IEA COMPASS
BRIEFS IN EDUCATION

Researching education, improving learning

NUMBER 10 SEPTEMBER 2020

INTERNATIONAL LARGE-SCALE ASSESSMENTS IN EDUCATION:

A BRIEF GUIDE

SUMMARY

International large-scale assessments (ILSAs) are one of the most important tools policymakers and other educational stakeholders have to inform evidence-based decision making for educational reform. Despite this, and the widespread use of ILSA data, results are sometimes misunderstood or misinterpreted. Here, we offer a brief guide to ILSAs and illuminate some of the important differences and commonalities within and across studies, limitations, and why they remain one of our most significant tools for education evaluation and reform. We focus on and compare the key studies, approaches, and structure of our own organization, the International Association for the Evaluation of Educational Achievement (IEA), with other ILSAs.

International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.
Website: www.iea.nl
Follow us:

@iea_education
IEAResearchInEducation
IEA

IMPLICATIONS

- IEA conducted the first ILSA study in 1958 but the mandates for ILSAs have changed over their history and new mandates are constantly developing. Most recently, focus has shifted to educational outcomes rather than inputs. ILSA results should be understood in the context of their changing remit.
- There are substantial differences between organizations and approaches to assessments. From the review process to decision-making and fees, fundamental differences should be accounted for when understanding results.
- Key ILSAs do share a common methodology marked by high-quality standards carefully defined to achieve each step in the ILSA process and, further, the data are accompanied by detailed supporting technical reports to assist in interpreting and reporting results.
- Limitations of ILSAs need to be acknowledged when understanding and reporting results. In particular, care must be taken when considering results from ILSAs as a holistic quality measure of the education system.
- Despite limitations, the assessments are unique, monitoring systems over time within a robust international framework, being largely independent of any single political system, and with data freely available to the public. When properly understood and analyzed, the data from ILSAs provide valuable opportunities to help inform policy decisions and research.

INTRODUCTION

International large-scale assessments (ILSAs) of education are empirical studies that assess educational abilities around the world. The data are used in various ways to help inform policymakers, educational researchers, and the general public. However, despite the widespread use of ILSA data, how to interpret and report results is often misunderstood. Different study results are sometimes reported or, interpreted to mean the same thing, yet there exists important differences that need to be accounted for when using results. As leaders in the field of international large-scale assessment, our intention for this brief is to provide context for a better understanding of ILSA results and; how they should be interpreted and reported—we discuss what ILSAs are, the history of their development, differences as well as commonalities in approaches, organization, and methodology, and important limitations—and to express why we believe that ILSAs are unique, important, and relevant tools for understanding educational systems and student achievement around the world, and informing evidence-based change.

WHAT ARE ILSAs?

ILSAs assess student achievement in specific disciplines and provide context for the results by collecting additional data at the student level. Further contextual details at the teacher, principal, and/or system levels may also be collected. To provide statistically valid results, a representative sample of schools (usually around 150 to 200 schools) are drawn from each participating country or education system, and a group of students are randomly drawn from within each of the sampled schools, either by sampling entire classrooms or by sampling students across classrooms.

BRIEF HISTORY

The first ILSA, IEA's Pilot Twelve-Country Study (Foshay et al. 1962), was launched by a group of researchers in 1958 (Husén 1983; IEA 2018). Scholars from various disciplines met at the UNESCO Institute for Education in Hamburg (Germany) and decided to launch a then exploratory study to test whether it was possible to compare learning outcomes across a range of different countries and cultures. They chose to assess student achievement in mathematics, assuming that it would be easiest to translate into different languages and was thus, more likely to result in valid comparisons across countries. Their aim was to find out what could be learned through international assessment, with the hope that countries could learn from each other. As Torsten Husén phrased it: "In general terms, international studies such as this one can enable educationalists (and ultimately those responsible for educational planning and policy making) to benefit from the educational experiences of other countries. It helps educationalists to view their own system of education more objectively because for the first time many of the variables related to educational achievement had to be quantified in a standardized way" (Husén 1967, pp. 13-14).

While the first ILSAs were conducted by researchers with quite minimal resourcing to satisfy academic interests in investigating education, by 1990 educational policymakers had begun to realize that ILSAs could potentially provide useful evidence-based data. However, their financial support for ILSAs demanded rapid outcomes. Where the first academic study reports were sometimes launched up to eight years after the study was conducted (Anderson et al. 1989), this wider interest led to a new pressure to publish results as soon as possible. An additional consequence was an interest in measuring trends in education systems, a challenge that IEA rose to meet with TIMSS, first conducted in 1995 and followed up by a second

The most well-known ILSAs are the **core studies** of the International Association for the Evaluation of Educational Achievement **(IEA)** and the Organisation for Economic Co-operation and Development **(OECD)**:

- IEA Trends in International Mathematics and Science Study **(TIMSS)**
- IEA Progress in International Reading Literacy Study **(PIRLS)**
- IEA International Civic and Citizenship Education Study **(ICCS)**
- IEA International Computer and Information Literacy Study **(ICILS)**
- OECD Programme for International Student Assessment **(PISA)**

Source: OECD 2019; IEA 2019

trend assessment four years later in 1999. The four year gap was chosen because TIMSS assessed grade four and grade eight students in 1995; in 1999, only grade eight students were assessed, the aim being to compare two cohorts of grade eight students' results, and assess the same cohort of grade four students four years later.

In the 1990s there was also an emerging international interest to have strong educational data to better understand economic growth. As such, the OECD, one of the leading international economic organizations, placed a greater emphasis on education and the measurement of educational outcomes, and launched its own assessment study, PISA, in 2000. Previously the OECD had only used other sources—including IEA studies and their annual publication Education at a Glance—for their educational work. However, the organization decided to focus on the skills needed to operate in a modern economy rather than on assessing what schools were teaching. Specifically, PISA claims to assess what the OECD believes 15-year-old students enrolled in school should know in reading, mathematics, and science literacy; the assessment is conducted every three years. Originally, the study focused on OECD countries, but an increasing number of non-OECD countries now take part in the assessment.

Toward the end of the 1990s and into the first decade of the 2000s, regional ILSAs were initiated including the LLECE (Laboratoria Latinoamericano de Evaluación de la Calidad de la Educatión) in Latin America, and PASEC (Programme d'Analyse des Systèmes Educatifs de la CONFEMEN) and SACMEQ (The Southern and Eastern Africa Consortium for Monitoring Educational Quality).

More recently, since 2015 when the Sustainable Development Goals (SDGs) were declared by the UN, a new emphasis for assessments has emerged. In contrast to the Millennium Development Goals, the SDGs focus on educational outcomes of education systems rather than on inputs like expenditure on education. This results in all countries being urged to report on the percentage of students reaching minimum proficiency levels and clearly constitutes a new challenge for ILSAs. In this regard, IEA is playing an active role through the implementation of the Rosetta Stone Project[1], in collaboration with the UNESCO Institute for Statistics, LLECE, and PASEC. The objective is to develop a concordance table that translates scores resulting from regional mathematics and reading assessments with the TIMSS and PIRLS scales.

DIFFERENT APPROACHES = DIFFERENT ASSESSMENTS

Although the best known ILSAs feature a number of similarities, there are also some substantial differences that need to be considered when comparing the results for different educational systems. In the following table we have compiled some notable differences that test consumers need to be aware of when comparing study results conducted by the IEA and OECD.

Table 1: Difference in approaches between IEA studies and PISA

ASPECT	IEA STUDIES	PISA
Study philosophy	Seeks to measure what is taught in schools and the contexts of learning.	Seeks to measure selected acquired skills of students towards the end of their compulsory education.
Content selection	The curricula of the participating countries are analyzed. Participating countries then jointly develop the assessment framework and test materials to ensure national interests are acknowledged.	OECD-selected experts determine the skills that they think students should have mastered for use later in life and assemble the study instruments accordingly.
Cohort selection	Samples are grade based (grades four, eight, or twelve) to reflect the structure of curricula and to establish a direct link to the subject teachers.	Samples are age based (15-year-olds for PISA) to assess a generation, independently of their school pathways or grade distribution.

1. https://www.iea.nl/studies/additionalstudies/rosetta

DIFFERENT ORGANIZATION

Beyond the studies themselves, the two organizations—IEA and OECD—differ in many regards (see Table 2). Most fundamentally, their missions are not equivalent when it comes to the relationship between results and policymaking. The scope of the OECD is larger than the field of education, originally grounded in the economic sector, and one OECD mission is to give its members recommendations in terms of policies to be implemented. IEA, as an association grounded in academic research in education, has no vocation to draft recommendations to its members but rather, to provide evidence on which each individual country can build adequate policies regarding its own context. The following table describes some of the organizational differences between the IEA and OECD.

Table 2: Organizational differences between IEA and OECD

ASPECT	IEA	OECD
Type of organization	Non-governmental association (with national members).	Governmental organization.
Method of carrying out ILSAs	Conducts studies cooperatively with partner organizations from its scientific network.	Initiates studies and tenders the studies for each cycle. Separates out different tasks (e.g., framework development, sampling, test platform).
Review process	All study publications are rigorously peer-reviewed.	Written and reviewed in-house and reviewed by board members of participating countries.
Participation	Open to all countries. No requirement to take part in studies.	The flagship PISA study was originally targeted to OECD member countries and is now open to all economies.
Decision making	Final content decided by the National Research Coordinators (NRC) of participating countries, each of which has an equal voice.	The organization structure reflects the original centrality of the OECD membership and voting is restricted to OECD members and select partner economies.
Fees	Each country makes an equal contribution to the international coordination costs.	Member countries pay different contributions based on their GDP. Non-member participants pay a flat fee.

COMMON METHODOLOGIES

Despite differences in organization and approach, ILSAs share a quite similar set of procedures and methods for implementation that have been developed and refined over the last 60 years, and have contributed to methodological advances in regional ILSAs and national assessment programs in many countries. Drawing from expertise from around the world, all major ILSAs mandate high-quality standards carefully defined to achieve each step in the ILSA process including: sampling rules, translation processes, field trial procedures, and psychometric modelling. These methods may appear quite complex to those unfamiliar with ILSAs, but, for transparency, the data are accompanied by detailed supporting technical reports and all ILSA data is made freely available online for researchers to download. Further, both IEA and OECD produce a large number of resources to help promote the secondary analyses of the data. These secondary analyses are possible and made rich by the presence of contextual data. A recent example is ICILS where contextual data helped to inform student digital literacy.

LIMITATIONS

There are some important limitations to consider in understanding and reporting ILSA results. Firstly, all ILSAs are cross-sectional studies. This means that they measure educational outcomes at one point in time for a specific population. Further, although some ILSAs such as TIMSS, PIRLS, and PISA report trends over time, they are not studies that follow individual students. This makes it challenging, if not unfeasible, to draw causal conclusions about student achievement and is a clear limit of the data with many policymakers seeking specific recommendation for change to help improve their educational system. Although some researchers seek to use advanced models to establish causal relationships, we personally maintain that such models are based on strong assumptions that are difficult, if not impossible, to achieve with ILSA data.

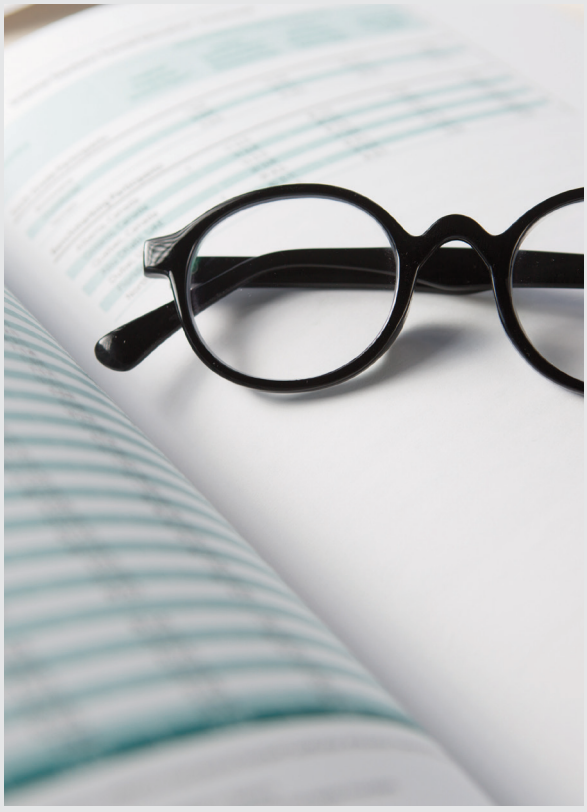
A second limitation of ILSAs is that the studies are not designed to measure individual students' achievement nor the results of individual schools but to reflect the educational results and relationship with background information within education systems. As such, the assessments are considered low stakes for schools, teachers, and students. Nevertheless, the low stakes aspect of the assessment has the advantage of lowering testing stress on students and schools. In fact, ILSAs require a fairly short testing time when compared to high stakes assessments and are only administered to a small representative portion of the population in the countries.

Finally, the content domains covered by ILSAs are not an exhaustive list of what is taught in schools. For example, IEA's TIMSS focuses on mathematics and science curriculum attainment at grades four and eight and, while critical subjects, strong attainment in these subjects alone cannot be considered to be a reliable measure of the overall health of an education system. Some have criticized ILSAs as exerting undue influence on education policies, with national approaches perceived as being replaced with a tendency to target national curricula toward better achievement in the subjects assessed by ILSAs. Consequently, care must be taken when considering results from ILSAs as a holistic quality measure of the education system, rather the focus should be on the results as important indicators of what students know and can do in specific subjects, and how students nationally compare to their peers internationally.

WHY ILSAs ARE IMPORTANT

Despite limitations, ILSAs are vital tools for education system improvement. The assessments are unique in that the information provided by the data can be used to monitor systems over time within a robust international framework, and they fall outside the governance of any one country thus being largely independent of any single political system. Further, the data is available to the public allowing researchers from around the world to explore the data and make use of it for their own research questions.

In many countries, research resulting from ILSAs has improved our understanding of how educational systems operate, informing policy decisions that are based on strong and reliable evidence. Various impact studies have shown that ILSA results have been used to support policymaking (e.g., Breakspear 2012; Schwippert & Lenkeit 2012; Wagemaker 2013) and as reported in the TIMSS and PIRLS Encyclopedias various educational improvements have been stimulated by evidence from ILSAs. When properly understood and analyzed, the data from ILSAs provide valuable opportunities to help inform policy decisions as well as research into education system improvement.



WHAT NEXT?

As ILSAs continuously modernize—most recently with a move toward computer-based assessments—new methodological opportunities and challenges await. This is why organizations conducting ILSAs are actively promoting research in the field of international assessment. For example, IEA sponsors academic journals, conferences, and thematic reports for outside researchers as well as employing its own research team to help further developments in the field of assessment. More generally, in its renewed strategy, IEA is placing a strong emphasis on research and innovation, including, for example, the promising topic of "process data"(i.e., digital traces left by students when passing an assessment).

Another focus of ILSA development is the exploration of larger and more complex dimensions, so-called "21st century skills." IEA has recently launched a curriculum study (21CS MAP) which aims to map these skills. This fundamental study will be based on what is taught in schools (the intended and implemented curricula) before developing a concrete assessment program.

CONCLUSION

In a growing interconnected world, we believe ILSAs can help us learn from others and, through comparison, better understand ourselves. IEA works diligently on assisting the policy and research community by training researchers and policymakers on how to interpret and analyze data, by writing and commissioning in-depth reports into the study findings, and by publishing quarterly briefs that are intended to be short digestible summaries of interesting study results. Such activities are undertaken in support of IEA's mission "to better understand education practices, processes, and policies in order to improve the quality of teaching and learning within and across systems of education." All IEA studies and reports follow the highest academic standards for social science research, including complete transparency about the testing process and robust peer review for studies and reports.

ABOUT THE AUTHORS

DR DIRK HASTEDT

Dr Dirk Hastedt is the Executive Director of IEA. He oversees IEA's operations, studies, and services, and drives IEA's overall strategic vision. Moreover, he develops and maintains strong relationships with member countries, researchers, policy makers, and other key players in the education sector. Dr Hastedt also serves as co-editor in chief of the IEA-ETS Research Institute (IERI) journal *Large-scale Assessments in Education*.

DR THIERRY ROCHER

Dr Thierry Rocher was elected to the IEA Chair position at the 59th General Assembly meeting in October 2018. Dr Rocher previously served as a Standing Committee member and General Assembly representative for France. He is the Head of the Office for Student Assessment (DEPP) at the Ministère de l'éducation nationale in France.

REFERENCES

Anderson, L.W., Ryan, D.W., & Shapiro, B.J. (1989). *The IEA classroom environment study*. Oxford, UK: Pergamon Press.

Breakspear, S. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*. OECD Education Working Papers No. 71. Paris, France: OECD Publishing

Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A., & Walker, D.A. (1962). *Educational achievements of thirteen-year-olds in twelve countries: Results of an international research project, 1959, 1961*. Hamburg, Germany: UNESCO Institute for Education. https://unesdoc.unesco.org/ark:/48223/pf0000131437.

Husén, T. (1967). *International study of achievement in mathematics: a comparison of twelve countries*. Vol. I–II. Stockholm, Sweden and New York, NY: Almqvist & Wiksell and John Wiley.

Husén, T. (1983). *An Incurable Academic: Memoirs of a Professor*. Oxford, UK: Pergamon Press.

IEA. (2018). *60 years of IEA 1958–2018*. Amsterdam, Netherlands: Author. https://www.iea.nl/index.php/publications/60-years-iea-1958-2018.

IEA. (2019). IEA studies [webpage]. https://www.iea.nl/studies/ieastudies

Lockheed, M. (2011). Reflections on IEA from the perspective of a World Bank official. In C. Papanastasiou, T. Plomp, & E. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories, Vol 2*. Nicosia, Cyprus: Cultural Center of the Kykkos Monastery.

Lockheed, M., Prokic-Bruer, T. & Shadrova, A. (2015). *The Experience of Middle-Income Countries Participating in PISA 2000-2015*. Paris, France: PISA/OECD Publishing.

OECD. (2019). PISA: Programme for International Student Assessment. http://www.oecd.org/pisa/

Schwippert, K., & Lenkeit, J. (Eds.). (2012). *Progress in reading literacy in national and international context: The impact of PIRLS 2006 in 12 countries*. Münster, Germany: Waxmann Verlag GmbH

Travers, K., & Westbury, I. (1989). *The IEA Study of Mathematics I: Analysis of the Mathematics Curricula*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Wagemaker, H. (2013). International large-scale assessments: From research to policy. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, p. 11. Florence, KY: Chapman and Hall CRC.

ADDITIONAL RESOURCES

Hastedt, D. (2016, November 7). The history and development of international assessments [Podcast]. FreshEd Podcast 49. http://www.freshedpodcast.com/dirkhastedt/.

Lockheed, M., & Wagemaker, H. (2013). International Large-Scale Assessments: Thermometers, whips or useful policy tools? *Research in Comparative and International Education, 8*(3), 296–306.

Singer, J.D., Braun, H.I., & Chudowsky, N. (2018). *International education assessments: Cautions, conundrums, and common sense*. Washington, DC: National Academy of Education.

IEA COMPASS

ABOUT IEA

The International Association for the Evaluation of Educational Achievement, known as IEA, is an independent, international consortium of national research institutions and governmental agencies, with headquarters in Amsterdam. Its primary purpose is to conduct large-scale comparative studies of educational achievement with the aim of gaining more in-depth understanding of the effects of policies and practices within and across systems of education.

ISSN: 2589-70396

Copies of this publication can be obtained from:

IEA Amsterdam
Keizersgracht 311
1016 EE Amsterdam
The Netherlands

By email: secretariat@iea.nl
Website: www.iea.nl

Follow us:
@iea_education
IEAResearchInEducation
IEA

Please cite this publication as:
Rocher, T., & Hastedt, D. (2020, September). *International large-scale assessments in education: a brief guide* IEA Compass: Briefs in Education No. 10. Amsterdam, The Netherlands: IEA