

Chapter 2

International Student Assessment: Aims, Approaches and Challenges



Miyako Ikeda and Alfonso Echazarra

Introduction

International student assessments are essential for improving education around the world. They fuel debate and provide powerful information and data to help educators and policy makers identify strengths and weaknesses of their school systems. Assessments serve to raise awareness and calls for accountability in the public eye. Participating countries in the Programme for International Student Assessment (PISA) led by the OECD (Organisation for Economic Co-operation and Development), for instance, have almost tripled over the last 20 years since the first assessment in 2000. Interest in international benchmarking of student performance continues to increase. In 2000, 31 countries participated in PISA, of which 28 were OECD members. Approximately 80 countries and economies took part in 2018 PISA. For PISA 2021 even more have already committed to participate. The other organisation which conducts major international student assessments on mathematics, science and reading is the International Association for the Evaluation of Educational Achievement (IEA). The Trend in International Mathematics and Science Study (TIMSS) was conducted by the IEA in 1995, 1999, 2003, 2007, 2011 and 2015, with another planned for 2019. The Progress in International Reading Literacy Study (PIRLS) by the IEA was conducted in 2001, 2006, 2011 and 2016, with another intended for 2021.

M. Ikeda (✉) · A. Echazarra
Organisation for Economic Co-operation and Development, Paris, France
e-mail: Miyako.IKEDA@oecd.org

International Student Assessments: An Overview

International student assessments, such as PISA and TIMSS/PIRLS, contribute to education policies and practices at the national level in three important respects. First, they provide reliable and internationally comparable indicators on student performance and other education outcomes and facilitate the monitoring of shifts over time. Second, analysing how international student assessments' data correlate with contextual information can contribute to improve education systems, schools and teacher quality. Finally, international assessments carry great value for the formulation of national policies and practices because they frame current educational debates and highlight internationally agreed-upon metrics and methodologies.

Major international student assessments are generally low stakes for individual students and schools. Survey designs and sampling are typically optimised to obtain results at the country or sub-national level, and assessment results are mainly intended for system-level analysis. Currently, major assessments share similar methodological and implementation steps in achieving their outcomes. First steps include developing a framework and designing appropriate instruments, creating a survey design and sampling plan, and establishing a standardised implementation procedure. These guide the development of operation manuals and various trainings for participants. This is followed by the translation and validation of survey instruments, the drawing of samples and the collection of data. The final steps are the processing and coding of data, the computation of weights, the development of scales for student performance and relevant background data, and the preparation of the database for public access. To maximise data access and information use, detailed publications and technical documents are prepared, published and disseminated online and in print, usually free of charge.

In general, international student assessments aim to collect data to benchmark student performance and to provide comparable indicators across participating countries. Student performance is measured through carefully developed tests based on an agreed-upon assessment framework. Many countries currently create and implement their own national education assessments to measure a variety of domains and interest areas. However, these assessments rarely provide results that allow for a direct and comprehensive international comparison. By participating in international student assessments, countries can compare their students and education systems directly with others. Participation in these assessments over a number of cycles also allows for the monitoring of trends and country performance over time. While current international assessments measure similar outcomes, they also look at different aspects of these outcomes. For instance, PISA focuses on the level of student preparedness for full participation in a society while TIMSS and PIRLS focus more on the level of student mastery of the school curriculum.

International student assessments, in addition to providing data on student performance, also collect contextual information on students, schools and school systems. Because data gathered in such assessments are limited in their capacity to

identify causal inferences, policy makers and educators must determine how best to improve student performance in their own countries. By correlating the contextual information with student performance, however, it is possible to identify student or school groups at risk. Correlational information also help in the examination of education policies and practices shared by high-performing students, schools and countries.

While results of international student assessments help policy makers, researchers and educators identify strengths and weaknesses of a given education system, such assessments provide valuable data and information for countries to learn from one another. For example, the topic that most interests researchers who use PISA data at this time is that of equity (Hopfenbeck et al. 2018). Internationally comparable indicators on equity have shown to be of great interest to them, more specifically the relationship between student socio-economic status and their academic performance. By showing that this relationship exists in essentially all participating countries, policy makers and researchers can use these results to better understand why this relationship is weaker or stronger in certain education systems. The debate on the possible trade-off between equity and overall education quality (e.g. average student performance) continues to be hotly disputed. However, PISA has shown that high education results and equity can be achieved by identifying those countries that have achieved both at the same time (OECD 2010, 2013a, 2016). Background contextual information collected also helps countries understand the roles of school organisation, teaching strategies and practices, the learning environment or parental support. In correlating background information with a variety of education outcome indicators, educators and policy makers can identify target groups that need further support and those policies and practices which are related to the outcomes. This is invaluable for policy makers and educators who must plan, adjust, implement and pursue effective and impactful policies and practices.

A number of education issues become more salient when education systems are held in comparison. The practice of grade repetition is an illuminating example because education systems around the world handle grade repetition and the challenges of diverse student populations differently. Some systems encourage or require students to repeat a grade if they are deemed unprepared for advancement. Other school systems allow students to advance automatically into the next grade regardless of performance and/or behaviour. In comparing rates of grade repetition across countries, educators can better understand how the quality and equity of their education systems are related to grade repetition. In this light, PISA results have shown that 15-year-old students are spread across wider range of grade levels in those education systems featuring grade repetition. Data shows that overall performance in these systems tend to be lower while the impact of socio-economic status on student learning outcomes is higher in those systems that feature more frequent grade repetition (OECD 2010). These PISA findings have consequently contributed to shape national policies on grade repetition. Belgium, France, Portugal or Spain, for example, introduced new policies that reduced grade repetition, lowering rates over recent years (OECD 2018a).

In addition to the important data results gained from international assessments, the frameworks themselves, the methodologies applied and the instruments designed also carry great value since they provide shared points of reference to policy makers, educators and researchers. Available to everyone online and/or in print, assessment frameworks provide the current definition of constructs and metrics to measure student performance. Detailed technical documents outline and describe survey design and methodologies, sampling, instrument development scaling approach, translation and verification processes, survey operation, and database structure and management. Those who are involved in developing and administering national assessments for their own countries often refer to these materials for insight and comparison, as well as for ideas and direction in determining their approach and their methodology. This has helped improve national assessments and fostered synergies between international and national assessments. This also sheds light on the distinct nature of each national assessment.

International student assessments have helped steer policy dialogues internationally and regionally. At this time, internationally comparable indicators on education contribute to the monitoring of the Sustainable Development Goals adopted by the United Nations in September 2015. More specifically, its fourth goal seeks to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. At the regional level, the European Union set objectives for its member education systems that applied PISA indicators as benchmarks (OECD 2013b; European Commission 2018). One target is to reduce the share of under-achieving 15-year-old students in reading, mathematics and science to less than 15% by the year 2020.

Approaches to International Student Assessment: PISA Case Study

Every 3 years since 2000, PISA has assessed the extent to which 15-year-old students near the end of compulsory education have acquired the knowledge and skills that are essential for full participation in modern societies. PISA aims to gauge whether students can reproduce the knowledge they have acquired in and outside of school. It also determines whether students can extrapolate from what they have learned and apply knowledge and skills in unfamiliar contexts. To maintain relevancy and impact, PISA has consistently broadened its assessment of competencies and dispositions with each assessment cycle.

Whereas the mastery of subjects such as mathematics, science and reading and their application are regarded by all as being essential, other skills and dispositions are also recognised for their importance. In response, PISA has included a new domain and/or topics with each of assessments. PISA 2003 included student self-assessment of their learning strategies. For 2006, PISA incorporated an assessment of student attitudes towards science. Both PISA 2003 and 2012 featured assessment sections concerning problem-solving skills. PISA 2012 also offered countries the possibility

of measuring financial literacy. In 2015, PISA assessed student ability to solve problems collaboratively. For PISA 2018, the assessment features a section concerning 15-year-old student capacity to examine local, global and intercultural issues, to understand and appreciate the perspectives and world views of others and to engage in open, appropriate and effective interactions with people from different cultures.

While PISA's main objective remains to provide reliable and comparable measurements of student performance internationally, PISA also collects contextual data that describe school systems and other important aspects. Such data help policy makers and educators improve and raise their national performance standards because they show a granular picture, helping establish relationships between student performance and a range of factors and influences, such as family background, student attitudes towards learning, their habits and their life in and outside of school. The assessment also surveys principals about the staff and material resources in their schools, aspects of school management and funding, the school's curricular emphasis, any extracurricular activities offered and the general context of instruction. Questionnaires for parents and teachers are also available for countries that are interested in learning more from their perspectives.¹

How PISA Differs from Other International Student Assessment Studies

Current international student assessments differ among one another in a number of key areas. For example, the most distinctive between PISA and TIMSS/PIRLS concerns what exactly is being measured. Wagemaker (2008) identified that the assessments embodied the differences in their histories and the aims of the two organisations. From its inception, PISA was created to monitor the extent to which students near the end of compulsory schooling had acquired the knowledge and skills essential for full participation in society. This aim falls within the broader mission of the OECD and embodies its mandate. When PISA was launched in 1997, the OECD initiated the Program Definition and Selection of Competencies (DeSeCo) to develop a conceptual framework which would define key competencies to guide the assessment. Through DeSeCo, the OECD collaborated with a wide range of scholars and specialists of a broad range of disciplines, as well as input from country representatives. They established that PISA results would contribute to valued outcomes for societies and individuals, help individuals meet important demands in a wide variety of contexts and be significant not only for specialists but for all individuals (OECD 2005). This was accomplished by identifying specific challenges and values common across countries and cultures, as well as acknowledging the diversity in values and priorities. Shaped by the DeSeCo framework for key competencies, the

¹ In PISA teacher data are linked to student data at the school level, but not at the individual student level.

first PISA assessment in 2000 took place, focusing predominantly on student competencies in the domains of reading, mathematics and science. The framework established a foundation and pathway for how additional competency domains, which are essential for student success in life, could be incorporated into future assessments (Rychen and Salganik 2003).

The IEA creates assessments focus instead on understanding the linkages between the intended curriculum (what policy requires), the implemented curriculum (what is taught in schools) and the achieved curriculum (what students learn), drawing on the concept of ‘opportunity to learn’ (<https://www.iea.nl/our-studies>). TIMSS and PIRLS are formulated and designed to focus on the teaching/learning process and to assess the extent of knowledge acquired by students after a fixed period of schooling. IEA’s interests lie ‘in addressing questions of efficiency and equity with respect to the ability of educational systems to deliver what is mandated by the curriculum’ (Wagemaker 2008).

PISA differs from those of the IEA in their sampling approach and questionnaire content. PISA applies an age-based sampling with a target population of 15-year-old students who are in grade 7 or above. In contrast, studies by the IEA apply grade-based sampling, regardless of student age. For example, the TIMSS grade 8 target populations are defined as ‘all students enrolled in the grade that represents 8 years of schooling counting from the first year of ISCED Level 1, providing the mean age at the time of testing is at least 13.5 years’ (LaRoche et al. 2016). In terms of the content coverage of questionnaires, IEA survey questions emphasise school curriculum. For example, in those countries which taught science as separate subjects at the grade 8 level (e.g. biology, chemistry, physics and earth science), the TIMSS 2015 survey had students respond to questions specific to each subject, in addition to other aspects of the curriculum and their home and school lives in general. Also, IEA studies had a dedicated questionnaire on curriculum that was completed by national research coordinators from participating countries that is specifically designed to collect information on the national contexts for learning (Hooper 2016). In contrast, the PISA 2015 student questionnaire focused on core subjects (e.g. science) and collects comparatively more information on non-academic outcomes (e.g. career expectations, well-being) and other contextual information concerning student life and well-being. Furthermore, country representatives complete a system-level questionnaire that feature education policy questions on such things as teacher training and support, the structure of the education system and school administrative aspects.

Challenges for the Future of International Student Assessments

Organisations leading international student assessments must ensure relevancy for policy makers, researchers and educators who aim to improve education. To fulfil expectations and address changing needs, international large-scale assessments, and

PISA in particular, are facing numerous challenges. An obvious challenge for international student assessments is managing time constraints. The 3-year PISA cycle provides timely information for countries but also requires the organisations, institutions and experts involved in the collection process to closely coordinate for a timely delivery of results. In addition, any innovation in the design and delivery of international student assessments must be balanced with the response burden for students and other education stakeholders. As with any international project, student assessments also face significant financial challenges. Organisations must keep operation costs, which are typically borne by taxpayers in participating countries, at reasonable levels. Beyond these obvious challenges, there are at least six other challenges for organisations leading international-level student assessments.

Responding to Economic, Social and Technological Changes

The competencies students required to succeed in world today are evolving at an ever-increasing rate. The Internet has changed the way people connect with one another and how individuals access information. People and markets are now interconnected globally in ways unimaginable only decades earlier. The labour market is increasingly seeking non-routine, interpersonal and higher-order skills (OECD 2013c; Frey and Osborne 2017). In response, the PISA assessment has also changed over time with the addition of new content areas and the use of technologies that have made the assessment process more streamlined and flexible. New assessment domains have included problem-solving, digital literacy, collaboration and global competence. Because the framework and questions of traditional domains (reading, mathematics and science) are updated every 9 years, the challenge is introducing new content areas while maintaining a consistency in traditional domains to measure trends over time.

The PISA assessment is now computer based rather than paper-pencil in a majority of countries. This change in 2015 came in response to the rise of digital literacy and the manner in which digital technology has been incorporated into daily life. A computer-based assessment is expected to expand the potential range of how questions can be presented to students and how responses could be given.

Despite these innovation efforts, more is needed to ensure the long-term relevance of international student assessments. Vital skills, such as creativity, entrepreneurship or communication skills, could be considered for assessment, as well as a number of traditional school subjects, such as art, history, geography and music. The ability of students to communicate in a foreign language is another important area that could be measured internationally. In this regard, the European Commission assessed the foreign language skills in 16 education systems in 2011, and a foreign language assessment is being considered as part of PISA 2024. Assessing these new domains for more than half a million students will certainly be a challenge fraught with complexity and debate. However, the inclusion of such subject domains helps

ensure the relevancy and worthiness of international student assessments in our changing world.

Making Assessment Relevant for All

An impressive number of education systems are currently participating in international student assessments. For instance, nearly 60 countries participated in TIMSS 2015 and nearly 80 countries and economies are taking part in PISA 2018. An increased number of education systems greatly enhances the value of benchmarking. However, the participation of middle- and low-income countries brings new challenges. The most significant is ensuring that international student assessments can accurately measure the knowledge, skills and learning contexts of a complex diversity of student populations. To address this challenge, the OECD initiated *PISA for Development* in 2014, a project designed to incorporate middle- and low-income countries into the main PISA assessment. This project adds more items at the lower end of the performance distribution, creates survey instruments that are more relevant to the context of these countries and offers countries the possibility of sampling out-of-school children. It also helps countries in survey implementation and national report development. Together, these initiatives have proven extremely beneficial and are being progressively incorporated into the main 2018 and 2021 PISA assessments. Greater flexibility and continual adjustments are still required to make international large-scale assessments equally relevant to all countries, particularly to middle-income countries.

Making Results Useful for All Stakeholders

Critics of international student assessments have highlighted that such projects appear to mainly serve policy makers and researchers. They argue that this has limited the relevancy of international student assessments for other important groups of education stakeholders, namely, schools, parents and teachers (Carabaña 2015). An OECD report evaluating the policy impact of PISA (OECD 2008) in fact cited that a majority of policy makers and researchers reported to be knowledgeable about PISA whereas only a third of parents and school principals reported similarly. To some extent, the relevance of international student assessments is limited by their design because they are not intended to present data at the individual classroom, school or even school district level or to have the results providing direct feedback to participating students and schools. To remedy this, linking international student assessments to national or regional assessments, either by having a subset of students taking both assessments or including some items from international student assessments in national assessments, can position students or individual schools on international scales and within the framework of international standards. This would

immediately raise the awareness of international student assessments, their role in international benchmarking, and increase the potential of raising learning standards.

In the case of PISA, whose natural target audience are policy makers, great efforts continue to be made in reaching out to other stakeholders who might benefit from its results by adjusting questionnaire content.

For example, the *PISA for Schools project*—in which a group of schools is invited to participate voluntarily—provides direct feedback to individual schools on the abilities and learning opportunities of their students using PISA as a benchmark. PISA has created special publications for other stakeholders, such as for parents (*Let's Read Them a Story*), for teachers (*Ten Questions for Mathematics Teachers and How PISA Can Answer Them*; *Qudwa: Global Teachers' Forum*) and for those who are interested in specific areas such as environment, gender and digital technology in education (*Green at Fifteen?*; *The ABC of Gender Equality*; *Students, Computers and Learning*).

Improving Test and Questionnaire Reliability

The test design and scaling procedures in international student assessments undergo regular update to improve and to incorporate current advances in the field. PISA 2015, for instance, increased the number of common items, transitioned from a paper-based to a computer-based assessment and applied a more flexible statistical model for scaling (e.g. two-parameter model or 2PL). The persisting challenge to improve reliability, however, lies in the cross-cultural comparability of questionnaire scales, most notably that concerning scalar invariance (i.e. whether the average of a certain indicator can be compared across cultures). PISA is addressing this issue in a number of ways, including working closely with field and technical experts, triangulating data sources whenever possible and being innovative in the questionnaire design. These include anchoring vignettes, using forced-choice items, reversed keyed items, including a reference point, and the use of various item formats in the field trial. In late 2018, in response to the clear need, the OECD gathered specialists and researchers from around the world for a conference that focused on novel approaches in the fields of measurement equivalence and invariance testing.

Drawing Causal Inferences

There is a recognised place of international student assessments in evaluating education systems and in the formulation of evidence-based policy decisions. However, the drawing of causal inferences from cross-sectional observational studies is problematic (Rutkowski and Delandshere 2016). PISA reports carry a disclaimer stating clearly that PISA cannot identify cause-and-effect relationships between policies/practices and student outcomes. Researchers can reduce the uncertainty

around causal inferences if analytical methods that rest on specific underlying assumptions are applied. These include using propensity score analysis (Hogrebe and Strietholt 2016; Kaplan 2016), instrumental variables (Pokropek 2016), applying a difference in differences approach (Rosén and Gustafsson 2016) or conducting cross-subject analysis with student-fixed effects (Bietenbeck 2014; Echazarra et al. 2016; Schwerdt and Wuppermann 2011).

In general, design of international student assessments can better accommodate causal analyses by integrating experimental and longitudinal studies within their frameworks. For instance, they can facilitate post-testing to measure the effectiveness of educational interventions or encouraging more countries to follow sampled students into the future. Countries that have implemented longitudinal studies following PISA studies include Australia, Canada, Denmark and Switzerland (OECD 2018b). Questionnaires focusing on collecting more information on all assessment subjects would increase the number of possible cross-subject analyses with student-fixed effects. Questionnaires could also include items that are retrospective with the aim of collecting data on the cumulative experience of students. This would provide a more holistic picture rather instead of a simple snapshot of student experience. Video studies within the international student assessment structure have proven to aid in better understanding classroom practice (Cuban 2013) with one example being the 1999 TIMSS Video Study. The upcoming TALIS (*Teaching and Learning International Survey*) Video Study will also look into the classroom dynamics in eight countries.

Enhancing Transparency and Communication

Organisations leading international student assessments have always prioritised transparency and communication. In every cycle, PISA makes the database, frameworks and questionnaires publicly available, explaining all technical aspects in a dedicated report (OECD 2017). It also presents the results in multiple formats (e.g. reports, country notes, PISA in Focus, blogs, working papers, slides, infographics). For illustrative purposes, PISA also releases a number of actual test questions. PISA is a collaborative effort in which the OECD and countries are supported by many actors, including international and national experts and specialised contractors. The governance structure of PISA requires a broad range of actors to participate, contribute and help in the design and implementation of the project. Rigorous technical standards guide activities at all stages of the assessment. Yet a surprising number of education stakeholders, most notably parents, teachers and principals, remain unaware of the process behind international student assessments, often viewing these important projects with suspicion. For instance, in an evaluation of the impact of PISA (OECD 2008), only a small share of parents, principals and media and business representatives reported to be aware of the manner in which the PISA assessment was planned, coordinated and implemented in participating countries. Clearly, organisations like the OECD and the IEA can improve in the way they inform the

public about their international student assessments. More effectively informing the public and their stakeholders about their projects, their purposes and their benefits will demystify the assessments and improve transparency. Simplifying pathways to the information on assessment designs and supporting explanatory and technical materials will broaden their reach and ensure that the data can contribute in meaningful and impactful ways.

Conclusion

International student assessments have done much to help improve education around the world. The OECD and IEA are but two major organisations that currently formulate, design and implement these assessments. Shaped by the missions of their organisations, PISA and TIMSS/PIRLS collect data for different purposes. To maintain relevancy, international student assessments must evolve with societal and public needs while remaining rigorous and robust. Researchers and policy makers have recognised the contribution that international student assessments have made to improve our knowledge base of education. However, organisations leading these assessments must work more deliberately to give all stakeholders the ability to fully access their results. If the right people employ the results of international student assessments correctly, student learning around the world can improve immensely.

References

- Bietenbeck, J. (2014). Teaching practices and cognitive skills. *Labour Economics*, 30, 143–153.
- Carabaña, J. (2015). *La inutilidad de PISA para las escuelas*. Madrid: Los Libros de la Catarata.
- Cuban, L. (2013). *Inside the black box of classroom practice: Change without reform in American education*. Cambridge, MA: Harvard Education Press.
- Echazarra, A., Salinas, D., Méndez, I., Denis, V., & Rech, G. (2016). How teachers teach and students learn: Successful strategies for school. In *Education working papers 130*. Paris: OECD Publishing.
- European Commission. (2018). *Education and training monitor 2018*. Luxembourg: Publications Office.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.
- Hogrebe, N., & Strietholt, R. (2016). Does non-participation in preschool affect children's reading achievement? International evidence from propensity score analyses. *Large-scale Assessments in Education*, 4(1), 2.
- Hooper, M. (2016). Developing the TIMSS 2015 context questionnaires. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, MA.
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353.

- Kaplan, D. (2016). Causal inference with large-scale assessments in education from a Bayesian perspective: A review and synthesis. *Large-Scale Assessments in Education*, 4(1), 7.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sample design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, MA.
- Organisation for Economic Co-operation and Development. (2005). *The definition and selection of key competencies: Executive summary*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2008). *The policies that inspired PISA – and the policies that PISA inspired: 2009 report on the evaluation of the policy impact of PISA, OECD document EDU/PISA/GB(2008)35/REV2*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: What makes a school successful? Resources, policies and practices* (Vol. IV). Paris: PISA/OECD Publishing.
- Organisation for Economic Co-operation and Development. (2013a). *PISA 2012 results: Excellence through equity: Giving every student the chance to succeed* (Vol. II). Paris: PISA/OECD Publishing.
- Organisation for Economic Co-operation and Development. (2013b). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2013c). *OECD reviews of evaluation and assessment in education: Synergies for better learning*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: PISA/OECD Publishing.
- Organisation for Economic Co-operation and Development. (2017). *Technical report*. Paris: PISA/OECD Publishing.
- Organisation for Economic Co-operation and Development. (2018a). *Education policy outlook 2018: Putting student learning at the centre*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2018b). *Equity in education: Breaking down barriers to social mobility*. Paris: PISA/OECD Publishing.
- Pokropek, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-scale Assessments in Education*, 4(1), 4.
- Rosén, M., & Gustafsson, J. E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-scale Assessments in Education*, 4(1), 5.
- Rutkowski, D., & Delandshere, G. (2016). Causal inferences with large scale assessment data: Using a validity framework. *Large-scale Assessments in Education*, 4(1), 6.
- Rychen, D. S., & Salganik, L. H. (Eds.). (2003). *Key competencies for a successful life and well-functioning society*. Gottingen: Hogrefe Publishing.
- Schwerdt, G., & Wuppermann, A. C. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review*, 30(2), 365–379.
- Wagemaker, H. (2008). Choices and trade-offs: Reply to McGaw. *Assessment in Education: Principles, Policy & Practice*, 15(3), 267–278.