

# Chapter 12

## Educational Assessment in Germany



Nele McElvany and Justine Stang 

Germany participates in several national (e.g. NEPS) and international (e.g. PISA, TIMSS) student assessments. This chapter thus presents an overview of Germany's participation in several national and international assessments. It also explains Germany's current educational monitoring system and how it has been affected by results of international student assessments. Results from the Progress of International Reading Literacy Study (PIRLS) are described in detail as an example of international student assessments in Germany. The study focuses on students' reading literacy, students' motivation, instructional quality, differences between boys and girls, and differences between students with and without a migration background. The chapter closes by discussing Germany's assessment policies, practices and outcomes.

### Introduction

#### *The Education System in Germany*

In Germany, the responsibility and cultural sovereignty for the education system lies primarily with the 16 federal states. Children, mainly aged three to six, may attend kindergarten. Subsequently, children are enrolled in school. In most states, the primary school lasts for four years. Therefore, the educational system is selective at a very early age: By the end of primary school, the decision, mainly based on students' achievement, is made to which type of secondary school children may go. The secondary schools are separated in lower- and upper-secondary schools (e.g.

---

N. McElvany (✉) · J. Stang  
Center for Research on Education and School Development (IFS), TU Dortmund University,  
Dortmund, Germany  
e-mail: [nele.mcelvany@tu-dortmund.de](mailto:nele.mcelvany@tu-dortmund.de)

grammar school). The first phase of secondary education typically lasts for 6 school years (grade 5–10), leading to several options of school leaving certificates. However, the length of compulsory education differs between the federal states. After compulsory schooling, another decision on the educational path has to be made. At this stage, the German education system is characterized by a wide range of education and training tracks, including, for example, upper-secondary school, which leads on to A levels or vocational schools. While 41.2% of all students graduated school with a school leaving certificate allowing them to study at a German university in 2016, about 6% left school without any formal graduation certificate (Autorengruppe Bildungsberichterstattung 2018).

### ***History of Participation in International Assessments in Germany***

Although (West) Germany started to participate in international student assessments early on, their scope was limited. In 1964, two German federal states took part in the First International Mathematics Study (FIMS) that assessed mathematics performance in secondary school students, mathematics teaching and the influence of social, curricular, and technological developments across 12 countries (Husén 1967; Schultze and Riemenschneider 1967). Shortly after FIMS, Germany joined parts of the Six Subject Study (English; political education); and in 1971, a representative sample of students from ten federal states (Schultze 1975) took part in the First International Science Study (FISS). These studies were conducted by the International Association for the Evaluation of Educational Achievement (IEA), founded in 1958 as an international cooperative of national research institutions, governmental research agencies, scholars and analysts. In 1971, the intergovernmental economic organization, the Organisation for Economic Co-operation and Development (OECD) also published an examination of educational systems in several countries including Germany.

After this early phase of Germany's (at least partial) interest in international educational comparisons, there was a long interval before Germany participated in an international large-scale study again with a representative student sample. In 1990–1991, 9- and 14-year-old Germans from both East and West German federal states were included in the International Study of Reading Literacy (IRLS or RL; Lehmann et al. 1995). Although the performance of German students was only just average, these findings attracted limited attention. In the same period, 10- and 13-year-olds were tested in 9 federal states in the 1989 and 1992 Computers in Education Study (ComPed; Lang and Schulz-Zander 1994; followed later by the Second Information Technology in Education Study). Participation in these large-scale assessments (LSA) marked the beginning of a new phase of educational monitoring that shifted from a nearly two-decade-long focus on issues such as individual school development or school tracking. The educational administrations' approach

on input orientation in these decades was met by an educational science often emphasizing on other approaches than empirical-quantitative evidence. This new phase continued with the German participation of representative samples of 13-year-olds and young adults in the Third International Mathematics and Science Study (TIMSS) from 1994 to 1996. The merely average performance outcomes raised awareness of the need for both empirical assessments of educational outcomes and improvements in teaching math and science in German schools. For education practice, one outcome was a large-scale model programme to ‘Increase of the efficiency of the math- and science instruction’ (SINUS) initially just for secondary schools but later also for primary schools ending in a transfer project period (Prenzel et al. 2009). For educational research, this led to the development of a strong area of research on math and science education that advanced instructional research compared to most other subjects by both quantity as well as empirical foundation. At the end of the millennium, Germany joined the Civic Education Study (CIVED) of representative samples of 14-year-olds in 28 countries (Händle et al. 1999). Again, although German students’ performance was only average, the results received little attention in Germany from either the general public or educational administrators. In 2009, Germany opted out of the study, and in 2016, only one federal German state joined the International Civic and Citizenship Education Study (ICCS).

The awareness of all stakeholders – educational policymakers, administrators, practitioners, universities educating future teachers, educational researchers, and last, but not least, the German public – was raised suddenly and definitively by the results of the Programme for International Student Assessment (PISA) study (Baumert et al. 2001; OECD 2001). The study revealed that 15-year-olds in Germany were performing far below expected levels in reading literacy (484 points and thus 16 points below the international OECD average), and that the distance to the group of high-performing countries was substantial. Additionally, the subgroup of very low performers was large (about 20%), as was the performance heterogeneity within Germany, and performance was particularly bad in the ‘reflecting and judging’ sub-scale. It also turned out that performance correlated more closely with socio-economic family background than in any other country under investigation. In the same vein, the average performance of students with a migrant background was much worse than that of native students. One year later, the results of the Progress in International Reading Literacy Study (PIRLS) were published and showed a more positive picture of German primary school students’ reading literacy than the PISA results for secondary school students (Bos et al. 2003; Martin et al. 2003): the average performance of German 4th-graders was the same as that in other participating European countries, the variation in performance was comparatively small (indicating homogeneity), and girls and boys did not differ as much in their reading skills as they did in many other countries.

Germany continued to participate in the regular waves of TIMSS (only 4th graders), PIRLS and PISA in the following years (see Table X.1). It also took part in the IEA’s International Computer and Information Literacy Study (ICILS) in 2013 and 2018. Furthermore, Germany also joined (1) the OECD LSA Programme for the

International Assessment of Adult Competencies (PIAAC) focusing on adult competencies in literacy, numeracy and problem solving within technology-rich environments (2011); (2) the Teacher Education and Development Study in Mathematics (TEDS-M; 2008; financed by a grant from the German Research Foundation to Humboldt Universität Berlin) focusing on future math teachers; and, currently, (3) the Teaching and Learning International Survey (TALIS-Video; 2018; financed by a grant from the Leibniz Society).

## Educational Monitoring and Policy Documents, Perspectives and Assessment Strategies in Germany

After a long phase of abstinence from international educational studies and disappointing results in newer LSA, the politics of educational monitoring in Germany were significantly changed towards a systematic overall approach that was supported by all 16 federal states in the last 20 years. This has also included a fundamental shift from a previously long-term emphasis on input to a new orientation towards output in the perceptions and measures of educational administrators and state governments. An integral part of this new approach has been the establishment in 2004 of an academic institute funded by all the German federal states: the Institute for Educational Quality Improvement [Institut zur Qualitätsentwicklung im Bildungswesen, IQB]. The institute's purpose is to ensure and improve the quality of education by operationalizing and evaluating educational standards and coordinating standard-based item development (see also Klieme et al. 2004). The change on the policymaking level has been accompanied by a substantial transformation of German educational science from the earlier domination of non-(quantitative) empirical approaches towards a strong empirical foundation in much current research. This development gained momentum through the establishment of a broad number of professorships dedicated to empirical educational research at most German universities and the subsequent formation of research groups in this field. Recent German assessment strategies based on the adjustment from an input to an output orientation are presented in three core documents agreed upon by the Standing Conference of the Ministers of Education and Cultural Affairs (KMK 1997, 2006, 2016), which includes ministries from all 16 federal states:

- 1997: Educational policymakers declared their aim to use empirical data from educational research to identify strengths and weaknesses in the educational system and to use appropriate measures; focus on secondary schools (Grades 9/10) and competencies in (German) language, mathematics, science, and foreign languages; and personal and social skills. Known as the '*Konstanzer Beschluss*', this marked the start of the shift towards an empirical approach. [Empirische Wende]
- 2006: Core studies and instruments were defined as a shared basis for an evidence-based educational policy oriented on the results (output) of educational processes (known as the '*Plöner Beschlüsse*').

2015: Update to the overall strategy, strengthening the need for explanatory next to descriptive knowledge, and identifying core areas of interest for further evidence to guide educational policy and practice.

All in all, four areas or tools have been identified and agreed upon as the current focus of educational monitoring in Germany:

1. Participation in international large-scale assessments (PIRLS, PISA, TIMSS)
2. Evaluation and implementation of educational standards [*Bildungsstandards*]: national assessments that enable comparisons across federal states and evaluate whether students meet the educational standards defined for specific subjects in specific grades; focus on the end of primary, secondary and continued secondary education with centralized tests at the end of the first two phases and the provision of a central pool of tasks for the final examination qualifying students for university entrance
3. Quality assurance on the school level: state-specific and cross-state implementation of assessments, making it possible to compare the performance of individual schools and classes in order to support instruction and school development (e.g. ‘VERA’ [*Vergleichsarbeiten*])
4. Bi-annual publication of a comprehensive national report on the status of educational system by the Federal Ministry of Education (BMBF) together with all German states

The six thematic areas of particular interest identified by the federal states in 2016 (KMK 2016) are:

- (1) Heterogeneity: individual support in heterogeneous learning groups including special needs and gifted students
- (2) Development of instruction: effects of instructional methods and didactic concepts, usage of evidence-based measures to ensure quality of instruction and school development
- (3) Relevance of teacher education and teacher deployment for students’ academic development
- (4) Effects of measures of school quality assurance
- (5) All-day schools: consequences for learning outcomes
- (6) Effects and strategies of school development: differences between schools in similar settings

## International and National Assessments Today

Today, Germany still participates regularly in core international large-scale student assessments and has joined some additional studies (see Table 12.1). In most cases, study implementation in Germany is commissioned by the Federal Ministry of Education and Research (BMBF) and/or the Standing Conference of the Ministers

**Table 12.1** Germany's current participation in multi-wave international large-scale assessments

Study	Short	International research coordination	National research coordination	Rhythm	Last assessment (and previous assessments)	Next assessment	Main assessment focus	Main target group	Number of participants
Progress in International Reading Literacy Study	PIRLS	IEA	TU Dortmund University	Every 5 years	2016 (2001, 2006, 2011)	2021	Reading	4th graders	~4000 students in ~200 classes
Trends in Mathematics and Science Study	TIMSS	IEA	University of Hamburg	Every 4 years	2015 (1995, 2007, 2011)	2019	Mathematics, science	4th graders	~4000 students in ~200 classes
Programme for International Student Assessment	PISA	OECD	Centre for International Student Assessment (ZIB)	Every 3 years	2018 (2000, 2003, 2006, 2009, 2012, 2015)	2021	Reading, mathematics, science	15-year-olds	~ 10,000 students in ~230 schools
International Computer and Information Literacy Study	ICILS	IEA	Paderborn University	Every 5 years	2018 (2013)	2023	Computer and information literacy	8th graders	~3000 students in ~150 schools
Programme for the International Assessment of Adult Competencies	PIAAC	OECD	GESIS Leibniz Institute for the Social Sciences	Every 10 years	2011/12	2021	Literacy, numeracy, problem solving in technology-rich environments	16- to 65-year-olds	~5000 persons

Note: IEA International Association for the Evaluation of Educational Achievement, OECD Organisation for Economic Co-operation and Development

of Education and Cultural Affairs (KMK). The national research coordinators for the different studies are located at various German universities.

Current national assessments in Germany include the IQB studies evaluating with representative samples from all federal states whether students meet set educational standards. Every 5 years, 4th graders are assessed in German language and mathematics (2011, 2016; upcoming: 2021); and every 3 years, 9th graders are assessed either in mathematics and science (2012, 2018; upcoming: 2024) or in German language or second language English/French (2009, 2015; upcoming: 2021). Yearly assessments of all students in 3rd and 8th grade – in at least German language and mathematics – are the responsibility of the individual federal states and serve the different purpose of helping teachers and administrators to further develop instruction and schools. Focusing on research evidence, Germany also started the National Education Panel Study (NEPS) in 2010 (Blossfeld et al. 2011). This multi-cohort longitudinal study is investigating how education develops from early childhood to old age and the effects education has on other aspects of life. Two starting cohorts (SC 2 and 3) began to follow students from the beginning of primary respectively secondary school education, whereas another starting cohort (SC 4) began with 9th-grade students. Since 2006, the status of the German educational system from early child care institutions, schools, professional education, university education up to adult further training is being reviewed every second year by a national education report commissioned by federal and state educational ministries (Klieme et al. 2003; Autorengruppe Bildungsberichterstattung 2018). These official measures are accompanied by a multitude of additional studies and reports (e.g. the yearly expert report by Aktionsrat Bildung 2018).

## **An Example of International Assessment Findings: PIRLS**

Reading literacy, the focus of PIRLS, is essential for succeeding in academic, working, and everyday life (McElvany et al. 2008). Reading literacy includes the ability to extract relevant information from texts and to understand, use and reflect on written texts (Mullis and Martin 2015). Several national and international studies have shown repeatedly that a substantial number of students have deficits in reading comprehension at the end of primary school (Baumert et al. 2001), thereby indicating the importance of measuring and monitoring students' learning in reading.

PIRLS monitors trends in the reading literacy of 4th graders. Since 2001, PIRLS has been administered every 5 years. Every student reads two texts, one literary and one informational, and works on 12–15 comprehension questions. In addition, students answer questions on their motivation, their attitudes towards reading, and their perception of instructional quality. Furthermore, parents, teachers and school principals complete questionnaires gathering information on students' reading comprehension and the school and family background.

Germany has taken part in every cycle since 2001, and will also participate in 2021 by surveying approximately 4000 4th-grade students from about 200 primary

schools in all 16 federal states. Germany's participation is part of the overall educational monitoring strategy agreed upon mutually and funded equally by the KMK and the BMBF. In Germany, PIRLS is being coordinated by the Center for Research on Education and School Development (IFS) at the TU Dortmund University.

In 2016, Germany's 4th-grade students scored an average of 537 points (Bos et al. 2017). Compared to other countries, Germany ranked in the lower middle range. Nonetheless, this mean score of 537 points is significantly higher than the international average (521 points) and is not significantly lower than the mean score for EU countries (540 points) or all OECD countries (541 points). In 2016, Germany was outperformed by 14 participating states (e.g. Sweden, Italy, Australia) and one benchmark state. In the long-term perspective from 2001 to 2016, there has been no significant change in German students' reading achievement (2001: 539, 2006: 548, 2011: 541 and 2016: 537 points). This result is comparable to other countries such as Sweden, Denmark, or Bulgaria. Even though there was no significant increase in students' reading achievement, the proportion of students on the highest competence level has increased (Competence level V: from 8.6% in 2001 to 11.1% in 2016). Parallel to this, however, there has also been an increase in low-level readers (under competence level III: from 16.9% in 2001 to 18.9% in 2016; Bos et al. 2017). This indicates that the heterogeneity of achievement has become larger. Indeed, students' achievement variance (78 points) was very high in 2016 and comparable to that in countries such as Hungary and Lithuania (Bos et al. 2017).

PIRLS differentiates between literary and informational reading literacy. In Germany, 4th-grade students had higher scores on literary reading literacy (542 points) than on informational reading literacy (533 points). This finding is comparable to 19 other participating countries, although the difference between literary and informational reading literacy is exceptionally large in Germany (Bos et al. 2017).

As in all other countries except Macao SAR and Portugal, girls in Germany (2016) showed a higher average achievement than boys did. Girls outperformed boys especially in literary reading literacy (18 points). The difference was notably smaller (5 points) for informational reading literacy. In 2001, the difference between girls and boys was almost the same.

Fourth-grade students from families with more than 100 books in the home had a 54-point achievement advantage over students with fewer books in the home. This result is similar for all participating states. Alongside Slovenia, Slovakia and Hungary, Germany is one of the four states in which social disparities have increased significantly since 2001. Furthermore, there is also a migration-related gap in reading achievement. Students who did not speak the test language at home had a lower average score than native speakers. In Germany, this difference amounts to 37 points. This means in effect that these students had a disadvantage of 1 year of learning. Students whose parents were born in Germany scored an average of 48 points more than students whose parents were born in a foreign country. All in all, achievement disparities between 4th graders with and without a migration background have remained constant over the last 15 years (Bos et al. 2017).



Related to student achievement is student motivation. In Germany, most 4th graders were highly motivated (2016:  $M = 3.18$ ,  $SE = 0.03$  on a 4-point scale ranging from 1 [*disagree a lot*] to 4 [*agree a lot*]). However, from 2001 to 2016, there has been a decrease in reading motivation, especially in students with reading difficulties. There were also gaps between girls and boys as well as between students with and without a migration background: girls had a higher reading motivation than boys did, and students without a migration background were more motivated than those with a migration background. Additionally, approximately 70% of students read once or twice a week in their free time. Again, girls read more than boys, and students without a migration background read more than those with a migration background (Bos et al. 2017).

To some extent, student achievement and student motivation are an outcome of instructional quality (Hattie 2008). Instructional quality can be divided into three major domains (Hamre and Pianta 2010): classroom management, cognitive activation and emotional support. Classroom management was perceived as very efficient by 39% of students; 60% of those students who rated classroom management as being efficient belonged to the group of high achievers. For the cognitive activation domain, 57% of students felt that they receive strong cognitive activation from their teacher. Most of these students (49%) were high achievers. The last domain, emotional support from the teacher, rated most positive: nearly 73% of students perceived that they got a strong emotional support from their teacher. Of this group, 55% were high achievers (Stahns et al. 2017).

The increasing heterogeneity of classrooms makes it necessary to determine which factors are most relevant for the acquisition of reading literacy under such changing conditions. Both the quantity and the quality of PIRLS data make it possible to examine these questions in detail. Drawing on PIRLS 2016, Hartwig et al. (submitted) combined student and teacher data to analyze whether the interplay between student heterogeneity in cognitive abilities and teachers' attitudes towards heterogeneity such as perceived costs and utility or instructional behavior (especially differentiated instruction) influence students' reading literacy. Hartwig et al. (submitted) found positive relations between teachers' perceived utility and students' reading literacy as well as negative correlations between differentiated instruction and reading literacy on the classroom level. In addition, they found that students' heterogeneity in cognitive abilities related positively to their reading achievement. After controlling for the mean cognitive abilities in classes, only the path between teachers' differentiated instruction and students' reading literacy remained statistically significant (Hartwig et al. submitted).

Germany benefits greatly from participation in PIRLS. PIRLS empirical data have been used to initiate different policies to increase equity in educational opportunity (Wendt et al. 2017). Based on the results of PIRLS 2001, 2006, 2011, and 2016, several measures should be implemented to establish equal opportunities independent of students' gender or migration background, to promote reading also beyond primary school, or to give equal support to both low and high achievers (Valtin 2017).

The fifth PIRLS cycle in 2021 will provide data spanning two decades. Additionally, PIRLS 2021 will offer several new initiatives. Through the transition to a digital format, PIRLS will also be assessing informational and literary reading digitally. In addition, ePIRLS, which was initiated in 2016, will measure students' online informational reading competencies. Another reform will be the national school panel: about 120 schools that already participated in PIRLS 2016 will be retested. This national school panel can be used to analyze longitudinal processes as well as developments on a school level. In Germany, PIRLS 2021 will focus on topics such as instructional quality while taking account of digitalization, multi-criteria goal attainment, and current topics such as the integration of refugee students and mainstream inclusion.

## Critical Discussion

Germany's performance in international and national assessments varies depending on the age group and the domain under investigation. Nevertheless, there are strong overall signs of a failure to achieve satisfactory success on such important goals as (a) increasing average results, (b) enlarging the high-performing group, (c) reducing the low-performing group, (d) reducing the correlation between socio-economic family background and performance level, and (e) providing more effective support for students with a migrant family background. The findings from international large-scale assessments have had and continue to have a substantial impact on many levels of education in Germany. Core consequences are the shift from an input to an output orientation and the subsequent continuous evaluation of the educational system and outcomes based on quantitative empirical data. These adjustments in approach have had comparable effects on educational policies, educational administration, educational practice, and even educational science. The 16 federal states of Germany have agreed on a joint overall strategy for education monitoring through international, national, and state assessments as well as continuous reporting. Substantial funds are being invested in educational monitoring including a national academic institute set up by the federal states in addition to their individual state institutes. The BMBF has launched a comprehensive framework for empirical educational research that is currently in its second phase focusing on (1) increasing educational equality by identifying and developing individual potentials, (2) dealing with diversity and strengthening societal cohesion, (3) supporting quality in the educational system, and (4) designing and using technological developments in education (BMBF 2018). Both quantitative and qualitative research are being encouraged, and more emphasis is being given to the practical relevance of possible research findings for implementation in educational practice. Educational research has also been supported significantly in recent years by the strong increase in university chairs of empirical educational research. Despite some disputes between the

(traditionally more quantitatively and empirically focused) educational psychology and the (habitually more theoretically or qualitatively and empirically oriented) educational science, recent years have seen a productive dialogue and cooperation between the disciplines involved in education. This development ultimately led to the founding of an interdisciplinary Society for Empirical Education Research (GEBF) in 2012 as well as a new interdisciplinary open-access online journal (Journal for Educational Research Online, JERO) in 2009. Furthermore, evidence-based thinking has changed teacher training, instructional approaches, and school development in many ways.

Nevertheless, there have also been criticisms of the German educational monitoring strategy and developments over the last two decades. These include its focus on a few core domains with the potentially negative consequences for other subjects such as the arts and history in terms of, for example, appreciation, attention, effort invested in further development, or funding. Worries have also been expressed about schools and teachers using teaching-to-the-test strategies (Volante 2004). On a more general level, criticism has questioned the utility approach underlying the selection of domains and the resulting shift away from the idea of education for the primary sake of human development. This also raises the question of how to define the desired outcome of education and how performance skills such as reading and math relate to other outcomes of educational processes such as social, emotional, or personal skills; personality; attitudes; and motivational characteristics. Similarly, there are concerns that educational research itself has been mainstreamed into a service discipline for educational administration with research money and positions being awarded only to researchers and research closely linked to political interests and priorities (Bormann 2015). Regarding methodological issues, critics have expressed concern over the general ability of standardized tests to measure complex domains and the exclusion of entire sub-areas due to the lack of any opportunity to measure them in the current frameworks. Finally, yet importantly, the costs related to the international and national assessments have been criticized, arguing that these funds could otherwise be invested directly in the educational system and in improving its quality.

Educational monitoring and assessment currently face multiple challenges. One is the shift from paper-and-pencil to computer-based assessments. The new mode of assessment based on digital devices opens up new opportunities regarding item formats and which skill sets can be investigated. However, there are many methodological issues regarding trend analyses and construct (in)equivalence that have yet to be resolved. Looking at the assembly of LSA in Germany, it also becomes clear that despite the acknowledged importance of early education, the earliest assessment is performed in grade 4 (at age 10) and no (international) standardized assessments are being implemented in early childhood. Important new developments apart from the digitalization of assessments include increased interest in the implementation of school panels in LSA (see, e.g. the school panel within PISA 2000–2009 in Germany). Hence, for example, the school panels planned for PIRLS 2016 and 2021 will make it increasingly possible to combine evidence on school effectiveness with measures that are directly relevant for school development.

## References

- Autorengruppe Bildungsberichterstattung. (Eds.). (2018). *Bildung in Deutschland 2018. Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung* [Education in Germany 2018. An indicator-based report with an analysis on the effects and earnings of education]. Bielefeld: wbv.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (Eds.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* [PISA 2000: Basic competencies of students in an international comparison]. Opladen: Leske + Budrich.
- Blossfeld, H.-P., von Maurice, J., & Schneider, T. (2011). The national educational panel study: Need, main features, and research potential. *Zeitschrift für Erziehungswissenschaft* [Journal for Education], 14, 5–17.
- Blossfeld, H.-P., Bos, W., Daniel, H.-P., Hannover, B., Köller, O., Lenzen, D., McElvany, N., Roßbach, H.-G., Seidel, T., Tippelt, R., & Wößmann, L. [Aktionsrat Bildung]. (2018). *Aktionsrat Bildung: Digitale Souveränität und Bildung* [Action council education: Digital sovereignty and education]. Münster: Waxmann.
- Bormann, I. (2015). *Unsicherheit und Vertrauen* [Uncertainty and trust]. *Paragrana*, 24, 151–163.
- Bos, W., Lankes, E.-A., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Eds.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* [First results from PIRLS. Student achievements at the end of the fourth grade in an international comparison]. Münster: Waxmann.
- Bos, W., Valtin, R., Hußmann, A., Wendt, H., & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick [PIRLS 2016: Important results]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kapser, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [PIRLS 2016. Reading literacy of primary school students in Germany in an international comparison] (pp. 13–28). Münster: Waxmann.
- Bundesministerium für Bildung und Forschung. (2018). *Rahmenprogramm empirische Bildungsforschung*. Bonn/Berlin: BMBF – Referat Bildungsforschung.
- Hamre, B. K., & Pianta, R. (2010). Classroom environments and developmental processes: Conceptualization, measurement, & improvement. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 25–41). New York: Routledge.
- Händle, C., Oesterreich, D., & Trommer, L. (Eds.). (1999). *Aufgaben politischer Bildung in der Sekundarstufe I. Studien aus dem Projekt Civic Education* [Tasks of political education at secondary level I. Studies from the project Civic Education]. Wiesbaden: Springer Fachmedien.
- Hartwig, S., Schwabe, F., & McElvany, N. (submitted). *The interplay between teacher attitudes towards heterogeneity, differentiated instruction, and students' reading competence*.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Husén, T. (1967). *International study of achievement in mathematics: A comparison of twelve countries, volume I–II*. Stockholm: Almqvist & Wiksell.
- 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. Eine Expertise* [The development of national educational standards. An expertise]. Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2004). *The development of national educational standards: An expertise*. Berlin: Federal Ministry of Education and Research (BMBF).

- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (1997). Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland – Konstanzer Beschluss – [Fundamental Considerations on Performance Comparisons within the Federal Republic of Germany – Constance Resolution – ]. Retrieved online: [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/1997/1997\\_10\\_24-Konstanzer-Beschluss.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1997/1997_10_24-Konstanzer-Beschluss.pdf) [05.03.2020].
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. [Comprehensive strategy by the Standing Conference for educational monitoring]. München: Luchterhand.
- KMK – Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. [Comprehensive strategy by the Standing Conference for educational monitoring]. Köln: Wolters Kluwer.
- Lang, M., & Schulz-Zander, R. (1994). Informationstechnische Bildung in allgemeinbildenden Schulen – Stand und Perspektiven [Information technology education in general education schools – status and perspectives]. In H.-G. Rolff, K.-O. Bauer, K. Klemm, H. Pfeiffer, R. Schulz-Zander (Eds.), *Jahrbuch der Schulentwicklung* [Yearbook of school development] (Vol. 8) (pp. 309–353). Weinheim: Juventa.
- Lehmann, R., Peek, R., Pieper, I., & von Stritzky, R. (1995). *Leseverständnis und Lesegewohnheiten deutscher Schülerinnen und Schüler* [Reading comprehension and reading habits of German students]. Weinheim: Beltz.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Kennedy, A. M. (2003). *Trends in children's reading literacy achievement 1991–2001: IEA's repeat in nine countries of the 1991 Reading Literacy Study*. Chestnut Hill: Boston College.
- McElvany, N., Kortenbruck, M., & Becker, M. (2008). Lesekompetenz und Lesemotivation: Entwicklung und Mediation des Zusammenhangs durch Leseverhalten [Reading literacy and reading motivation: Their development and the mediation of the relationship by reading behaviour]. *Zeitschrift für Pädagogische Psychologie* [Journal of Educational Psychology], 22, 207–219.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 assessment framework* (2nd ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website. <http://timssandpirls.bc.edu/pirls2016/framework.html>
- OECD. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264195905-en>.
- Prenzel, M., Stadtler, M., Friedrich, A., Knickmeier, K., & Ostermeier, C. (2009). *Increasing the efficiency of mathematics and science instruction (SINUS) – A large scale teacher professional development programme in Germany*. Kiel: Leibniz-Institute for Science Education (IPN).
- Schultze, W. (1975). *Die Leistungen im Englischunterricht in der Bundesrepublik im internationalen Vergleich* [Learning achievement in English in Germany in an international comparison]. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung (Sonderheft Mitteilungen und Nachrichten).
- Schultze, W., & Riemenschneider, L. (1967). *Eine vergleichende Studie über die Ergebnisse des Mathematikunterrichts in zwölf Ländern* [Learning achievement in math in Germany in an international comparison]. Frankfurt a.M.: Deutsches Institut für Internationale Pädagogische Forschung (Sonderheft Mitteilungen und Nachrichten Nr. 46/47).
- Stahns, R., Rieser, S., & Lankes, E.-M. (2017). Unterrichtsführung, Sozialklima und kognitive Aktivierung im Deutschunterricht in vierten Klassen [Classroom management, social climate, and cognitive activation in German lessons in fourth grade]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [PIRLS 2016. Reading literacy of primary school students in Germany in an international comparison] (pp. 251–278). Münster: Waxmann.

- Valtin, R. (2017). Einordnung der IGLU-2016-Befunde in das europäische Rahmenkonzept für gute Leseförderung. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kapser, E.-M. Lankes, N. McElvany, T. C. Stubbe & R. Valtin (Eds.), *IGLU 2016. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* [PIRLS 2016. Reading literacy of primary school students in Germany in an international comparison] (pp. 315–328). Münster: Waxmann.
- Volante, L. (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. Retrieved from: <https://files.eric.ed.gov/fulltext/EJ848235.pdf>
- Wendt, H., Walzebug, A., Bos, W., Smith, D. S., & Bremerich-Vos, A. (2017). Germany. In I. V. S. Mullis, M. O. Martin, S. Goh, & C. Pendergast (Eds.), *PIRLS 2016 Encyclopedia: education policy and curriculum in reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/encyclopedia/>