Looking Back at Five Rounds of PISA: Impacts on Teaching and Learning in Germany

Christine Sälzer and Manfred Prenzel

The 3rd December 2013 was quite a happy day in Germany. When the results of PISA 2012 were released to the public, the picture looked gratifying and the conclusion to be drawn was that Germany did well in educating their youth and preparing young people for their lives after compulsory schooling. The fact that Germany performed well above the OECD average in all three PISA domains was considered good news, although there was still a gap between Germany and the internationally top-performing countries. If PISA 2012 had been the first round, the results probably would have been received less positively. In contrast, with reference to the results of the first round, known as the German "PISA-shock", it becomes quite clear what a distance lies between the student performance in PISA 2000 and in PISA 2012. The above-average results in PISA 2012 can therefore be seen as a common achievement between several players in national education policy, research and practice.

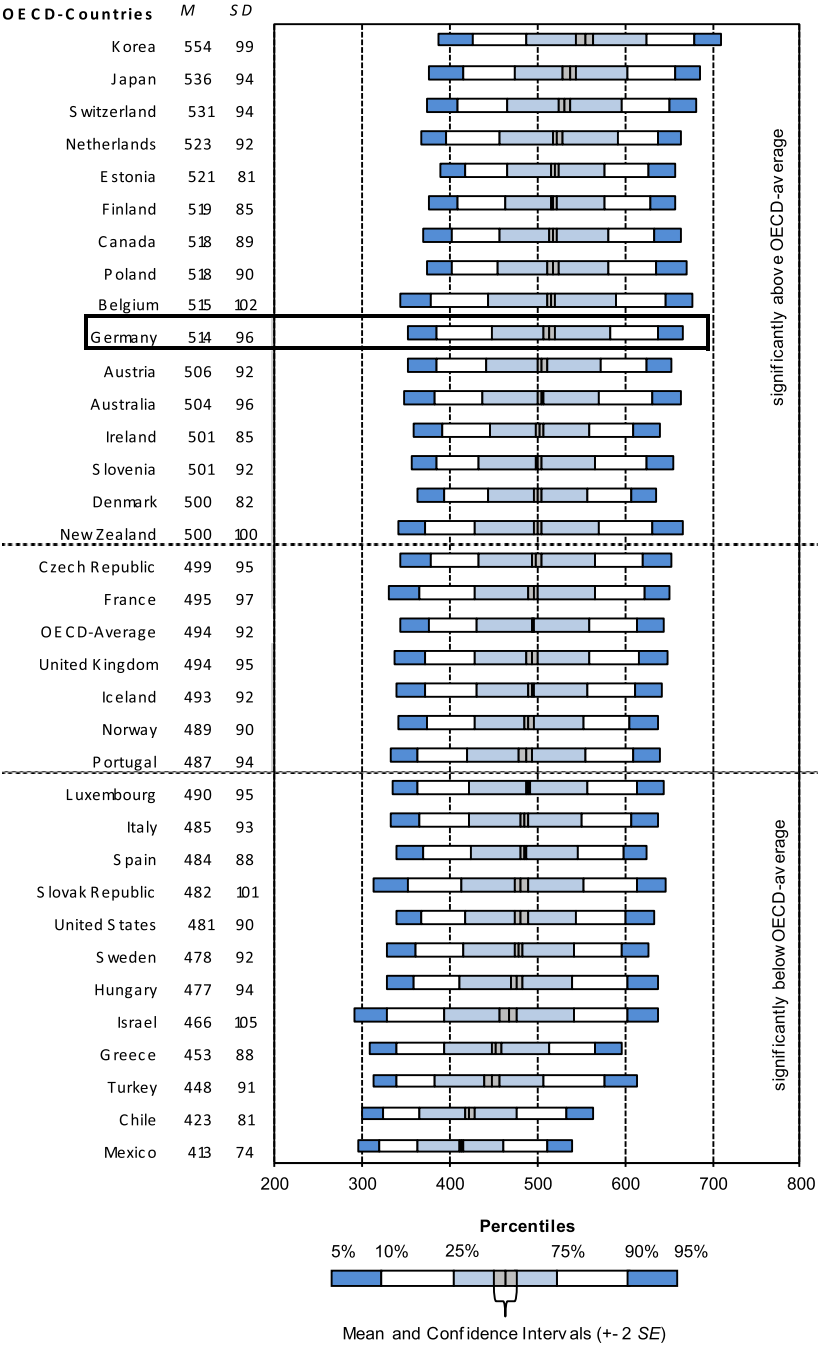
This article discusses the development of educational outcomes in PISA between 2000 and 2012. The question guiding the paper is: Which factors have contributed to a uniquely positive development and overall improvement of student performance? First, a few central results of PISA 2000 and PISA 2012 will be presented. Afterwards, three aspects of the deliberate change in Germany will be highlighted: a thorough diagnosis of the situation, an intense discourse between policy makers, the public and educational researchers as well as some specific examples of measures that have been taken. The article will be closed by an integrating discussion and implementations for the future.

PISA: Different Pictures of Educational Outcomes in Germany in PISA 2000 and PISA 2012

The picture revealed by PISA in Germany in 2001 yielded an overall achievement significantly below the OECD average in all three test domains. While the OECD average scores in all three domains were 500 in PISA 2000, Germany scored 484 in Reading, 490 in Mathematics and 487 in Science. Huge disparities in student performance were found according to social background, migration status and gender. Germany is by constitutional law divided into 16 federal states (so-called Länder), each holding its own responsibility for education. Surprisingly, an oversample of the PISA 2000 cohort showed enormous gaps between the highest and the lowest performing federal states (Baumert et al., 2002) within the country. This gap totalled 64 points on the PISA Mathematics scale, which corresponds to approximately two years of schooling (Ehmke et al., 2006). In other words, inside Germany, differences were found that covered nearly the range of OECD countries. Furthermore, the grading standards varied considerably between states, indicating that the feedback students received for their performance differed both according to the state they lived in and the school they attended within their state. For Germany, PISA 2000 provided data that had been unavailable before and hence allowed international benchmarking of characteristics of the educational system that shed light on the dimensions of statistical parameters that stood alone before PISA 2000. For example, the percentage of students who had repeated a school year in Germany was as high in no other country (Baumert et al., 2001; Krohne, Meier and Tillmann, 2004). Twenty four percent of all 15-year-old students had been retained and another 12 percent had started school one year later than usual, hence a total of 36 percent of students in Germany had a lagged school biography. PISA 2000 drew a picture of educational quality in Germany that was surprising and hence necessitated the need to get more data and reliable information about the situation.

Twelve years later, however, this picture looked considerably more positive (Prenzel, Sälzer, Klieme and Köller, 2013). Over the years, students in Germany have attained a level of performance that continuously exceeded the OECD average. Taking into account that the educational reforms and changes at the system level cannot be measured after just one or two PISA cycles, the development during more than a decade serves as a plausible indicator for monitoring progress. The positive trend in Germany is obvious, and at the same time the OECD average has not improved over the years. In part, this may be due to a changed combination of the

Figure 1: Percentile plots of mathematical competence in OECD countries



OECD countries; this group has been growing since PISA 2000, when it comprised 28 countries and PISA 2012, when it consisted of 34 countries. The constantly positive development of the average performance in Germany over five PISA rounds is quite unique among OECD countries, as the starting point (e.g., 490 points in Mathematics) was below the OECD average and still German students performed better and better in each of the following PISA rounds. The same applies to the other two domains, Reading (PISA 2000: M = 484; PISA 2012: M = 508) and Science (PISA 2000: M = 487; PISA2012: M = 524). Figure 1 shows the distribution of mean performance in Mathematics over all OECD countries in PISA 2012.

Besides an above average performance, the results of PISA 2012 showed that disparities due to social background and migration status have considerably decreased in Germany. The gain in student achievement goes back particularly to a large improvement in the achievement of low performing, low socioeconomic status and immigrant students. The number of students who do not surpass a level of proficiency that is worryingly low, has also decreased in size. Grade retention is still a common practice in Germany, but compared to 36 percent of students in PISA 2000 who had repeated a class at least once or started attending school later than usual, this percentage is now at 20 percent (Sälzer, Prenzel and Klieme, 2013; OECD average: 12.4 percent).

At this point, one can ask whether the improvement was obtained by merely focusing on student assessment and narrowing down the curriculum (Berliner, 2011), but also through enforced training or even drilling students (Ho, 2006). After all, PISA and other large-scale student assessments stand for a quite modern concept of educational 'efficiency' that can be measured through data collection (Carnoy, 2014). Taking into account non-cognitive outcomes also, PISA allowed analyses to control such assumptions. With regard to student characteristics and attitudes, we see that students in Germany have a high self-concept with regard to Mathematics (Schiepe-Tiska & Schmidtner, 2013) and they feel a strong sense of belonging towards their school (Sälzer et al., 2013). This had already been the case in PISA 2003, when Mathematics was the major domain of assessment for the first time. Problematic aspects such as late arrival or student absenteeism from school are, compared to the OECD average, marginal in German schools. As in the vast majority of participating countries, girls achieve significantly higher in reading competence than boys, while boys outperform girls in Mathematics (Gebhardt et al., 2013). While there are significant disparities between students with an immigrant background and those without one (Gebhardt et al., 2013), the relationship between

socio-economic background and student competence in Germany lies within the OECD average range (Müller and Ehmke, 2013). All in all, the results of PISA 2012 were positive news that revealed little cause for serious concern regarding a high price that is paid for improved performance, such as anxiety or lack of identification with school.

Looking back at the development between PISA 2000 and PISA 2012, the results of the latest PISA study seem somewhat impressive. Given that Germany had a difficult start with PISA, resulting in the well-known "PISA-shock", it is quite clear that good student performance is an achievement of a functioning educational system rather than a gift, a matter of the structure or curriculum. Numerous efforts have been taken, targeting the objective of improving student performance, reducing disparities and establishing a national concept of education rather than 16 Länder-specific educational systems. The following paragraphs highlight three central measures that we assume to have supported the positive development of Germany's performance in PISA.

A Through Diagnosis of the State of the Educational System

Along with the Third International Mathematics and Science Study (TIMSS), PISA 2000 was one of the first international large-scale student assessments that Germany participated in. For a long period of time, there was no doubt in Germany that the school system and its outcomes, especially in the domains of Mathematics and Science, were highly commendable. From the beginning, large scale assessment studies such as PISA have been conducted at research institutions in order to ensure high quality data analyses that go beyond a descriptive survey and provide both policy makers and the public with in-depth findings. These findings were meant to be used as a foundation for informed decisions or actions to be taken in order to improve educational quality in Germany.

The international comparisons as delivered by PISA provide benchmarks that help countries align themselves within a scale that indicates a country's position in an international context. How is a country's performance compared to other countries which have similar educational systems? How is it compared to countries with a fundamentally different structure? Without international benchmarks, many findings of studies within countries would have a less distinct meaning, for example 'high' satisfaction of teachers with their current job would stand-alone and one would be unable to tell whether this is good or bad news. In this regard, PISA provided large contributions to analysing the state of the educational system in Germany and to identifying its strengths and weaknesses.

From the beginning, numerous studies were attached to PISA in order to make sure the results of PISA would be useful for the German situation of a federal state that has 16 educational systems. National test instruments were administered as a supplement to the PISA tests in order to ensure an adequate measurement of nationally relevant curricular competences on the one hand and a valid interpretation of the international results of PISA on the other hand. Additional tests in Mathematics and Science were used to enhance the international design of PISA 2000, where both Mathematics and Science appeared as minor test domains for the first time. Reading was the major domain and was broadly investigated at the international level.

The German supplement was designed and administered so that already in the first cycle of PISA, a deep and differentiated analysis was possible as well as studying the association of international PISA items with items that were based on German curricula. These extra tests took place in a second testing day, along with additional tests capturing cross-curricular competences. Such overarching competences like communication, cooperation or school-related problem-solving were considered to be relevant indicators of students' readiness for life as a citizen and an autonomous part of society (Baumert et al., 2001).

Besides the cognitive part of PISA, researchers in Germany put an emphasis on student attitudes, experiences and beliefs as well. These constructs were measured using a student questionnaire, which in Germany was expanded by a number of additional scales and items. Among others, peer relationships were measured, and a nationally enhanced school questionnaire comprised of a number of questions that captured, for example, quality assurance and cooperation of schools with other institutions was used. A parent questionnaire had not been part of the international survey design of PISA 2000, but German researchers had developed one in order verify student responses with regard to their family background and in order to collect data on individual educational biographies.

In addition to the mentioned supplementary tests and questionnaires, the sample of PISA participants was extended as well. Taking into account the federal structure of Germany, politicians and researchers were interested in having a sample representative of the Länder. Such a sample would enable a thorough description of the situation within the Länder and comparing the results between the Länder. Germany actually conducted two PISA-studies: PISA-I (PISA International) and PISA-E (PISA Extended; Baumert et al., 2002; Prenzel, Baumert et al., 2005). While the international sample comprised between 200 and 250 schools, the sample representative at the Länder level consisted of around 1500 schools. Both

samples were overlapping, i.e. the PISA-I-schools were a subset of the PISA-E-schools. Besides this oversample at the school level, Germany enhanced the student sample within schools. Although the age of school enrolment in Germany corresponds to the OECD average (6 years; Sälzer et al., 2013), there are several measures commonly used in Germany which lead to quite a wide range of grade levels attended by 15-year-old students. Usually, the German PISA sample is drawn from students attending grade 7, 8, 9, 10 or 11. In order to get a more robust idea of the competences and characteristics of students towards the lower secondary level, an oversample of about ten students per 9th grade in each of the sampled schools was drawn. From PISA 2003 on, complete ninth grades (two per school) were drawn in addition to the age-based student sample of 15-year-olds (Prenzel et al., 2004). This additional sample was widely used to explore processes during lessons in order to analyse teaching and various interactions in classrooms (e. g., Seidel, 2002). One example for a comprehensive research and development undertaking is the so-called SINUS project, which will be described below.

At the same time, numerous research initiatives in the field of education have been activated. The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) launched a Priority Programme focusing on educational quality in schools. The main focus of this programme was to explore school-related and outside school conditions for developing competences in Mathematics, Science and cross-curricular fields. One crucial aspect of the Priority Programmes is a nationwide cooperation between the participating researchers, particularly with regard to the federal structure of the country. Results from the Programme revealed possible reasons for the weaknesses of the German educational system that had recently been discovered by PISA and TIMSS and they investigated what could be done in order to improve educational outcomes in the future. More than 30 studies within the Programme took different perspectives on the educational quality in schools and focused on topics such as the effects of single-sex schools, evaluating training on self-regulated learning and problem-solving or instructional quality. All these topics were immediately related to aspects and issues that were identified to be either a weakness of the educational system in Germany or that were lacking reliable data and research findings.

To summarise, a lot of effort has been taken in Germany to improve insights into educational structures, processes and outcomes. The need for sound empirical research was clearer than ever after both TIMSS and PISA had discovered the devastating state of the educational system. Not only the participation in international comparisons, but also an extensive

national enhancement of these studies marked the beginning of a new era that did not settle for a common sense perception of educational effectiveness, but required reliable and solid evidence. Such evidence was the foundation for an informed discourse among policy makers, researchers and the public, as described in the following section.

Discourse Between Policy Makers, Researchers and the Public[1]

Collecting and analysing empirical data on educational institutions, processes and outcomes provided many institutional and political players with profound evidence that can help in different ways when decisions have to be made regarding the educational system. Doing this at an international level was quite new in Germany after PISA 2000. Policy makers, researchers and the public now had a common ground to start from, which could be used to prioritize the choice of options that were identified with regard to strengths and weaknesses of the German educational system at the secondary level. In this sense, PISA is said to have contributed to a rediscovery of internationally and nationally comparative education in the German discourse (Ertl, 2006). The public reaction to PISA was overwhelming in Germany. Hitting the headlines for weeks after the release of the results in December 2001, PISA entered everyday discussions and enduring debates on the quality of schools, teaching and teachers in Germany. Since then, education has made it into the focus of public attention where it used to be something that was taken for granted. Questions such as whether the school a child is attending matters for their achievement, development or well-being arose had rarely been asked before (Baumert, Trautwein and Artelt, 2003).

PISA affected Germany as a whole country. The problem revealed by PISA had to be addressed from a nationwide perspective. Germany had to face this problem in order to improve the quality of schools and educational outcomes. As education was, and still is, the responsibility of the Länder, this situation was new. And yet, Germany had to face this problem in order to avoid a repetition of the disastrous results in international LSA studies. The "PISA-shock" ignited a broad reform agenda in Germany, which is best seen in a document called *PISA 2000 – Central Fields of Action* by The Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germa-

1 Referring to the 'public' in this article means the sphere outside formal politics, policy discourse or academia. In that regard, the 'public' comprises of citizens in a country, here: Germany, who discuss about educational issues from diverse perspectives, but not necessarily with a professional background or knowledge.

ny (KMK). This document is part of a long-term strategy (KMK, 2006) and states that Germany will participate in international LSA studies in the long run. Seven fields of action were identified that are to be transferred into specific measures to improve the educational situation in the whole country. Priority was given to (1) early support for improving language skills, (2) better links between pre-school and elementary school, (3) improvement of elementary schooling by a steady advancement of competences in reading, Mathematics and Science, (4) more support for children at risk, especially for those with an immigrant background, (5) measures of consequent advancement and quality assurance in teaching and school using nationwide standards and evaluation, (6) professional teacher education as well as (7) all-day schooling. One example of how these fields of action were addressed is the implementation of national educational standards. A group of researchers and experts in the field of education developed a framework for the development and implementation of standards (Klieme et al., 2003). This framework served as a kind of roadmap for combining societal objectives, scientific findings about competence development as well as concepts and procedures of test development.

Implementation of Overarching Measures

Besides a thorough diagnosis of problems within the educational system and a widespread discourse on education, the third consequence to be taken in Germany was to implement nationwide, overarching measures inside and across schools. One of these overarching and nationwide programmes, implemented after the TIMSS study in Germany, was the so-called SINUS-programme (*Enhancing the Efficiency of Teaching in Science and Mathematics*). SINUS was a model programme pursuing a sustainable improvement of the professional development of Mathematics and Science teachers. The programme suggested 11 modules as the core of SINUS which aimed to improve teaching and learning by advancing the development of a thorough change in Mathematics tasks (Prenzel, Friedrich & Stadler, 2009). The structure of modules enabled schools to select their own set of modules which could be combined and, later on, enhanced by new modules. This "new culture" of Mathematics tasks comprised of a much broader range of mathematical competencies (Niss, 2003) and focused on securing a basic understanding and fostering of cumulative learning in Mathematics. Teachers could find highly elaborated recommendations that helped them identify strengths and weaknesses in their own teaching as well as ideas and impulses for developing further approaches. SINUS intended to engage a large number of teachers who would commit to working in teams and use the modules to improve

tasks, materials and teaching approaches. It can therefore be considered as an example of teacher professional development (Oster Meier, Prenzel and Duit, 2010). Dissemination of approved successful modules took place first within the school, and then spread to other schools in the regional and later national school networks. The principle of SINUS was to encourage teachers teaching the same subject to cooperate within their school and, in the long run, between schools as a school network. Such cooperation would be coordinated at a local, a regional or a cross-regional level. In doing so, SINUS generated an enormous pool of materials available online to all interested teachers as well as in the form of manuals, books or teachers' magazines. Over the years, SINUS had developed from a starting size of 180 secondary schools in 1998 to 1750 schools in 2003, followed by a modified programme for primary schools.

SINUS was guided by five principles which best describe its approach. First, problematic areas should be handled by working on 11 so-called modules. These modules referred to an expertise which had identified certain problematic issues found in Mathematics and Science teaching in Germany (BLK Projektgruppe "Innovationen im Bildungswesen", 1997) and contained suggestions for developing one's teaching practices. Second, cooperation among teachers within the same school as well as within school networks (so-called school sets) was encouraged and fostered. Quality development and quality assurance was the third principle, meaning that schools were to establish routines in developing and maintaining quality-related factors. Fourth, the work of participating teachers was scientifically inspired and supported. Materials, professional development courses and counselling were meant to accompany the 11 modules mentioned. Finally, SINUS was enhanced by a number of research studies to delve into conceptual questions or questions of implementation, but also to evaluate the programme. Along with these five principles, SINUS pursued objectives at three different levels. Professional development of teachers was the first objective, meaning among others that teachers should learn to cooperate by sharing materials and test items, visiting each other during lessons or combining forces for preparing lessons. Cooperation has been found to be a crucial feature of effective schools (Sammons, 1999), but was not very common in German schools (Terhart, 2001) and should therefore be promoted. The second objective envisaged the level of teaching in lessons, which implied the need for improvement. Modules focusing on this second objective comprised of elements such as developing a new culture of tasks during lessons, scientific working or learning from mistakes (Prenzel, Carstensen, Senkbeil, Ostermeier and Seidel, 2005). The third objective is aimed at student learning

and achievement, seen as a learning outcome. Modules referring to this objective serve as a content-oriented framework for scaffolding in-depth learning, positive attitudes and high interest in Mathematics and Science.

SINUS schools were tested in 2003, using some of the PISA test items in order to compare student performance between SINUS schools and, as a control group, PISA schools (Prenzel, Carstensen et al., 2005). Results showed that students attending SINUS schools, at least in some school types, performed significantly better than students at non-SINUS (i.e. PISA) schools and also showed higher interest and more positive attitudes towards Mathematics and Science. To summarise, comprehensive schools and lower-secondary level schools (Hauptschule, offering a secondary-level I certificate after 9 years, HSA) profited the most from SINUS. This is in line with the findings of PISA 2012, where it turned out that the competence improvement over the past twelve years went back mostly to school types besides the Gymnasium, which are set up less academically than a Gymnasium. Therefore it seems as if especially those school-types with the most significant need for development have improved by participating in SINUS (Prenzel et al., 2004; Prenzel et al., 2006).

Another relevant overarching measure in Germany was national educational standards. Introducing national educational standards in Germany was one of the most fundamental measures to be taken as a consequence of the "PISA shock". In 2003 and 2004, the KMK published educational standards referring to primary level and secondary level I (e.g. KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik, 2003; KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik, 2004). At the primary level, the core of these standards referred to the school subjects of German language and Mathematics. At secondary level I, focus was set on German language, Mathematics, and the first foreign languages learned at school (English or, in some Länder, French). Standards were differentiated between a lower-secondary school-leaving certificate (HSA) after nine school years and an intermediate secondary school-leaving certificate (MSA), which is usually attained after ten school years (Pant et al., 2014).

For the first time, education was considered from a perspective of standards, competences and outcomes rather than from a curricular point of view and an input-oriented perspective. These standards being educational *objectives* means that the underlying concept of a young learner at a certain age is one that defines a competence that is expected from each individual who completes a certain level of schooling (Klieme et al., 2003). Such competences are described from a mastery point of view, i.e. they

indicate requirements towards teaching and learning at school and they name desirable learning outcomes for students. In that regard, Germany moved from an input-oriented perspective to an output-oriented point of view towards the educational system. Another important aspect that comes with this concept of educational standards is that they are verbalized and illustrated so teachers get a clear idea of what the curricular description of a standard or competence means. In addition, the description of standards can be understood as a framework for the development of tasks and tests (Klieme et al., 2003). Schools are expected to ensure that the competences defined in the national educational standards can be achieved by every student, regardless of social background. This idea shifts the role of schools from one that is responsible for successfully completing Länder-specific curricula to one that is responsible for ensuring the development of defined levels of competence that are valid across the Länder and across school-types. In order to sustainably develop and test the new educational standards in Germany, a specific research institution was founded in Berlin: the Institute for Educational Quality Improvement (IQB) at the Humboldt-Universität zu Berlin. One of the main responsibilities of the IQB is to administer sample-based comparative assessments of state-level (Länder) educational performance in order to assess the extent to which educational standards are being met across the 16 federal states of Germany. Along with several international Large-Scale Assessments such as PISA, TIMSS or PIRLS (Progress in International Reading Literacy Study), the so-called IQB National Assessment Studies examine student competences at the end of secondary level I within Germany. The first IQB National Assessment Study was conducted in 2009, in parallel with PISA 2009, and tested student competences in the subjects of German and the first foreign language at secondary level I (Köller, Knigge and Tesch, 2010). The next IQB National Assessment Study followed, along with PISA 2012, and examined student competences in Mathematics, Biology, Chemistry, and Physics (Pant et al., 2013).

One more aspect adds to the quite special situation in Germany's educational system. In Germany, there is only one secondary school-type that is prevalent throughout the country. All 16 Länder have the Gymnasium system, which qualifies students to attend tertiary education after completing eight or nine school years, respectively. So far, the PISA performance of Gymnasium students has been consistently high (Baumert et al., 2001; Klieme et al., 2010; Prenzel et al., 2004; Prenzel et al., 2007; Prenzel et al., 2013). As Gymnasium schools are the only school type that has not changed due to reforms over the PISA cycles and is prevalent in all federal states in Germany, only Gymnasium schools can be analysed from

a trend perspective. Therefore, it is not possible to describe the proficiency development of other school types over time. It can, however, be assumed that if the student performance at Gymnasium schools has not changed and the average performance of all students in Germany has significantly improved since PISA 2000, it must be due to an increase in student performance at other school types besides the Gymnasium (Prenzel et al., 2013). On one hand, this is good news, as the national educational standards published in 2003 and 2004 explicitly refer to the lower secondary and the intermediate secondary school-leaving certificate that can be obtained in school types other than Gymnasiums. So it seems that these school types have successfully implemented the objective lying behind the educational standards, namely to prepare students to transfer knowledge and processes to other, partly unknown situations and contexts and master a defined minimum of requirements that should be met by a typical young person leaving compulsory schooling. On the other hand, the fact that the highest achieving students who are grouped in Gymnasium schools have not improved their average performance from cohort to cohort as have students in other school types is not such good news. National educational standards with a focus on Gymnasium graduates (*Allgemeine Hochschulreife*, i.e. *Abitur* or high-school diploma) have recently been published and their effects will be visible in several years' time (e.g., KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik, 2012; KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik, 2013). Taking into account that the percentage of students per age cohort attending a Gymnasium has increased since PISA 2000 (from less than 30 percent to 36 percent) and still the average performance of this school type is consistently high, this expansion of the Gymnasium has been successful. However, the results of PISA 2012 also showed that high achieving students with good potential can and should be supported more in German schools, especially at the Gymnasium.

Both the SINUS programme and national educational standards tried to overcome boundaries between federal states in Germany and develop a spirit of cooperation and a common understanding of what education should pursue, how it could be done effectively and what good teaching actually means. Thinking of the teaching profession as one of the most responsible positions in society and with a lot of potential for professional development, SINUS chose a modular approach and encouraged teachers to overcome their being used to individualism and cooperate with their colleagues. At the same time, setting national educational standards reformed the understanding of proficiency, away from an

achievement-based orientation towards a mastery orientation. Mastery of educational standards, as understood in the German case, involves a regular assessment of student competences. These competences are defined in the educational standards rather than in the curriculum.

Summary and Discussion

This paper dealt with the recently published results of the fifth round of PISA in Germany and their meaning with regard to the development of the educational system since the "PISA-shock" that followed the first round of PISA in 2001. Comparing the pictures in 2001 and 2013, revealed that students in Germany have notably improved their average performance. It is at least in part for this reason that the positive results of PISA 2012 have been received quite happily in Germany. If it were not for the "PISA shock", results like those in PISA 2012, where students in Germany performed significantly above the OECD average in all domains, would not be very spectacular, but rather conforming to everybody's expectations. The paper then discussed how this quite uniquely positive development of student performance over five rounds of PISA could be explained. Three main aspects of the deliberate change in Germany have been elaborated: a thorough diagnosis of the state of the educational system, an intense discourse between policy makers, researchers and the public, as well as the implementation of overarching measures across the country.

With regard to the diagnosis PISA provided and the discourse it stimulated, Germany's educational system has benefited a lot from this study. Intense and, at times, heated debates have been important motivators the German development since PISA 2000. Knowing about problematic aspects within the system and struggling for better educational outcomes that are not achieved by a successful teaching of the test, but by merely by activating sustainable teaching, has proved to be healthy and fertile. However, PISA is limited with regard to what types of data and information it can deliver. PISA provides high-quality indicators of certain aspects of educational systems worldwide, such as specific strengths and weaknesses. It is thus a solid foundation for educational benchmarking at an international level. PISA does not however go beyond that. Solutions and strategies to implement changes have to be identified elsewhere. In PISA, countries can learn from one another if they wish, although some suggestions of publications in the PISA context miss the point by focusing too strictly on structural indicators of educational systems that may have some effects on student development and the explanation of variance, but do not necessarily take into account the underlying processes or country-specific coherences that go beyond explaining variance in student

performance (OECD, 2010). In Germany, some structural modifications have been undertaken, such as a reconsideration of the multi-tier secondary schooling system that is and has been prevalent in all federal states. However, these modifications have not touched the Gymnasium and the general structure of several secondary schooling types which children at age 10 are assigned to. For such reasons, flanking educational research is needed to enhance the potential that lies within PISA data. In Germany, this has been undertaken at the political level by the seven fields of action that have been stated by the Standing Conference of the Ministers of Education and at the academic level by accompanying research attached to PISA and by allocating the national project management to research institutions. If project management does not stop at the point where the collected data are submitted to another institution, but continues with country-specific analyses of PISA data, the value added to the participation in international large-scale assessments widely increases. Concerning the third pillar of deliberate change after PISA 2000, implementing overarching measures was the consequence that touched schools and teachers most directly. Introducing national educational standards initiated a shift from an input-oriented to an output- or outcome-oriented perspective in education. Measures to improve school quality, such as internal and external evaluation, standardised policies for Mathematics or teacher mentoring (OECD, 2013) are less prevalent in Germany than in most other OECD countries, but many schools already use these instruments and it can be assumed that more will follow when they hear about improvements and progress from their colleagues.

Can one conclude from this that Germany is done with PISA, now that student literacy is above the OECD average and many problematic aspects have been improved? Not quite yet. The uniquely positive development of Germany rests on several pillars: a nationwide strategy ('fields of action') bringing 16 educational systems together, national educational standards, overarching projects such as SINUS, but also a steady scientific evaluation and enhancement of large scale assessment studies and a continuous diagnosis of the state of the educational system. These aspects have become part of a regular public, scientific and political discourse. Such discourse was not prevalent before PISA in Germany and it now contributes to a common understanding of education as a goal to achieve and a project to work on. Education does not come by itself and it cannot be "provided" to students. Students find educational offers in schools and they have to make use of them in order to become educated young individuals, ready for a position in modern societies. Taking into account that the trend of PISA results (and, hence, of educational outcomes) in Germa-

ny is pointing upwards, there are also several challenges to be addressed. Although disparities have decreased, e.g. between immigrant and non-immigrant students or between students of high and low socio-economic background, these disparities still exist and need to be reduced further. Gender gaps, indicating that girls outperform boys in reading and boys outperform girls in Mathematics, have to be closed. This is especially relevant for Gymnasium schools, where many students achieve excellent results. It is within this group of top-achievers in Mathematics that girls are underrepresented . In general, strong and high achieving students require more support in Germany, e.g. through enrichment or specific school programmes offering a setting to develop excellent skills, but also high interest and motivation with regard to literacy skills as they are proposed and measured in PISA. With regard to supporting strong students as well as average or low achieving students, teaching practices and teaching quality need to be improved. Tasks that stimulate students to think, to be creative and interactive learners and to be able to transfer skills from the classroom to the real world are needed and these tasks have to be implemented in lessons. Given that other PISA participant countries have developed less positively than Germany, whether they have undergone structural reforms or not, this last aspect of teaching practices and teaching quality improvement seems to be more important by far than the question of comprehensive or tracked school systems providing a better learning environment for students. After all, it is not a matter of the system, but of the classroom processes when it comes to fostering student learning and preparing them for life as citizens in modern societies.

References

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., and Weiss, M. (eds.) (2002) *Pisa 2000 - die Länder der Bundesrepublik Deutschland im Vergleich: [PISA-E].* Opladen: Leske + Budrich.

Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., and Weiss, M. (eds.). (eds.) (2001) *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske + Budrich.

Baumert, J., Trautwein, U. & Artelt, C. (2003) Schulumwelten -- institutionelle Bedingungen des Lehrens und Lernens. In: Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J., and Weiss, M. (eds.). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich, pp .261-331.

Berliner, D. (2011) Rational responses to high stakes testing: the case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education*. 41(3), pp. 287-302.

BLK PROJEKTGRUPPE „INNOVATIONEN IM BILDUNGSWESEN" (1997) Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts". Materialien zur Bildungsplanung und zur Forschungsförderung: Vol. 60. Bonn: BLK, Geschäftsstelle.

Carnoy, M. (2014) Globalization, educational change, and the national state. In: Stromquist, N. P., and Monkman, K. (eds.) *Globalization and education. Integration and contestation across cultures*. Lanham, MD: Rowman & Littlefield., pp. 21-38.

Ehmke, T., Blum, W., Neubrand, M., Jordan, A. & Ulfig, F. (2006) Wie verändert sich die mathematische Kompetenz von der neunten zur zehnten Klassenstufe? In: Prenzel, M., Baumert, J., Bum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. and Schiefele, U. (eds.) *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres.* Müster: Waxmann, pp. 63-86.

Erth, H. (2006) Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education*. 32 (5), pp. 619-634.

Gebhard, M., Rauch, D., Mang, J., Sälzer, C. & Stanat, P. (2013) Mathematische Kompetenz von Schülerinnen und Schülern mit Zuwanderungshintergrund. In: Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (eds.) *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann, pp. 275-308.

Ho, E. S.-C. (2006) High stakes testing and its impact on students and schools in Hong Kong: What we have learned from the PISA studies. *Korean Journal of Educational Policy*. 3 (1), pp. 69-87.

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. Bonn: BMBF.

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik (2003) Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss (10. Jahrgangsstufe). Beschlüsse der Kultusministerkonferenz.

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik (2004) Bildungsstandards im Fach Mathematik für den Hauptschulabschluss (9. Jahrgangsstufe). Beschluss vom 15.10.2004. Beschlüsse der Kultusministerkonferenz.

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik (2006) *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. München: Wolters Kluwer.

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik (2012) Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife: Beschluss der Kultusministerkonferenz vom 18.10.2012. Available from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf (Accessed: 2nd May 2014).

KMK = Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik. (2013) Konzeption zur Implementation der Bildungsstandards für die Allgemeine Hochschulreife: Beschluss der Kultusministerkonferenz vom 10.10.2013. Available from: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2013/2013-10-10_Konzeption_Implementation_Bildungsstandards-AHR.pdf (Accessed: 2nd May 2014).

Köller, O., Knigge, M., and Tesch, B. (2010) *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann. Available from: http://www.e-cademic.de/data/ebooks/extracts/9783830923398.pdf (Accessed: 2nd May 2014).

Krohne, J. A., Meier, U., and Tillmann, K.-J. (2011) Sitzenbleiben, Geschlecht und Migration - Klassenwiederholungen im Spiegel der PISA-Daten. *Zeitschrift für Pädagogik*. 50 (3), pp. 373-391. Available from: http://www.pedocs.de/volltexte/2011/4816 (Accessed: 2nd May 2014).

Müller, K., and Ehmke, T. (2013) Soziale Herkunft als Bedingung der Kompetenzentwicklung. In: Prenzel, M., Sälzer, C., Klieme, E. & Köller, O. (eds.) *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann, pp. 245-274.

Niss, M. (2003) Mathematical competencies and the learning of Mathematics: The danish KOM project. In: Gagatsis, A., and Papastavridis, S. (eds.) *3rd Mediterranean conference on mathematical education*. Athens, Greece: Hellenic Mathematical Society and Cyprus Mathematical Society, pp.115-124.

OECD (2010) *What Makes a School Successful? PISA 2009 Results.* Paris: OECD Publishing.

OECD (2013) *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices. Volume IV.* Paris: OECD Publishing.

Ostermeier, C., Prenzel, M., and Duit, R. (2010) Improving Science and Mathematics instruction - The SINUS-Project as an example for re-

form as teacher professional development. *International Journal of Science Education*. 32 (3), pp. 303-327.

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., and Pöhlmann, C. (eds.) (2013) *IQB-Ländervergleich 2012: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I*. Münster: Waxmann.

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., and Pöhlmann, C. (2014) *The IQB National Assessment Study 2012: Summary*. Münster: Waxmann.

Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., and Pekrun, R. (eds.) (2007) *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. and Schiefele, U. (eds.) (2004) *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland: Ergebnisse des zweiten internationalen Vergleichs*. Münster, New York: Waxmann.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. and Schiefele, U. (eds.) (2005) *PISA 2003. Der zweite Vergleich der Länder in Deutschland -- Was wissen und können Jugendliche?* Münster: Waxmann.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. and Schiefele, U. (eds.) (2006) *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Müster: Waxmann.

Prenzel, M., Carstensen, C. H., Senkbeil, M., Ostermeier, C., and Seidel, T. (2005) Wie schneiden SINUS-Schulen bei PISA ab? *Zeitschrift für Erziehungswissenschaft*. 8 (4), pp. 540-561.

Prenzel, M., Friedrich, A., and Stadler, M. (eds.) (2009) *Sinus-Transfer. Von SINUS lernen: Wie Unterrichtsentwicklung gelingt (1st ed.)*. Seelze: Kallmeyer.

Prenzel, M., Sälzer, C., Klieme, E., and Köller, O. (eds.) (2013) *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann.

Sälzer, C., Prenzel, M., and Klieme, E. (2013) Schulische Rahmenbedingungen der Kompetenzentwicklung. In: Prenzel, M., Sälzer, C., Klieme, E., and Köller, O. (eds.) *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann, pp.155-187.

Sammons, P. (1999) *School Effectiveness: Coming of age in the twenty-first century*. Lisse: Swets & Zeitlinger.

Schiepe-Tiska, A., and Schmidtner, S. (2013) Mathematikbezogene emotionale und motivationale Orientierungen, Einstellungen und Verhaltensweisen von Jugendlichen in PISA 2012. In: Prenzel, M., Sälzer, C., Klieme, E., and Köller, O. (eds.) *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster: Waxmann, pp.100-121.

Seidel, T. (2002) *Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Prozesse beim Lernen -- Eine Videostudie im Physikunterricht.* Münster: Waxmann.

Terhart, E. (2001) *Lehrerberuf und Lehrerbildung: Forschungsbefunde, Problemanalysen, Reformkonzepte.* Weinheim: Beltz.