., & Dardis, G. J. (2002). Accountability as a deterrent to self-enhancement: the search for mechanisms. *Journal of Personality and Social Psychology*, 83(3), 592–605.
- Shavit, Y., & Blossfeld, H.-P. (Eds) (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Westview.
- Simola, H. (2005). The Finnish miracle of PISA: historical and sociological remarks on teaching and teacher education. *Comparative Education*, 41(4), 455–470.
- Sivesind, K. (2002). Task and themes in communication about the curriculum: the Norwegian compulsory school reform in perspective. In M. Rosenmund, A.-V. Fries & W. Heller (Eds), *Comparing Curriculum-making Processes* (pp. 319–332). Peter Lang.
- Sivesind, K. (forthcoming). *Reformulating Reforms*. Universitetet i Oslo, Institutt for lærerutdanning og skoleutvikling.
- Sivesind, K., Bachmann, K., & Afzar, A. (2003). *Nordiske læreplaner [Nordic curricula]*. Læringssenteret.

- Sivesind, K., Langfeldt, G., & Skedsmo, G. (2006). *Utdanningsledelse. Cappelen akademisk*.
- Sjøberg, S. (2007). PISA and 'real life challenges': mission impossible? In S. T. Hopmann, G. Brinek & M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 203–224). Lit Verlag.
- Slagstad, R. (1998). *De nasjonale strategier [The national strategists]*. Oslo Pax Forlag.
- Slavin, R.E. (2006). *Educational Research in an Age of Accountability*. Allyn & Bacon.
- Stack, M. (2006). Testing, testing, read all about it: Canadian press coverage of the PISA results. *Canadian Journal of Education*, 29(1), 49–69.
- STM 37 (1990–1991). *Om organisering og styring i utdanningssektoren [On organisation and management in the education sector]*.
- STM 29 (1994–1995). *Om prinsipper og retningslinjer for tiårig grunnskole: ny læreplan [On principles and guidelines for the ten-year primary school: a new curriculum]*.
- STM 47 (1995–1996). *Om elevvurdering, skolebasert vurdering og nasjonalt vurderingssystem [On pupil assessment, school-based assessment and the national assessment system]*.
- STM 28 (1998–1999). *Mot rikare mål. Nasjonalt vurderingssystem for grunnskolen [Towards worthier goals. National assessment system for the primary school]*.
- STM 17 (2002–2003). *Om statlige tilsyn [On nationalized supervision]*.
- STM 30 (2003–2004). *Kultur for læring [Culture for learning]*.
- Sutherland, D., & Price, R. (2007). *Linkages Between Performance and Institutions in the Primary and Secondary Education Sector OECD Economics Department Working Papers No. 558*. OECD.
- Swanson, C.B., & Stevenson, D.L. (2002). Standards-based reform in practice: evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, 24(1), 1–27.
- Telhaug, A.O. (2005). *Kunnskapsløftet: Ny eller Gammel skole? [The knowledge promise: New or old school?]*. Cappelen Akademisk.

- Telhaug, A.O. (2006). *Skolen mellom stat og marked: norsk skoletenkning fra år til år 1990–2005* [School between state and market: Norwegian educational thought from year to year 1990–2005]. Didakta.
- Telhaug, A.O., & Mediås, O.A. (2003). *Grunnskolen som nasjonsbygger: fra statspietisme til nyliberalisme* [The primary school as nation-builder: From state pietism to neo-liberalism]. Abstrakt.
- Turmo, A., & Lie, S. (2004). Hva kjennetegner norske skoler som skårer høyt i PISA 2000? [What characterises Norwegian school achieving high scores on PISA 2000?]. *Acta Didactica*, No 1/2004. Universitetet i Oslo, Institutt for lærerutdanning og skoleutvikling.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. University of Chicago Press.
- Uljens, M. (2007). The hidden curriculum of PISA: the promotion of neo-liberal policy by educational assessment. In S. T. Hopmann, G. Brinek and M. Retzl (Eds), *PISA zufolge: Hält PISA, was es verspricht? [According to PISA -Does PISA Keep What It Promises]* (pp. 295–304). Lit Verlag.
- US Department of Education (2007). *Building on Results: A Blueprint for Strengthening the 'No Child Left Behind Act'* (Washington, DC: US Department of Education). ERIC ED 495 309.
- Wallerstein, I.M. (2004). *World-Systems Analysis: An Introduction*. Duke University Press.
- Watermann, R., Stanat, P., Kunter, M., Klieme, E., & Baumert, J. (2003). Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen: Das Disseminationskonzept von PISA-2000 [School feedback in the context of school performance assessments: The dissemination concept of PISA-2000.] *Zeitschrift für Pädagogik*, 49(1), 92–111.
- Watson, S., & Supovitz, J. (2001). Autonomy and accountability in the context of standards-based reform. *Education Policy Analysis Archives*, 9(32).
- Weigel, T.M. (2004). *Die PISA-Studie im bildungspolitischen Diskurs. Eine Untersuchung der Reaktionen auf PISA in Deutschland und im Vereinigten Königreich* [The PISA study in the education policy discourse. An analysis of the reactions to PISA in Germany and the United Kingdom].