Transformations of the State
CRC 597

TranState Working Papers

Governance by Comparison – How Ratings & Rankings Impact National Policy-making in Education

Kerstin Martens
Dennis Niemann

No. 139

Universität Bremen • University of Bremen
Jacobs Universität Bremen • Jacobs University Bremen
Universität Oldenburg • University of Oldenburg

Staatlichkeit im Wandel • Transformations of the State
Sonderforschungsbereich 597 • Collaborative Research Center 597

Kerstin Martens
Dennis Niemann

Governance by Comparison – How Ratings & Rankings Impact National Policy-making in Education

TranState Working Papers

No. 139

Sfb597 „Staatlichkeit im Wandel“ − „Transformations of the State“
Bremen, 2010
[ISSN 1861-1176]

Kerstin Martens, **Dennis Niemann**
Governance by Comparison – How Ratings & Rankings Impact National Policy-making in Education
(TranState Working Papers, 139)
Bremen: Sfb 597 „Staatlichkeit im Wandel“, 2010
ISSN 1861-1176

Universität Bremen
Sonderforschungsbereich 597 / Collaborative Research Center 597
Staatlichkeit im Wandel / Transformations of the State
Postfach 33 04 40
D - 28334 Bremen
Tel.:+ 49 421 218-8720
Fax:+ 49 421 218-8721
Homepage: http://www.staatlichkeit.uni-bremen.de

Diese Arbeit ist im Sonderforschungsbereich 597 „Staatlichkeit im Wandel“, Bremen, entstanden und wurde auf dessen Veranlassung unter Verwendung der ihm von der Deutschen Forschungsgemeinschaft zur Verfügung gestellten Mittel veröffentlicht.

Deutsche Forschungsgemeinschaft
DFG

Governance by Comparison – How Ratings & Rankings Impact National Policy-making in Education

ABSTRACT

How can international comparisons have an impact on one country while others are not affected at all? This paper examines the power of ratings & rakings (R&R) using the example of the OECD's PISA study (Programme for International Student Assessment) and its differential impact on national education policy making. We argue that R&R have an impact if the evaluated topic is framed as crucial for national objectives and if, at the same time, a substantial gap between national self-perception and the empirical results can be observed. After assessing the media impact of PISA on 22 OECD countries, we illustrate our theoretical argument through the use of examples of two poorly performing countries who demonstrated entirely opposite reactions: Germany and the U.S. While the German system of secondary education was strongly affected by the international comparison and underwent comprehensive changes, the U.S. did not respond to its below-average ranking at all. The theoretical concepts of self-perception and framing offer explanatory power to delineate the different reaction patterns.

CONTENTS

1. INTRODUCTION 1
2. NATIONAL (MEDIA-) REACTIONS TO PISA 3
3. GOVERNANCE BY COMPARISON – A THEORETICAL APPROXIMATION 6
   3.1 Framing an issue as problematic 8
   3.2 Substantial gap between self-perception and results 9
4. TWO SIDES OF THE SAME (PISA-) COIN – ASSESSING THE CONTRARY REACTIONS OF GERMANY AND THE U.S. 11
   4.1 Germany – the rediscovery of education policy-making due to the PISA shock 11
   4.2 The U.S. – or how a country disregards its PISA results 15
5. CONCLUSION AND OUTLOOK 18
REFERENCES 20
BIOGRAPHICAL NOTE 23

Governance by Comparison – How Ratings & Rankings Impact National Policy-making in Education[1]

1. INTRODUCTION

When the OECD's Programme for International Student Assessment (PISA) released the results from its third cross-country comparison of the knowledge and skills of high school students in early December 2007, a contentious and hysterical debate about education was (again) unleashed in Germany. PISA is the largest international comparative education study surveying the competencies and skills of 15 and 16-year-olds in reading, mathematics, and science at the end of compulsory schooling in some 57 countries. It first began in 2000 and has since been conducted every three years with the results being published the following year.[2] Being developed and carried out by the Organisation for Economic Cooperation and Development, PISA is obviously not ideologically neutral, but rather evaluates education from an economic perspective and promotes, according to this paradigm, related learning techniques. From this viewpoint, PISA is much closer to the pre-existing Anglo-American understanding of education than to that of continental Europe.

The impacts of the comparative results vary substantially across the participating member countries. The first PISA report, released in December 2001, caused a "shock" in Germany: the country that traditionally prided itself on its education system and its contributions to western science and philosophy ranked at the lower end of the comparative spectrum. The expected superiority of the German education system empirically turned out to be no more than just mediocrity. German students did poorly in all of

---

[1] The research on U.S. reactions to PISA was conducted during a two-month stay in Washington in summer 2008. Kerstin Martens would like to thank the American Institute for Contemporary German Studies (AICGS) and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst – DAAD) for having enabled this stay with a generous grant. For the empirical part, 25 interviews were conducted with politicians, officials of administrative institutions, education researchers, representatives of think tanks, trade unions, lobby groups and the like. For comments on earlier drafts of this article, we would like to thank Hans Krause Hansen, Arthur Mühlen-Schulte and the participants of the Copenhagen GARNET workshop on "Powers in Numbers: Exploring the Use of Ratings, Rankings and Benchmarking Schemes in Global Governance" and those of the panel on "The Power of Numbers: Exploring the Use of Ratings, Rankings, and Benchmarking Schemes in Global Governance" at the International Studies Association meeting in New York City in 2009, in particular Timothy J. Sinclair. For further commentaries we also thank Sotiria Grek and Harald Müller. For preparing the media analysis we would like to thank Priya Fielding-Singh, Marie Popp and Gesa Schulze.

[2] For further information on PISA, see: www.pisa.oecd.org.

the three tested subjects. Whereas Germans had always tacitly assumed that they belonged to the world's leaders in education, this international data now revealed that their performance was well below average in the twenty-first century. In essence, the PISA-experience unleashed a comprehensive reform initiative in the German education system.

In many other industrialized countries, by contrast, the PISA results remained unpublicized, triggering neither public discussion nor any reforms in education policy – despite the fact that these countries were faced with poor or even poorer results than Germany. The U.S. is a prime example for a complete lack of response to PISA: proud of being the leading nation in tertiary education, the prime destination for foreign students and with eight universities among the top ten of the well-known Shanghai Ranking, it does not seem to be concerned when it comes to secondary education. It too finds itself at the lower end of the PISA ranking, but no debate about improving the quality of schools or teaching has taken place after such disastrous results.

How can an international comparison have an impact on one country while others remain entirely unaffected? Or to put it differently: under what circumstances can such ratings and rankings (R&R) produced by an international organization (IO) have influence in terms of causing severe national reactions leading to substantial reforms of the education sector? This paper examines the governance by comparison that IOs exert through the example of the OECD's PISA study and its diverging impact on national education policy-making. In a rather explorative approach, we propose that R&R have an impact if the evaluated topic is framed as crucial (for state purposes) and if, at the same time, a substantial gap between the national self-perception and the empirical results can be observed. By examining the particularly prominent case of the PISA study, our study contributes to the general discussion regarding the authority of IOs – such as the OECD – in influencing national policy-making (Barnett and Finnemore 2004; Reinalda and Verbeek 1998; Joachim et al. 2008; Hawkins et al. 2006).

The paper is structured as follows: In the next section, the diversity of national reactions to PISA is exemplified via the analysis of media coverage in 22 contributing OECD countries. The aim of this section is to show that the extent of public reactions related to PISA cannot be linked directly to the actual position of a country in the PISA ranking. Even countries with a similar performance vary substantially regarding their public reactions. Considering the empirical observations, we then secondly develop a theoretical framework in order to understand the power of IOs in conducting R&Rs that offers explanatory power in assessing different reactions in similarly performing OECD countries drawing from rationalism and sociological institutionalism. We suggest that, in analyzing varying reactions, it is decisive if an issue is framed as crucial and if a gap between self-perception and empirical results exists. We do not seek to establish causal

linkages nor do we claim to present a dense theoretical model. In fact, we provide a thick description of a social phenomenon and suggest conditions that might account for an explanation. Therefore, this working paper is intended as a first, preparatory step in developing a causal explanatory model.

By examining the two crucial cases of poor performing PISA countries that vary substantially regarding their (public and political) reactions – Germany and the U.S. – the different reaction schemes are illustrated and linked to the theoretical assumptions in the third section of this article. In the German case, we show that the OECD's PISA study has managed to transform poor results into the activation of broad education reform dynamics by revealing a gap between self-perception and evidence as well as by generating a link to other crucial issues of state performance. In contrast, these mechanisms have failed in the case of the U.S. regarding PISA but were present in prior education debates.

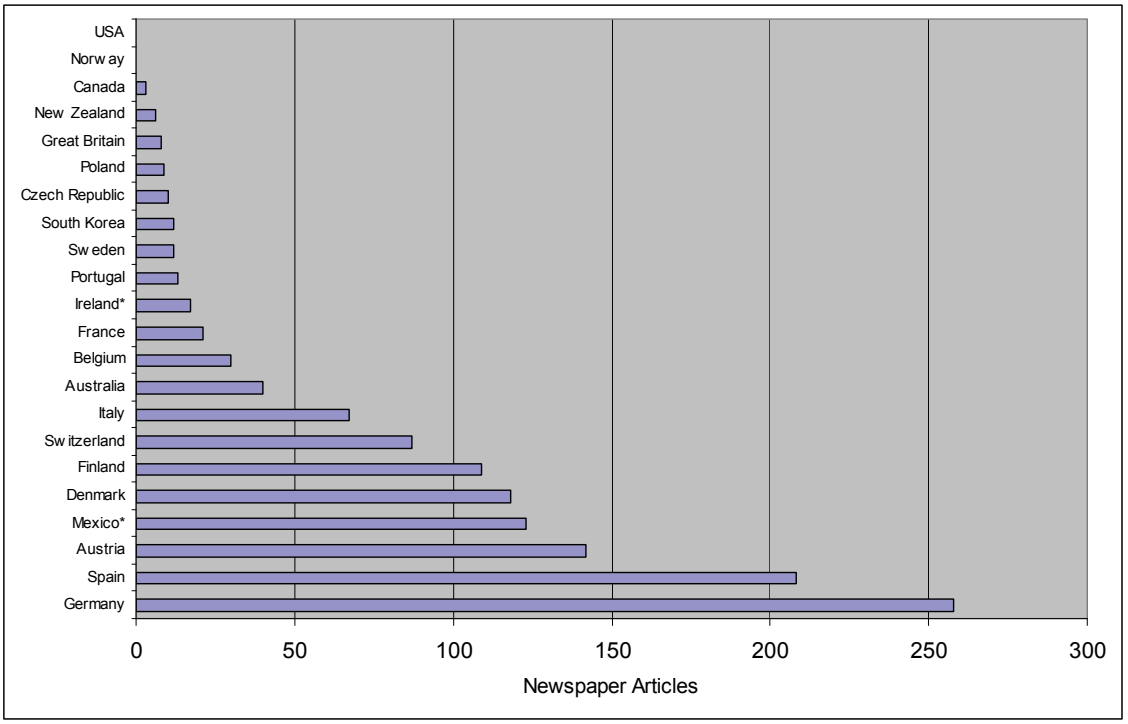
2. NATIONAL (MEDIA-) REACTIONS TO PISA

For an assessment of national reactions to PISA, the media attention this comparison received in participating countries provides a first (but not exclusive) indicator to analyze the domestic impact of this study, in particular in the absence of detailed studies on its influence.[3] But media reaction should not misleadingly be equated with reform reactions. Whether a country responds medially or politically are two different things. And high media reaction is no guarantee for political reform actions. Generally, the saliency of an issue is commonly reflected by its perception in the media which displays what topics are the subject of public controversy. For instance, if there are many articles in national newspapers dealing with an issue, it can be interpreted as a particularly salient topic in a current national discourse. Looking at the public reactions to PISA in diverse national contexts, a different degree of educational saliency becomes obvious: The media coverage between the participating countries regarding PISA is as diverse as the results. In some countries, such as the U.S., Norway, Canada, New Zealand, and Great Britain, the media has not paid any significant attention to the OECD's education comparison or its specific country's results in it. Hence, PISA did not initiate a public debate on education. In other countries, such as Germany, Spain, Mexico, and Austria, PISA has been the subject of broad media coverage (see fig. 1). In these countries, in contrast, the publication of PISA has been accompanied by an extensive public discourse on edu-

[3] PISA is still a comparatively new phenomenon. Although the OECD's education policy has received frequent academic attention (see in particular Papadopoulos 1994, Henry et al. 2001), in-depth impact studies on PISA's influence on national education policy are still rare. Exceptions are Ertl 2006; Tillmann et al. 2008; Grek 2009. See also http://www.knowandpol.eu for ongoing work on this issue.

cation which culminated – in some cases – in education reforms to overcome the highlighted deficits. In countries with almost no public reactions to negative PISA results, political reforms are not likely to occur.

Figure 1: PISA related media coverage[4]

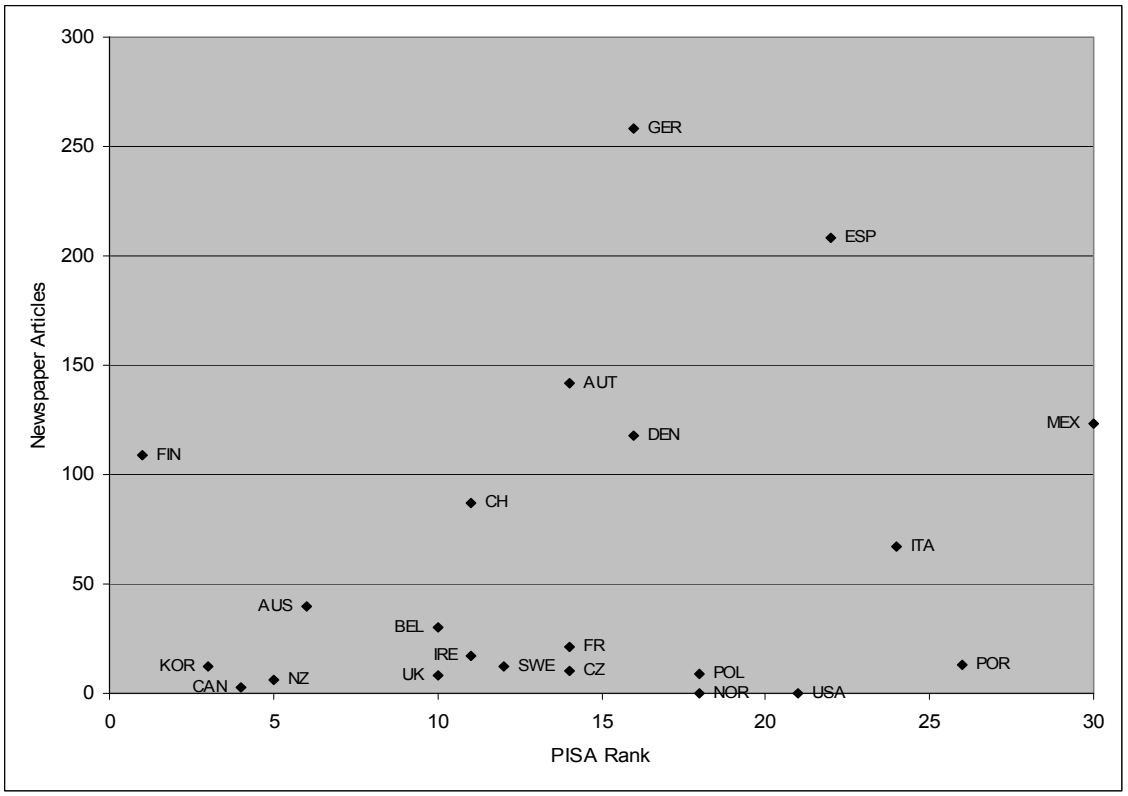


What are the driving rationales behind these very different reactions to the same event? One would intuitively expect a strong correlation between the results of a participating country in the study and the media feedback to it. More precisely, particularly good or particularly poor results should positively correlate with comprehensive media coverage of education topics. One could easily assume that the empirical evidence of lagging be-

[4] To assess the PISA related media coverage in each of the 22 OECD countries, one national high-quality daily newspaper with high circulation was analyzed in the period from December 2001 to November 2008. The newspapers were scanned via the online media database “factiva” (http://global.factiva.com) for articles dealing with the keywords “OECD” and “PISA”. The national quality newspapers analyzed (in parentheses) were: Australia (The Australian), Austria (Die Presse), Belgium (De Standaard), Canada (National Post), Czech Republic (Mladá fronta Dnes), Denmark (Politiken), Finland (Suomen Tietotoimisto), France (Le Monde), Germany (Süddeutsche Zeitung), Great Britain (The Times), Ireland (Irish Independent), Italy (Corriere della Sera), Mexico (Reforma), New Zealand (The Press), Norway (Bergens Tidende), Poland (Rzeczpospolita), Portugal (Journal de Notícias), Republic of South Korea (The Korea Herald), Spain (El País), Sweden (Svenska Dagbladet), Switzerland (Neue Zürcher Zeitung), USA (New York Times). *In the cases of Mexico and Ireland the respective newspapers were only available from January 2003 (Ireland) and May 2004 (Mexico) onward.

hind other industrialized countries in matters of education would be the starting point for a public debate. But this argument does not hold true empirically. A closer look shows that countries which are “neighbors” in the PISA league table are not inevitably interchangeable in respect to their media reception (see fig. 2). Although poor results seem to increase the education issue’s likelihood of becoming a substantial object of public discourse, as one can easily conclude regarding the cases of Germany, Spain, Austria, and Mexico, this phenomenon is not universally valid: In other poor performing countries, like the U.S., Norway, Poland, and Portugal, a position at the lower end of the PISA league table has not led to a public outcry reflected in the media coverage. As an overall trend, the worse the rank, the more differentiated the reactions are. The same holds true for those countries which did well in the PISA study: With the exception of the Finnish case, education policy has generally not become a hot topic in the media discussion of countries that are among the “PISA winners” ranked at the top of the league table. Performing well does not entail the need for further analysis or discussion but performing poor does not necessarily attract attention either. PISA did not trigger a great media reaction in countries such as Canada, South Korea or New Zealand

Figure 2: PISA ranking related to media coverage[5]



[5] For each country the average rank of PISA 2000, 2003, and 2006 is displayed. Additionally, in each of the three circles the average of the three sub-disciplines (reading, mathematics, and science) was assessed.

which are the top two to four performers in the study. Hence, the rankings alone seem to have only a very restricted influence on the extent of public debate, and other mechanisms must account for explaining the linkage between PISA and national reactions.

How then can the differences in reaction to the OECD's PISA study be explained? When do R&Rs lead to severe public responses and subsequent reforms? In the following section, we set up a parsimonious theoretical framework for assessing the impact of R&Rs. We argue that there are basically two ways in which R&Rs exercise impact on national policy-making: when a topic is framed as crucial for state purposes and if the gap between national self-perception and empirical evidence is too substantial to be ignored.

3. GOVERNANCE BY COMPARISON – A THEORETICAL APPROXIMATION

Since the 1990s, comparative depictions have become prominent phenomena in modern globalized societies. In particular, comparative illustrations in the form of R&Rs increasingly occur in many policy fields and issue areas. Often, the outcomes of such R&Rs influence future decisions of states, market actors, and individuals alike. For example, states are ranked by the non-governmental organization Transparency International for their level of corruption, and foreign aid at times depends on such evaluations. Companies are eager to find themselves ranked at the top of the FTSE4 Good Corporate Responsibility Index to ensure the confidence of buyers and investors. In the academic world, universities – in the United Kingdom, for example – are ranked for their scientific achievement during Research Assessment Exercises whereas their score may determine future funding. Comparisons of this kind are popular techniques in modern societies because they provide succinct information in a short period of time and in an easily digestible way: without being an expert in the specific field, anybody can understand that being ranked #5 is different than #77, or that a positive value in any rating differs from a negative.

Thus, R&Rs are far from new. What is fairly new, however, is their application and continuous expansion in academic fields and to political institutions. Today, R&Rs are used for states, state performance in certain policy fields, and state institutions. The results have become very popular and often initiate public debates, as R&Rs create an air of competition around performance or policy by attributing relative positions. Whereas many comparative analyses are applied without the agreement of those being ranked (for example, the corruption index), there are other forms of rankings and ratings to which explicit consent must be given.

Furthermore, comparison as a form of governance implies a scientific approach to political decision making. The most effective (rationalist) or most appropriate (sociological institutionalist) decisions should be established through objective criteria and

evaluation (March and Olsen 1998). The parties evaluated are implicitly pressured to converge towards those practices, forms of organization or behavior that are regarded as *best* (either most effective or most appropriate) in line with the specific criteria of the respective framework of comparison. From a governance viewpoint, comparisons thus illustrate that power is not wielded solely by traditional regulatory activities, but also by such *soft* comparative ratings and rankings as those of IOs (Abbott and Snidal 2000). These soft forms of governance by comparison emerge in the transnational sphere and have the potential to influence established institutionalized practices at the national level by creating standards and establishing “best practices” which then produce pressure to improve. They are representative for the shift from government to governance in the field of education (Grek 2009).

In this paper, we claim that IOs such as the OECD gain a powerful position when given the tool of governance by numbers (Miller 2001). That is to say, even if the OECD were merely given the task of producing information to guide state-based decision making, the direction of the actual implementation would nevertheless remain outside the hands of states, which may have a guiding impact on policymakers. From a more general point of view, the case study of the OECD and its PISA study shows that IOs that generate and publish comparative evaluations gain power by setting standards independent of the original motives states had for delegating them the task of objective evaluation (Martens 2007). That is, even if R&Rs may originate from a rationalist approach (states seeking effective policy-making and a reduction of transaction costs), these procedures can develop further within the framework of an international organization that ultimately creates its own standards of performance. In doing so, an IO like the OECD takes over the task of comparison and evaluation, applies procedures in its own way and establishes new internationally shared standards (thereby adhering to a sociological institutionalist logic) which were unforeseen by states (Barnett and Finnemore 1999; 2004).

But why do R&Rs then have such different impacts on states that have been ranked similarly? For the purposes of this paper, we refer to the two extreme cases of public and political reactions to the OECD’s PISA study: Germany and the U.S. Both countries ranked on the lower end of the PISA league table. But whereas in Germany the often-cited PISA shock occurred and subsequent multiple education reforms were launched, no such thing happened in the U.S. – there was neither a public debate nor any related reform attempt. Thus, the measurement as such is not the decisive matter for response; what counts is how the results are perceived (Van Dooren 2008). Derived form the theories of rationalism and sociological institutionalism, the two dimensions of the national *framing of the issue as problematic* and the relation between *self-perception and actual*

results offer explanatory insight for understanding why states react differently to similar results in education R&Rs.

3.1 Framing an issue as problematic

First, from a sociological institutional interpretation inspired by constructivist views, participation in international comparative R&Rs and their influence should result from perceptions shared between states regarding the purpose of education policy and regarding their role with respect to citizens and other states. Thus, a sociological institutional approach asserts that willing submission to comparisons with peers is not the result of a functional, utilitarian calculation; the main purpose is rather – putting the emphasis on the emergence of normative structures of social interaction – the diffusion of a specific practice or mode which influences the behavior of actors (see, for example, Meyer and Ramirez 2000), especially of those being compared. From this point of view, the objective of R&Rs is to establish normative criteria for appropriate behavior. Moreover, IOs such as the OECD are themselves seen as actors capable of producing shared norms, values and standards (Finnemore 1993). Especially when the membership is homogenous – as in the OECD to which only highly industrialized countries are admitted – norms and standards receive a high degree of communality (Rittberger and Zangl 2006).

Drawing on this explanation as to why states participate in international R&Rs, the potential effect of R&Rs on national policy-making can be derived. In order to have an impact it is necessary that the issue under evaluation be nationally perceived and discursively constructed as crucial. Education is generally not framed as an end in itself. It is usually associated with certain superordinate state objectives like micro- and macroeconomic prosperity, technological and cultural advance, or even security policy (Martens and Weymann 2007). Hence, a poor performance in education can be discursively linked to anticipated future disadvantages in other domains. The central point here is that pressure for improving national education performance can only emerge from poor results in an international comparison if the topic “education” is linked to another issue that is nationally perceived and framed as crucial for general state objectives. For example, the issue of education is nowadays increasingly understood against the background of economic performance. Thus, bad results in PISA which indicate a general low quality of the national education system are equated with a risk to overall economic prosperity. When overall economic performance is framed as crucial, education policy becomes framed as crucial as well, and the need for improving education quality is seen as imperative.

On the other hand, if poor results in education performance are not evaluated in the light of crucial issues or linked to a policy field that is highly relevant, the pressure for

reforms is not present to a comparable extent. If PISA results are not linked to an endangerment of the economy because great variation in results among students is not interpreted as critical as long as a highly educated elite remains, it is also unlikely that matters of education will cause encompassing national reactions. But not only linkages to the economic dimension can be connected to education. Matters of security can also be linked to education, like in the U.S. during the Cold War when the discovery of poor performance in the education sector was directly translated into a threat to national security and triggered pressure to improve the education system.

In brief, R&Rs only have an impact regarding how the issue at stake is framed: if an issue is challenged by linking it to a relevant national topic, national reform pressure emerges. From the constructivist point of view, IOs thus act as standard setters and can play an important role in shaping the understanding of an issue as crucial and linking it to other policy fields. If IOs successfully reveal and promote the linkage between certain issues, they might re-frame the national understanding of a topic and alter its domestic evaluation. Among other things, the impact of PISA on Germany can be attributed to the OECD's emphasis on the connection between education and economic performance.

3.2 Substantial gap between self-perception and results

Second, from a rationalist point of view, it should be argued that comparative R&Rs reduce transaction and information costs in liberal societies and liberalized markets (see for example, Moravcsik 1993). The need for better and more efficient performance on the part of education systems as the result of increasing demands on labor forces operating within a global market cannot be solved individually by states. As such, comparative studies are an instrument for displaying different solutions to a particular problem. Only an international comparative analysis reveals the "best practices" available which can then be copied by poorer performers. Thus, states approach IOs such as the OECD with the task of evaluating their education system from a comparative perspective. They are considered suitable bodies with the capacities needed for developing and applying objective criteria in order to conduct an international assessment between different states.

To be evaluated does not mean that a self-impression is created simultaneously by the specific ranking. Overall, R&Rs themselves basically do not constitute a whole new perception of an issue. Rather, preexisting self-impressions regarding expectations about the performance collide with empirical results of R&Rs; or in other words, the empirical results do not lead to what Espeland and Sauder (2007) called a "self-fulfilling prophecy" – on the contrary. Generally, if a divergence between expected results and empirical findings is revealed, R&Rs might lead to a re-evaluation of the self-

impression regarding the evaluated topic or trigger reform pressure in order to meet the expectations in the future.

In this context, different interrelations between the self-impression and the actual results are conceivable whereby each entails different consequences for creating pressure to improve in the evaluated discipline. For example, the empirical findings are better than expected. In such a setting no pressure occurs at all. Also, the results of an evaluation match the anticipated performance. Meeting the expectations does not mean that pressure for improvement is absent but rather that the possibility of the occurrence of a “shock” event is not very likely.

Moreover - and this is the significant dimension - the gap between self-impression and empirical evidence can be negative: the performance is worse than expected. In this case, the awareness of the findings might cause a shock that also elicits high pressure for improvements. The case of Germany illustrates this point, where the broad public (including most policy-makers) was deeply convinced that the German education system was among the best in the world. The results of the PISA report dramatically contrasted with this self-perception by showing the deficits in comparison to other industrialized countries.

In brief, R&Rs are most likely to have an impact if they reveal a negative gap between self-perception and actual results. From the perspective of the rational institutionalist reasoning of increasing effectiveness, comparative evaluations conducted by IOs can reveal externally if a country is lagging behind and what measures for improvement have to be taken into account. In addition to its achievement in framing education as a crucial issue, the OECD also successfully managed to reveal a mismatch between performance and national self-perception. In terms of effectiveness, PISA established benchmarks for education systems and showed that some education systems require encompassing reforms to meet the standards of other industrialized states.

Summing up the arguments of the theoretical model, an issue which receives a performance evaluation contrary to the expected results (in a negative sense) also has to be defined as problematic in order to trigger high pressure for improvements and reforms. For instance, if in an international comparison the issue X of country A is ranked on the lower end of a league table, it only causes high pressure for improvements if a negative gap between rank and self-impression is existent and at the same time X is defined as a crucial factor of country A.

Taken together, R&Rs by themselves are not automatically able to exert influence on national policy-making. The mechanisms of self-perception and framing substantially shape how R&Rs are received on the national level and whether they have consequences.

4. TWO SIDES OF THE SAME (PISA-) COIN – ASSESSING THE CONTRARY REACTIONS OF GERMANY AND THE U.S.

In the light of the theoretical approximation, the differential impact of PISA can be descriptively illustrated by referring to the two extreme cases of poor performing countries identified in the media analysis – Germany and the U.S. While Germany significantly changed its national education policy-making due to its negative PISA performance and as a consequence of the emerged public discourse, the U.S. did not seem to be bothered by PISA, neither medially nor politically. Unlike Germany, the U.S. neither experienced an education shock through PISA nor did it frame the issue of education as problematic.

4.1 Germany – the rediscovery of education policy-making due to the PISA shock

In Germany, the publication of the first PISA results in late 2001 led to the often cited “shock” by pinpointing Germany’s status as a laggard – and not, as expected, as a leader – in the quality of education compared to its industrialized peer countries. Education instantly became one of the most frequently discussed themes in Germany. OECD testing items have since received high coverage in the media, and the term “PISA” itself has become a synonym for testing, ranking, and rating. Furthermore, the era in which education was not a political issue relevant for winning elections came to an end.

The intensity and extent of the new education debate in Germany after PISA was quite astonishing, however. Even though previous international comparative studies concerning education performance in the mid 1990s[6] had already sensitized at least the elites in education policy-making to the mediocrity of German students in mathematics and sciences (Heinze 2002: 18), the broad public was not aware of the situation. Thus, the publication of PISA finally “led to a public outcry in Germany” (Ammermüller 2004: 2). The media picked up the PISA report and subjected it to a general debate about the German education system itself and the necessity for far-reaching reforms. Contrary to previous impulses, PISA managed to put the issue of education back on the public and political agenda. Generally, with the double-edged process of evaluating the national system and comparing it to other countries, the OECD created an immense pressure on German policy-makers to improve the education system. However, this pressure did not directly impact the political level but rather mobilized the broad public which in turn called for far-reaching reforms.

Most importantly, the first PISA report revealed that the performances of German students at age 15 were significantly below the OECD average in all areas of academic competence (reading, mathematics, and science) (Baumert et al. 2001; Kiper and Katt-

[6] For instance the Trends in International Mathematics and Science Study (TIMSS).

mann 2003: 15-17).[7] In addition to the poor general results, it also became obvious that Germany is one of the OECD countries with the highest level of performance variation across students. In no other industrialized country is academic success so strongly determined by the socio-economic or migration background of the respective child as in Germany (Kiper and Kattmann 2003: 32; Ertl 2006: 620). This means that the linkage between the social background and school performance is exceptionally high. Compared to other OECD countries, the German educational system seems to be unable to reduce existing social inequality (Loeber and Scholz 2003: 246). Hence, the aim of education to provide every child – independent of his/her socio-economic background – with basically the same opportunities for academic success and advancement becomes more or less nullified. This aspect successively became a fundamental characteristic of the German PISA debate. Although Germany improved slightly in the PISA studies of 2003 and 2006 (Prenzel et al. 2005; Prenzel et al. 2007), it was still far from ranking among the top countries in education quality.

However, the central question still remains: why did the PISA results surprise Germany to such an extent and cause such a shock? Overall, the field of education policy had not been a prominent political issue in Germany until the mid 1990s when almost two decades of non-reform and disregard came to an end as education was carefully rediscovered as an issue for policy-making. Beforehand, far-reaching and encompassing education reforms had been introduced in the late 1960s and early 1970s after the national study by Georg Picht (“*The German Education Catastrophe*”, 1964) revealed huge deficits in the German education system. Back then, education policy was highly politicized and subject to widespread public attention accompanied by extensive reform projects. Later on, the predominance of other policy issues, such as foreign and security policy as well as labor market policy, marginalized education policy to the extent that it became an issue dealt with mainly among experts. A linkage between education and other policy fields was not existent.

Moreover, German education policy-making was resistant to external shocks related to comparative studies in the period from the late 1970s to the mid 1990s because Germany simply did not participate in any such international evaluations which could have indicated any deficits and disadvantages. At that time, a tradition of evaluating the education system or relating educational performances to other countries simply did not exist. The self-perception of being at the top level in educational performance was not supplemented by empirical evidence but rather perpetuated by a historical understanding which regarded German educational capability as outstanding.

---

[7] Overall, Germany was ranked (averaged) 20th in PISA 2000, 15th in 2003, and 13th in 2006 (which was mainly due to an 8th rank in the sub-discipline of science).

Ultimately, in 2001 PISA finally demonstrated Germany's laggard status in education performance and German policy-makers were hence inevitably faced with the shortcomings of the education system compared to other industrialized countries as well as with a public demand for rapid improvement of the quality of education. As a consequence, the new century heralded a new enthusiasm for educational reform. Nearly all aspects of education were closely scrutinized and evaluated. Even the "sacred cow" of the German educational system, its three-tiered school system, was scrutinized, questioned, and, in some of the German *Länder*, reformed.[8] Furthermore, the liberal education tradition in Germany was challenged by the ostensible economic attitude PISA promoted (Niemann 2010).

In light of the negative empirical PISA results in December 2001 – the same month that PISA was officially publicized and almost three decades after the last comprehensive education reforms – the responsible German policy-makers in education agreed within the Standing Conference of the Ministers of Education and Cultural Affairs (KMK) upon an action plan which would provide a framework for substantial reforms and guide the long range of reforms in the *Länder*.[9] Identified as being one of the blind spots of Germany's education system and having had a substantial influence on poor PISA performances, early education received increasing attention in order to create a better basis for further education. Thus, the political discourse was not restricted exclusively to secondary education, and the German *Länder* have since embarked on a strategy to reform early childhood education in order to counterweight the socio-economic background of children (Carey 2008: 17-21).

Also, measures of quality assurance are being enhanced on the basis of binding standards and output-oriented evaluation. In order to improve the quality of education, in 2002 the KMK agreed upon the introduction of educational standards and the establishment of a central agency for monitoring compliance with these standards (Nieke 2003: 201). This new development was also accompanied by an increased implementation of comparative tests and evaluation criteria in schools. Due to the introduction of education standards, academic and school performances can now be better scrutinized and hence a stronger emphasis is placed on the output dimension of education today.

In essence, the OECD not only successfully promoted its ideas and recommendations concerning education but also influenced Germany on the level of policy-making to adapt OECD instruments for assessing school performance. This, in turn, also shaped

---

[8] It is important to note that the German *Länder* are qua Basic Law responsible for matters of education.

[9] The proposed reform measures were not "invented" within the few days after the publication of PISA. In fact, compiled reform schemes already existed, though eventually through PISA they made it on the agenda (Tillmann et al 2008: 379).

the way education policy-making is generally conducted nowadays in Germany. Due to the shift toward output orientation in education policy-making stimulated by the OECD's PISA study, political stakeholders increasingly rely on scientific advice and consult experts in order to identify the paramount problems and develop solution strategies. The already existing but rudimentary debates about quality development, quality assurance, matters of teachers' education, and curricular issues were increasingly fostered by PISA. Today, education policy-making does not simply focus on the fulfillment of certain principles but instead integrates the new governance mode of "evidence-based policy-making". In general, policy-makers are supported by scientific advisors and can draw upon their expertise. For instance, the KMK has introduced a scientific advisory council, which provides expertise regarding implications derived from comparative studies.

Not only international comparisons became a focal point of interest. Assessments of the performances of the German *Länder* were also increasingly subject to evaluation. The supplement PISA-E study compares and evaluates the academic performances among the individual *Länder*. Within a very short period of time, Germany established a culture of comparison in education policy-making. The changes on the structural level which can be evaluated against the background of OECD activities encompass the empirical turn that emphasizes the role of evaluation, including output orientation and changes in school settings, in order to overcome the highlighted weaknesses of education (e.g. dependence of academic performance on students' socio-economic backgrounds).

Taken together, the conducted reforms in Germany's education system generally reflect a paradigmatic shift that comprises the enhanced orientation towards an output perspective on education, as promoted by the OECD, and abandons the previously predominant input dimension: Instead of primarily focusing on the investments in education, the prevailing question is now: "What does the education system produce?" For example, even the ideas and discursive framing regarding education changed substantially in Germany due to the impact of OECD's PISA surveys. Whereas prior to the recent debate education was first and foremost seen as a microeconomic issue (which focuses on training individuals for a profession) and as a citizenship issue with respect to self-fulfillment, the post-PISA debate is increasingly characterized by shifting the emphasis to the macroeconomic dimensions of education. The interpretation that education is central to the economic performance of a whole country and that poor performance in education also might entail the danger of jeopardizing future economic prosperity was further developed and became successively accepted. Bad education quality is equated with bad economic performance and a risk of a declining national welfare and prosperity.

4.2 The U.S. – or how a country disregards its PISA results

In the U.S., by contrast, the three cycles of the PISA study and their results have remained virtually unheard of. The general public and the media did not take notice of this study or its results—despite the fact that the U.S. performed as badly as Germany, being ranked below average among participating countries. In fact, the U.S. scored worse in the 2006 study than it had in the previous two studies: in the latest PISA 2006 study, the U.S. only ranked 24th in mathematics among the 30 OECD member states. However, despite these poor results, there have been few reactions in the broader public. Even when extending the media analysis PISA is not an issue in the U.S: of eight major U.S. daily and weekly newspapers, only eight articles over the period of 2001 to 2008 actually covered PISA at all—and of these eight articles only three deal with the poor results of U.S. students.[10] As a comparison, during the same time period the German daily newspaper *Süddeutsche Zeitung*, for example, published 253 articles. Why did the PISA study not have the same effect in the U.S. as it did in Germany? Or to put it differently: why hasn't there been a PISA shock in the U.S.?

Part of the answer is that the poor PISA results are in fact nothing new for the U.S. Ever since the Soviet Union launched the first satellite in 1957, education policy has been an issue on the political agenda. In a way, Sputnik demonstrated the Soviet Union's technological superiority over the U.S. or, at the very least, its equality in the field. Sputnik also showed that the Soviet Union possessed strong intercontinental missiles and could threaten the U.S. with its atomic bombs. This technological capacity challenged the West's claim of pre-eminence which had so far been thought to be secure.[11] The reasons for this Western "lag" were primarily found in the education system. According to experts, the main problem was the prevailing circumstances in schools which excluded too many people from participating in social progress.

The Sputnik shock triggered a crisis in the self-image of Americans who had naturally considered themselves inhabitants of the technologically most progressive nation on earth. Democracy and capitalism had been widely considered as natural competitive advantages accounting for the U.S's technological superiority. The fact that the communist Soviet Union with its planned economy was now one step ahead of the U.S. in terms of space-related issues was deeply shocking to Americans. As a result of this

---

[10] The newspapers examined include: New York Times, Newsweek, International Herald Tribune, Washington Post, USA Today, The Wall Street Journal, and Los Angeles Times. The research was also conducted with "factiva", again with the keywords 'OECD' and 'PISA'.

[11] As an immediate consequence of this Sputnik shock the U.S. intensified its endeavors to gain pre-eminence in the technological race into the universe. It accelerated the Western missile programs and led to the formation of NASA.

shock, the majority of the American population rapidly accepted the demand for a basic school reform intended to help the U.S. succeed over the Soviet Union in their technological race into the universe. Thus, the Sputnik shock led to a comprehensive reform of the American education system. The natural sciences seemed to be particularly in need of reform since the Soviet Union trained twice or even three times as many engineers as the U.S. For this reason, U.S. president Dwight D. Eisenhower, who had labeled education policy as more important than missile production in one of his speeches, introduced the *Federal-aid-to-Education-Program*. The government provided $1.6 billion for this program which financed the reform of the education system over a period of four years. The money was used to quadruple the National Science Foundation's annual budget to 134 million dollars, grant 20,000 scholarships, improve the teacher training, and finance the construction of new school buildings.

With these measures, new emphasis was placed on the inclusion of underprivileged classes in the hope of discovering new resources for the education system. One of the measures for opening up these resources was the introduction of preschools. New school busses were to integrate children from remote areas into the education system. Furthermore, curriculums were restructured, and courses which had previously dealt with housekeeping or direct vocational education, were substituted by math, physics, and chemistry. However, the advancement of humanities, such as political and historical science or philology, was also part of Eisenhower's program. It aimed to generate wise leaders who could apply the technological achievements for the collective good of the American nation. John F. Kennedy and Lyndon B. Johnson further advanced the introduction of television education programs, the end-to-end integration of education institutions and libraries in order to grant better accessibility to education, as well as the *New-Math-Program* which was to familiarize children with abstract mathematics from very early on.

In the early 1980s, in the U.S. the government then enlisted a national commission to conduct a study of the American educational system 25 years after Sputnik to review the progress made since. However, the resulting report of 1983, "A Nation at Risk: Imperatives for Educational Reforms", found devastating outcomes: with 23 million adults and 17 percent of juveniles being illiterate, the American educational system was abysmal. The report triggered broad public concern about the quality of education, identifying the need for the monitoring of schools, standards, and teachers. The resulting "Great School Debate" (Gross and Gross 1985) soon attracted the attention of the highest political officials in the country: during the height of the Cold War, President Ronald Reagan and political stakeholders considered this state of affairs a national security risk and made school reform a top priority of his administration. Since then, education matters have been on the top end of each presidential administration's list of reforms. In

fact, the 2008 presidential elections are the first since the 1980s where other issues, in particular the war against terrorism and the state of the economy, outranked education in terms of importance.

In a way, the U.S. thus had its “education shock” earlier than Germany. And, unlike Germany, which was hit by an external international organization, the U.S. “shock” was triggered by the perception of education as a problematic issue and a question deeply linked to the nation’s security. Since the 1980s, a series of educational reforms aiming at creating standards, providing for accountability, enabling school choice, and improving the quality of teaching have been initiated and implemented. Most importantly, these reforms have lead to outcome-oriented types of educational teaching, based on the belief that setting high standards and establishing measurable goals can improve individual results in education.

The latest of these reforms is the current No Child Left Behind Act (NCLB) of 2002.[12] The law reauthorized a number of federal programs aiming to improve the performance of U.S. primary and secondary schools by increasing the standards of accountability for states, school districts, and schools, as well as providing parents more flexibility in choosing which schools their child can attend. The Act requires states to develop basic skills assessments to be given to all students in certain grades if those states want to receive federal funding for schools. However, NCLB does not assert a national achievement standard; instead, standards are set by each individual state, in line with the principle of local control of schools.[13] Upon the implementation of this law, Congress increased federal funding of education: No Child Left Behind received an increase from $9.7 billion in 2001 to $25.0 billion in 2007 (Kosar 2005: 191).

Thus, within the domestic sphere, the U.S. has taken a path toward more standardization and comparative evaluations, but international comparisons such as PISA remain unnoticed. Whereas ratings and ranking within the U.S. play a decisive role today as regards school choice, the international standing of the U.S. as a whole does not receive much attention. The irony, however, is that it was in fact the U.S. that gave the incentive for the PISA study, which ended up having a far greater impact on other countries’ education policies than on the U.S.’ policies. When “A Nation at Risk” hit the U.S., the Reagan administration not only initiated domestic reforms; it also wanted international comparative data on the state of education in the industrialized world in general. The

---

[12] For details how this act came about, see Kosar (2005: chapter 6).

[13] The Tenth Amendment to the United States Constitution specifies that powers are not granted to the federal government; education powers are reserved for the individual states. However, NCLB is seen as the greatest increase in federal power of education since the Elementary and Secondary Education Act of 1965. See, for example, McGuinn (2006). For the evolution of the federal role in school education, see McGuinn (2005).

idea for an international comparative study was then brought to the OECD. Being the major financial contributor to the OECD's budget for education, the U.S. pushed the international organization to significantly change its work in education and compile internationally comparable statistical data. In response to tremendous pressure from the U.S., the International Indicators of Educational Systems (INES) project was established in 1988 and produced regular publications on education indicators, of which the PISA study is the latest offspring.[14]

In brief, despite the reforms of the last 25 years since "A Nation at Risk" was published, schools in the U.S. have improved only to a limited extent – despite high intensions and great financial means. The various low PISA scores for the U.S. did not trigger a shock to the country, since it was basically already known that the U.S. education system needs improvement. Thus, unlike the German case, there was simply no gap in the U.S. between the self-image of the country concerning the performance of its secondary education systems and the actual results in the PISA study. In a way, it would have instead been a positive shock for Americans had the U.S. had performed better in the PISA study. Moreover, secondary education is currently not framed in the context of another issue at stake for the nation, whereas in the late 1950s and beyond, education was interpreted as a policy field of relevance for the national security.

5. CONCLUSION AND OUTLOOK

In order to explain differences in national reactions (including media coverage and follow-up political reforms) to PISA, it is not sufficient to simply refer to a nation's actual position in the ranking. As our media analysis has shown, there is no obvious correlation between the rank of a state and its reaction. In fact, even states that are evaluated similarly vary substantially regarding their national responses. In other words, a poor ranking is not sufficient to trigger a national education debate and initiate reforms. Therefore, to explain the power such numbers can exert, we need an approach which goes beyond a simple rank-reaction model. In our paper, we conceptualized R&Rs with respect to a theoretical approach which we called governance by comparison. We argued that two dimensions need to be considered when seeking to understand under what conditions R&R have an impact on national policy-making, namely that an issue – in this case, education policy – needs to be framed as problematic within the national context and that the relation between the self-perception and the actual results need to be taken into account.

In the empirical part, we applied our theoretical concept against the two countries which were on average ranked similarly poorly in the OECD's PISA study but reacted

[14] For more details on the OECD's role in education, see Martens (2007).

oppositely in the most extreme manner: Germany with the greatest public debate about education and the U.S. with the least public debate. Our analysis showed that Germany witnessed its PISA results as different to its self image as an education nation (self-perception) in times of the knowledge-based society in which education became a significant ‘natural resource’ (problematic framing). In particular, the poor results demonstrated problematic aspects of the German education system such as fostering social inequality and system inefficiency in integrating children of migrant families. In contrast, in the case of the U.S., PISA did not trigger any debate about reforming the education system as the study did not bring forward any new information for the country about the poor quality of secondary education (self-perception). In addition, education was framed as a security issue during the Cold War (problematic framing), an external condition which is no longer given.

Our explorative theoretical framework is an attempt to find an explanation for puzzling different responses to comparative rankings. The next step would be to establish causality for our described phenomena. In further research, it would need to be evaluated against additional arguments. First, both states possess different political cultures as regards their reception of IOs and their governance. The U.S. perceives itself much more as independent and does not feel obliged to international commitments. From this point of view, it is not surprising that the American secondary education sector is not affected by the PISA survey as conducted by the OECD. On the other hand, it was the U.S. that initiated PISA and could therefore be expected to recognize its significance and react to it. In addition, the more that competences for education are located on the local or private level, the less should international initiatives bother the country as a whole. While in Germany, state authorities (on the level of the *Länder*) are responsible for education policy-making, the U.S. system is organized much more locally and privatized. PISA, however, assesses national education performance, and as a result, it is difficult to clearly transfer poor results to the local level in the U.S. In contrast, poor PISA results of German students impacted the education system as such. Second, education policy generally seems to be a second order policy field that is not perceived to be as fundamental as say, for instance, security and the economy. That means that if other prominent issues are more relevant than educational matters, first order issues take precedence. Perhaps the PISA results were ignored in the U.S due to the prevalence of security issues after September 11th and subsesequent problems related to the economic recession. In Germany, those topics were not as important, and thus there were fewer distractions away from the focus on education.

Beyond the actual effects and non-effects PISA exerted on both states, our study provides insights about the authority the OECD (or similar expert organizations) can have with the governance instrument of comparison. PISA is an example of the varia-

tion in the reception of international stimuli at the national level. As shown in the case of the U.S., PISA more or less stated the already obvious, and thus its technique of governance by comparison was not perceived as a groundbreaking revelation of new phenomena; in contrast, in Germany, the comparison told a “new story” and provided essential insights. As a precondition, an IO needs to be perceived as a legitimate authority and expert organization in order to be successful in influencing policy-making at the nation state level. Our analysis of the German and the U.S. reactions to the OECD’s PISA study can, however, only be a first step in grasping the influence of governance by comparison can have. It would be a promising endeavour to compare reactions to PISA in other OECD countries regarding the concepts of self-perception and framing in light of their respective reactions. Expanding the scope of governance by comparison to other policy fields beyond education would be expedient in clarifying to what extent the power of the OECD (or other IOs) depends on its expert status.

REFERENCES

Abbott, Kenneth W. and Duncan Snidal (2000) 'Hard and Soft Law in International Governance', *International Organization* **54**(3): 421-56.

Ammermüller, Andreas (2004) 'PISA: What Makes the Difference?', *Diskussionspapiere der DFG-Forschergruppe* **04/07**: 1-30.

Barnett, Michael and Martha Finnemore (2004) *Rules for the World. International Organizations in Global Politics,* Ithaca: Cornell University Press.

Barnett, Michael and Martha Finnemore (1999) 'The Politics, Power, and Pathologies of International Organizations', *International Organization* **53**(4): 699-732.

Baumert, Jürgen, Petra Stanat and Anke Demmrich (2001) 'PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie', in Deutsches PISA-Konsortium, ed. *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*, 15-68, Opladen: Leske + Budrich.

Carey, David (2008) 'Improving Education Outcomes in Germany', Economic Department Working Papers, OECD.

Ertl, Hubert (2006) 'Educational standards and the changing discourses on education: the reception and consequences of the PISA study in Germany', *Oxford Review of Education* **32**(5): 619-34.

Espeland, Wendy Nelson and Michael Sauder (2007) 'Rankings and Reactivity: How Public Measures Recreate Social Worlds', *American Journal of Sociology* **113**(1): 1-40.

Finnemore, Martha (1993) 'International organizations as teachers of norms: the United Nations Educational, Scientific, and a Cultural Organization and science policy', *International Organization* **47**(4): 565-97.

Grek, Sotiria (2009) 'Governing by numbers: the PISA ‘effect’ in Europe', *Journal of Education Policy* **24(**1): 23-37.

Gross, Ronald and Beatrice Gross, eds. (1985) *The Great School Debate,* New York: Touchstone.

Hawkins, Darren G., David A. Lake, Daniel Nielson and Michael J. Tierney, eds. (2006) *Delegation and Agency in International Organizations*, New York: Cambridge University Press.

Heinze, Rolf G. (2002) 'Die Bildungskatastrophe - Ein alter Hut?', in Bernd Fahrholz, Gabriel Sigmar and Peter Müller, eds. *Nach dem Pisa-Schock. Plädoyer für eine Bildungsreform*, 16-21, Frankfurt am Main: Hoffmann und Campe.

Henry, Miriam Bob Lingard, Fazal Rizvi and Sandra Taylor (2001) *The OECD, Globalisation and Education Policy,* Oxford: Pergamon.

Joachim, Jutta, Bob Reinalda and Bertjan Verbeek, eds. (2008) *International Organizations and Implementations*, New York: Routledge.

Kiper, Hanna and Ulrich Kattmann (2003) 'Basiskompetenzen im Vergleich - Überblick über Ergebnisse der PISA-Studie 2000', in Barbara Moschner, Hanna Kiper and Ulrich Kattmann, eds. *PISA 2000 als Herausforderung,* 15-37, Baltmannweiler: Schneider Verlag Hohengehren**.**

Kosar, Kevin R. (2005) *Failing Grades. The Federal Politics of Education Standards*, Boulder and London: Lynne Rienner Publisher.

Loeber, Heinz-Dieter and Wolf-Dieter Scholz (2003) 'Von der Bildungskatastrophe zum Bildungsschock - Zur Kontinuität sozialer Benachteiligung durch das deutsche Bildungssystem', in Barbara Moschner, Hanna Kiper and Ulrich Kattmann, eds. *PISA 2000 als Herausforderung,* 241-85, Baltmannsweiler: Schneider Verlag Hohengehren**.**

March, James G. and Johan P. Olsen (1998) 'The Institutional Dynamics of International Political Order', *International Organization* **52**(4): 943-969.

Martens, Kerstin (2007) 'How to become an Influential Actor – the ‘Comparative Turn’ in OECD Education Policy', in Kerstin Martens, Alessandra Rusconi and Kathrin Leuze, eds. *New Arenas of Education Governance – The Impact of International Organisations and Markets on Education Policy Making*, 40-56, Houndsmills: Palgrave Macmillan.

Martens, Kerstin and Ansgar Weymann (2007) 'The Internationalization of Education Policy - Towards Convergence of National Paths?' in Achim Hurrelmann, Stephan Leibfried, Kerstin Martens and Peter Mayer, eds. *Transforming the Golden Age Nation State*, 152-172, Houndsmills: Palgrave Macmillan.

McGuinn, Patrick (2006) 'Swing Issues and Policy Regimes: Federal Education Policy and the Politics of Policy Change', *Journal of Policy History* **18**(2): 205-240.

McGuinn, Patrick (2005) 'The National Schoolmarm: No Child Left Behind and the New Educational Federalism', *Publius* **35**(1): 41-68.

Meyer, John W. and Francisco O. Ramirez (2000) 'The World Institutionalization of Education' in Schriewer, Jurgen, ed. *Discourse Formation in Comparative Education, 111-132,* Frankfurt am Main: Peter Lang.

Miller, Peter (2001) 'Governing by numbers: why calculative practices matter', *Social Research* **68**(2): 379-396.

Moravcsik, Andrew (1993) 'Preferences and Power in the European Community: A Liberal Intergovernementalist Approach', *Journal of Common Market Studies* **31**(1): 473-524.

Nieke, Wolfgang (2003) 'Schulreform: Wie reagiert die Fachpolitik auf bildungswissenschaftliche Bestandsaufnahmen? ' in Toni Hansel, ed. *PISA - und die Folgen? Die Wirkung von Leistungsvergleichsstudien in der Schule*, 197-214, Herbolzheim: Centaurus Verlag.

Niemann, Dennis (2010) 'Turn of the Tide – New Horizons in German Education Policy through IO Influence' in Martens, Kerstin, Alexander Nagel, Michael Windzio and Ansgar Weymann, eds. *Transformation of Education Policy – the Impact of the Bologna Process and the PISA Study in Comparative Perspective,* Houndsmills: Palgrave Macmillan.

Papadopoulos, George (1994), *Education 1960-1990. The OECD Perspective,* Paris: OECD.

Picht, Georg (1964) *Die deutsche Bildungskatastrophe: Analyse und Dokumentation,* Olten and Freiburg: Walter-Verlag.

Prenzel, Manfred, Cordula Artelt, Jürgen Baumert, Werner Blum, Marcus Hammann, Eckhardt Klieme and Reinhard Pekrun (2007) *Pisa 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*, Münster and others: Waxmann.

Prenzel, Manfred, Jürgen Baumert, Werner Blum, Rainer Lehmann, Detlev Leutner, Michael Neubrand, Reinhard Pekrun, Jürgen Rost and Ulrich Schiefele (2005) *PISA 2003. Der zweite Vergleich der Länder in Deutschland - Was wissen und können Jugendliche?,* Münster, New York, München, Berlin: Waxmann.

Reinalda, Bob and Bertjan Verbeek, eds. (1998) *Autonomous policy making by international organizations*, London: Routledge.

Rittberger, Volker and Bernhard Zangl (2006) *International Organization. Polity, Politics and Policies*, Hampshire, New York: Palgrave Macmillan.

Tillmann, Klaus-Jürgen, Kathrin Dedering, Daniel Kneuper, Christian Kuhlmann and Isa Nessel (2008): *PISA als bildungspolitisches Ereignis*, Wiesbaden: VS Verlag für Sozialwissenschaften

Van Dooren, Wouter (2008), 'Performance indicators: a wolf in sheep’s clothing?', Paper presented at the Wissenschaftszentrum Berlin at the conference on “Changing Educational Accountability in Europe”, June 2008.

BIOGRAPHICAL NOTE

Kerstin Martens is project director at the Collaborative Research Center “Transformations of the State”, University of Bremen.

Telephone: +49 421 218 8733
E-Mail: kerstin.martens@sfb597.uni-bremen.de

Dennis Niemann is research fellow at the Collaborative Research Center “Transformations of the State”, University of Bremen.

Telephone: +49 421 218 8707
E-Mail: dennis.niemann@sfb597.uni-bremen.de

Address: University of Bremen, Collaborative Research Center „Transformations of the State“, Linzer Strasse 9a, D 28359 Bremen