



UiO : **CEMO** – Centre for Educational Measurement
University of Oslo

WORKING PAPER

What Teachers Can't and Can Learn From International Assessments

THREE THINGS TO CONSIDER

- *These tests are not designed to give individual or school feedback*

ILSAs are not design for providing individual, class, or school feedback. And we are simply not good enough at assessing students for all purposes. ILSAs are designed to assess and compare national systems. As members of that system, participation can be useful and provide worthwhile information even if there is no specific feedback at the school, class, or school level.

- *There is no such thing as a value free assessment or score*

Each assessment measures different things for different purposes. It is important for any consumer of ILSAs to understand the purpose of the assessment, what is being assessed and the agenda of the organization. Although these notions aren't complicated, they are often overlooked in mainstream conversations around international assessments. As educators, we should be aware of the differences and help inform policy makers and the public of these differences.

- *Teachers should and can enter the debate!*

A number of critics take issue with several aspects of international educational surveys, including the comparability of diverse educational systems (e.g., Bracey, 2000; Ercikan, 2003), achievement rankings (e.g., Goldstein, 2013) and methods used to estimate achievement (Kreiner & Christensen, 2014; Rutkowski, 2011). Others such as Sjøberg (2012) have called into question the possibility of constructing an assessment that can be used across countries to assess “real life” situations. Sjøberg along with other scholars (e.g. Guardian, 2014) have been active in the popular press raising questions about what PISA and other ILSAs can tell us about our educational systems often warning of test misuses. What is often lacking, however, is the voice of teachers. We, as an educational community, need to hold each other accountable for using the data correctly. That said, simply because some people misuse the data does not mean we should entirely reject this important educational tool. As educators, we encourage you to become involved in the process and explore the large amount of resources that these assessments can provide!

AUTHORS:

David Rutkowski,
PhD

Leslie Rutkowski,
PhD

Centre for
Educational
Measurement,
University of Oslo

[www.uv.uio.no/
cemo/english](http://www.uv.uio.no/cemo/english)

Introduction

The measurement field has come a long way in assessing student achievement at the individual, classroom, national, and international level. Specifically, at the international level, assessments have grown tremendously over the past few decades. Today, two-thirds of all countries with populations greater than 30,000 have participated in one or more international or regional large-scale assessments¹. PISA alone has led to an increase in the number of assessed students internationally from 174,896 in 2000 to 510,000 in 2012. New versions of PISA such as PISA for Development, targeting developing economies, and PISA for Schools, focused on providing participating schools with internationally comparable PISA results, will only increase these numbers and bring international assessments to new contexts and audiences.

The highly aggregated scores resulting from these assessments are often the most cited in the media largely because they seem “clear” and “easy” to understand. In fact, the ease of assigning and interpreting a simple aggregated score, in and of itself, is a plausible explanation for the increased popularity of international large-scale assessments (ILSAs). For example, the OECD average mathematics score for the 2012 cycle of PISA was 494, with the following results in Nordic countries that participated: Finland 519, Iceland 493, Norway 489, and Sweden 478. We dig a bit deeper into score meanings subsequently, but first let’s think about the overall score and everything it appears to communicate. For example, the only Nordic country that did better than the OECD average was Finland. While Sweden’s performance was lowest overall in the Nordic region. With this in mind, imagine hypothetical headlines in Finland: “Finland outperforms all Nordic countries in math”; in Sweden: “Swedish outperformed by Nordic neighbors”; in Norway and Iceland: “Norwegian and Icelandic students perform below OECD peers.” With this one very simple and easily digestible number the media and policy makers can create narratives that are supported by a seemingly “objective” achievement measure, the metric of which is the same across all countries. A single summary math score of school-enrolled 15-year-olds wields incredible power! As the saying goes, however, the devil is in the details. And, in the case of international assessments, there is no shortage of details.

One obvious issue is that ILSA results (and most education tests) are being sold as objective and unbiased assessments of entire educational systems. And although a good share of the burden falls to the media, they aren’t the only stakeholders who oversell ILSA results. To that end, OECD leadership enthusiastically promotes PISA as an objective measure of an educational system’s performance. In a TED talk, Andreas Schleicher (2013), the PISA architect and Director of the Directorate for Education at the OECD claimed that:

PISA shows what is possible in education, it takes away excuses from those who are complacent, and it helps countries see themselves in the mirror of the educational results and educational opportunities delivered by the world’s leaders in education. (para. 9)

However, PISA technical documentation explicitly warns against making such inferences with PISA data as the test results do not support causal claims and do not specifically test educational systems. Nonetheless, OECD spokespeople seem to disregard their own warnings, as demonstrated in Schleicher’s statement. As such, it is difficult to solely blame the media for using the assessment in inappropriate ways.

¹Lockheed, Prokic-Breuer, & Shadrova, (2015)

In truth, ILSAs are far from perfect (as any measurement expert would certainly agree). Further, we as a testing community have yet to develop an assessment that serves all purposes or accurately measures all levels or aspects of an educational system. As such, when attempting to understand what teachers can learn from any national or international assessment it is imperative that we first examine what each assessment was intended for and thus what information any testing organization can reasonably provide to the public, media, and other stakeholders.

In what follows, we consider our readership to be practicing educators in Nordic countries. As such, we begin with a general description of the three largest international assessments: the Programme for International Student Assessment (PISA); the Trends in International Mathematics and Science Study (TIMSS); and the Progress in International Reading Literacy Study (PIRLS). The brief then continues with a more in-depth discussion around some questions that we often encounter when speaking with educators and school leadership internationally regarding how ILSAs can assist their own practice.

What is an ILSA?

To be as concise as possible we can define ILSAs as educational tests and questionnaires that measure educational achievement in select content domains for pre-specified samples of students enrolled in educational systems around the world. The samples of students are intended to represent the population to which they belong. And the tests that students take are intended to be an adequate representation of what students know and can do in the relevant content areas (e.g., math, science, and reading). The results of these assessments are used to compare educational systems (normally these systems are countries) and to inform policy, practice, and educational research both *nationally* and *internationally*.

Who administers the assessments?

As with any international endeavour the testing world is complicated and includes many actors and clients at all levels, from international to local. However, in general there are two organizations that are the clear leaders in the world of international assessment: the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD).

IEA

The IEA is an independent, international cooperative of national research institutions and government research agencies that aims to provide high-quality data capable of increasing policymakers' understanding of key factors influencing teaching and learning. Since its founding in 1958, the IEA has conducted more than 23 research studies on cross-national achievement. The organization is headquartered in the Netherlands with study centers in Germany and the United States. The IEA is grounded in the belief that the diversity of educational philosophies, models, and approaches that characterize the world's education systems constitute a natural laboratory in which countries can learn from one another.

OECD

The mission of the OECD is to promote policies that will improve the economic and social well-being of people around the world. Established in 1961, the OECD is headquartered in Paris, France, and has 34 member countries that represent the world's most industrialized nations. The OECD analyzes and compares data, sets international standards, and recommends policies to governments around the globe. The OECD began assessing educational systems in 2000 and, unlike the IEA, explicitly uses results from their assessments to make policy recommendations to participating countries.

What are the largest ILSAs?

ILSAs and international surveys in education have grown in terms of topics covered and participants (see appendix). Yet, three assessments currently stand out as the most popular measures of primary and secondary educational outcomes. The following is a brief description of each.

PISA

PISA is an assessment that tests content knowledge but is not limited to school-based curricula. Instead, PISA assesses applied knowledge and literacy and emphasizes assessment of the functional skills as defined by the OECD. The guiding question asked by PISA is "How well can students nearing the end of compulsory schooling apply their knowledge to real-life situations?" The subject areas tested on PISA are reading literacy, mathematics literacy, science literacy, financial literacy, and collaborative problem solving. PISA also includes measures of general competencies such as learning strategies.

PIRLS

PIRLS collects data to provide information on trends in reading literacy achievement. PIRLS includes an array of questions that investigate the experiences young children have at home and in school in learning to read. The assessment is offered to fourth-grade students because fourth-grade represents an important transition point in children's development as readers. In many countries, children are expected to have learned to read by fourth grade and are beginning to transition from learning to read to reading to learn. Because new countries participate in PIRLS each cycle, PIRLS also provides baseline data for new participating countries. In addition, PIRLS collects an array of information about the reading curriculum and instruction in each participating country.

TIMSS

TIMSS provides data about trends in mathematics and science achievement. The content assessed in TIMSS is based on an internationally agreed upon common curriculum in math and science. TIMSS collects detailed information not only about student achievement in math and science, but also about teacher preparation, resource availability, and the use of technology.

Table 1: General information on the big three international assessments:

	PISA	TIMSS	PIRLS
Subjects Assessed	Reading literacy, mathematics literacy, and science literacy, financial literacy and collaborative problem solving	Mathematics and science	Reading
Sample	15 year olds in school	8th grade and 4th grade students	4th grade students
Frequency	Every 3 years	Every 4 years	Every 5 years
Last Assessment	2015	2015	2016
Countries	65 countries and economies in 2012	66 countries and 14 sub-national entities in 2011	55 countries and 7 sub-national entities in 2011

Who participates in ILSAs?

Participation in PISA, TIMSS and PIRLS normally takes place at the national level although some sub-national regions and states also participate separately in order to compare themselves internationally. So, for example, there are a handful of U.S. states as well as national regions such as the Basque region of Spain that participate in one or more of the assessments. To participate, countries pay fees to the organization. Then each country assembles a team of people who will collect, compile, and send the data to the testing organizations for scoring. In general, the data is both collected and scored with satisfactory integrity. Of course, given the large variation of participating countries some problems do arise and generally these issues are noted in the tables of the international report and technical documentation (IEA studies) or in the technical documentation (OECD and IEA studies).

As shown in Table 1 there are a large number of countries that participate in these assessments and the numbers keep growing. For more information on the full list of participants for each study, please see the helpful links section of this brief for links to each assessment. For our purposes, Table 2 describes Nordic participation over the last two completed cycles of each assessment.

Table 2: Participation in the big three international assessments among Nordic countries:

	PISA		PIRLS		TIMSS			
	2009	2012	2006	2011	2007 4th	2007 4th	2011 4th	2011 8th
Denmark	x	x	x	x	x		x	
Finland	x	x		x			x	x
Norway	x	x	x	x	x	x	x	x
Sweden	x	x	x	x	x	x	x	x
Iceland	x	x	x					

Who actually takes the assessment and how are individual participants selected?

One important distinction between IEA's TIMSS and PIRLS studies and the OECD's PISA study is that IEA studies are grade based while the OECD studies are age based. Thus, PISA differs from the TIMSS and PIRLS approach in that it samples 15-year-olds enrolled in school (age based), regardless of how many years of schooling students have received. On the other hand, TIMSS and PIRLS measures a sample of students at the relevant grade level (grade based), regardless of the student's age. Although this brief explanation is somewhat oversimplified, the following provides a general overview of the two sampling techniques.

To obtain a representative sample of the population both the IEA and OECD receive a list of all schools in each country. Although sample sizes vary depending on specific country needs, in general about 150 schools and approximately 4,500 students are selected. Each country (or other sub-national educational system) provides a list of eligible schools from which each organization randomly selects schools to participate. An important distinction between the OECD and IEA is how students are selected within each school. For the PISA study the schools provide a list of all 15 year olds in the school (regardless of grade level). Thirty five students are then selected at random from this list (except in the case that there are fewer than 35 eligible students in a school. Then all 15-year-olds are selected). For TIMSS and PIRLS, which use a grade based sample, generally one or two entire classrooms at either the 4th or 8th year of schooling are selected.

Each approach has its own advantages. For example, a grade-based sample ensures that all students who take the assessment have had a similar amount of schooling, while an age-based sample allows inferences on the skills that are attained through the first 15 years of life, regardless of whether the measured skills are attained in or outside school. In some countries, such as Norway, where most students are the same age in each grade, there isn't much of a difference between the two test populations (e.g., between TIMSS and PISA). But, in other countries, such as the United States, where 15-year-olds can be in as many as five different grades, the differences between IEA and OECD studies can be meaningful. Even if most students in your country are from a single grade it is important to keep in mind that the same does not hold for all countries that you compare yourself with.

Due to sampling alone, assessments (e.g., TIMSS and PISA) should not be viewed as interchangeable. In other words, testing organizations are purposely assessing different populations in most countries (PISA = 15-year-olds in possibly different grades or TIMSS = 8th grade students at possibly different ages). Not surprisingly, both the IEA and OECD advocate for their own approach. For example, the OECD claims that an age-based sampling procedure is used to ensure valid international comparisons of educational performance that can happen within and outside the classroom whereas the IEA claims that a grade based sample ensures that all students had the same opportunity to learn the information being tested. There are valid arguments for and against each approach.² Important for the general ILSA consumer is that each assessment measures a *representative* sample of the population of interest. Nevertheless, the defined populations are fundamentally different across the big three studies. It makes is all the more important, then, to *only* interpret test results for the population from which the sample was drawn. In the case of PISA interpretations are relevant for 15-year-olds enrolled in school. In contrast, interpretations should be limited to students in either the 4th or 8th year of schooling for IEA studies.

² For more information please see: (McGaw, 2008 and Wagemaker, 2008)

What can the results tell us?

As these assessments are designed to measure national (or sub-national) systems, the majority of information that can be gleaned from these assessments is admittedly most relevant at the national (or sub-national) level, rather than at the local level. Further, because of the assessment design and methods used to estimate achievement, achievement should only be reported at the national/sub-national or aggregate subgroup level (e.g., boys' achievement compared to girls' achievement). It is not appropriate to report achievement at the student, class, or school level.

Perhaps one of the most unfortunate aspects of international testing programs has been an over-reliance by media and policy makers on the ranking of aggregated country/system achievement scores, or what are typically called *league tables*. In general, the aggregated national score actually tell us very little about our educational systems, particularly with respect to the context and correlates of educational achievement. Taking PISA as an example (we could have also chosen TIMSS or PIRLS), the OECD average mathematics achievement score was 494 points. Further, Finland scored 519, Norway scored 489 and Sweden scored 478. With these scores in mind, the following statement most accurately represents the results for Norway.

On the 2012 PISA assessment 15-year-olds that are enrolled in school in Norway scored an average of 478 points on the mathematics portion of the assessment, which focuses on what the OECD believes students should know to participate in the global economy.

This, of course, is much different than many of the headlines and discussions we see in popular media, political discussions, and OECD reports. For example, one newspaper reported that “Norway left behind Denmark and Finland in new PISA survey.”³ Similarly, the OECD reported that based on PISA results “Sweden should urgently reform its school system to improve quality and equity.”⁴ These overreaching statements regarding tend to overshadow the useful information that PISA and other ILSAs *can* provide. In what follows we attempt to outline some areas where we think ILSAs provide an important voice to international educational conversations.

ILSAs as a means for general comparison

The first and perhaps most common use of ILSA results are as a way for national policy makers to see how their country is doing in comparison to others. When comparisons are appropriate this can be helpful. For example, Norway, Denmark, Iceland, Sweden, and Finland share a great deal in common, yet they have implemented different ways of educating their citizens. ILSAs allow an accessible way to better understand how one's system “stacks up” against another, in light of country-wise differences. Although “seeing how you are doing” can be interesting and is often the most cited aspect of ILSAs, it is unfortunately also the most superficial and normally tells us very little about our educational system. More troubling is that comparisons can be dangerous and even misleading. For example, comparing the educational system of Iceland with a population of approximately 320,000 to the U.S. with a population of approximately 320,000,000 may not be relevant for either system. Yet they appear together in the league tables. We

³ <http://www.tnp.no/norway/panorama/4151-norway-left-behind-denmark-and-finland-in-new-pisa-survey>

⁴ <http://www.oecd.org/sweden/sweden-should-urgently-reform-its-school-system-to-improve-quality-and-equity.htm>

contend that although the “horse race” often fills the headlines around the world, it is the least relevant aspect of ILSAs and should be discounted in any consideration of ILSA results.

ILSAs as a way to learn from others

ILSAs also allow us to learn from others. These assessments provide a great deal of information on how other educational systems operate. Moreover, the assessments enable national policy makers and researchers to raise and then explore important policy and research questions. For example, why do immigrant students in country *X* outperform immigrant students in country *Y*? Or, why do girls in country *Z* have a higher self-perception of their math proficiency when compared to girls in country *V*? Exploring what national systems are doing to improve immigrant students’ achievement or girls’ self-perception of math can help policy makers and researchers design appropriate policies for their national context. Learning from others does not suggest a simple adoption of other’s policies. Rather, looking outward offers an opportunity to examine what others are doing differently and exploring whether those differences can be adapted to one’s local context.

Learning from others is an area where teachers can also find useful information from ILSAs. As noted before, these assessments are designed to measure at the national (or sub-national) level. However, teachers can review the various reports that result from ILSAs and relate them to their own context. For example, if the results from the latest ILSA cycle suggest that girls in a given teacher’s country have lower self-perception of math proficiency than other countries, a useful exercise could be to examine if the finding holds in his or her own classroom. Finding consistent evidence of low self-perception among girls certainly makes it reasonable to further investigate the types of practices in other countries to foster a better sense of self-perception. Additionally, there is a clear opportunity for teachers to engage with policy makers and to request assistance. ILSAs are not a one way-street, solely designed to provide evidence for the policy and research community in order to evaluate educators. We believe that educators can also use the information provided by ILSAs as empowerment tools, using ILSA findings to demonstrate strengths and areas in need of more support.

ILSAs as a way to understand ourselves

Most educators teach where they were educated. This provides a great deal of advantage, particularly with respect to connection to the local culture. Yet, being raised in a particular educational system can lead many to look inward rather than outward. ILSAs can, and have in many countries, resulted in conversations around existing educational structures. In other words, ILSAs have increased dialog about what we do and do not do well. Of equal importance, ILSAs have encouraged public debate as to their usefulness for making cross-country comparisons. In fact, we contend that critical debates about ILSAs have made many reengage around issues of education and better understand their own systems. Looking outward to other systems is often as important as looking inward. Comparison helps us better understand ourselves and allows us to continually reevaluate our own way of doing things focusing both on what seems to be working and what can be improved.

Beware of uncertainty!

It is important to understand that there is both uncertainty and error associated with any student assessment. And given the scale of the studies issues around uncertainty and error are only amplified with international assessments. Although a detailed description of all error sources is not possible in a short brief, we would like to point out that there are errors that you can and cannot account for with statistical models. For example, testing organizations draw a sample of students to take the test (rather than testing the entire population), which gives rise to inherent error national score estimates. *Sampling error* is usually accounted for by some margin of error around the estimated achievement scores. In the case of Norway's mathematics PISA score, this error range is greater than 6 points on the PISA scale. Although the single overall estimate of achievement places Norway below the OECD average, once we account for sampling error, we see that Norway is no different than the average of OECD countries in terms of their achievement. Although we can explicitly account for sampling error, other sources of error are more difficult to account for and are, reasonably, cause for concern. Here, we focus on two types of error that are less accounted for in typical analyses: Cultural differences in item characteristics and background questionnaire consistency.

Cultural bias

In general, the methods used to estimate achievement rely on an important assumption: students with similar knowledge and skills find the test questions equally difficult across countries. In other words, an item should be equally difficult for equally-proficient students in Norway, Kazakhstan, and Shanghai. However, research has shown that this assumption does not hold and can actually affect the ranking of middle-performing educational systems, including most Nordic countries⁵. In other words, violating the assumption of equivalent item difficulty across populations has a meaningful impact on the scores we use to rank countries. Because ILSAs scores are subject to this error and rankings are consequently susceptible, it is important to consider achievement for groups of countries (e.g. high, medium, low performing), rather than focusing on individual rankings. In other words, it is rarely appropriate to say that country *X* rose or fell three places on the international league tables between cycles.

Background questionnaire consistency

The background questionnaires from ILSAs serve a number of purposes, with the most obvious being to help us better contextualize achievement. For example, we can investigate differences in achievement between boys and girls, students that have high self-efficacy in the subject, and students with high and low socio-economic status. Some assessments even include home questionnaires, such as PIRLS, where both students and parents filled out questionnaires. Further, some questions are common between students and their parents, making a comparison between their respective responses possible. As an example, parents and students were asked to indicate the approximate number of books they have in their home – a variable that is commonly used as part of a socioeconomic/sociocultural measure in international studies – on a scale from one (none to 11 books) to five (more than 200 books). We would expect that the responses from the parents and children would be highly correlated, if not precisely identical; however, in many

⁵ (Rutkowski, Rutkowski, & Zhou, 2016)

educational systems this is not the case.⁶ Norway, for example, has a raw correlation of .44, suggesting that less than half of the parents and students agree on this important and widely used variable. As such both policy makers and researchers should use considerable caution when including either the parent or student response in an analysis, particularly as a predictor in achievement.

A final important fact to keep in mind is that due to the complexity of the study design, scores cannot be compared at the individual participant or even school level. In other words, it is never appropriate to compare achievement scores from one student to another student or one school to another school using these assessments. A primary reason this is not possible is because of the scope of the content areas and as a result, no student takes the entire test. Rather, each student takes part of the test and highly sophisticated methods are used to leverage small pieces of information across the sample to derive a reasonable estimate of achievement for the population. This does not indicate that results are not valid or reliable at aggregated levels. Nonetheless, it leads to high levels of uncertainty at the student, class, and school levels. For teachers, this is especially important to keep in mind if your school is selected to participate in one of the ILSAs. Unfortunately, there is generally no information to offer students, teachers, or their schools on overall performance at such disaggregated levels. We offer one exception to this statement in the appendix, where we suggest the possibility of using background questionnaires as a means for local measurement and possible comparison at the national or international level.

Conclusion

ILSAs play an important role in our understanding of educational systems both nationally and around the world. As we have discussed in this brief, ILSAs allow us to learn from others, and better understand ourselves. As such, we see clear value in the results; however, we also believe that restraint in interpreting the results should be exercised by all parties especially when it comes to making policy recommendations solely from the results. With both the promise and pit-falls of ILSAs in mind, we contend that teachers need to play an active role in helping the community and policy makers contextualize the results. Although, complicated analysis of ILSAs certainly exists in most policy and public circles, ILSAs are used in very basic ways. Practitioners are an important voice in the contextualization of ILSAs and can also play an important role in bringing to light when the assessments are being misused.

Helpful links:

- | | |
|--|--|
| <ul style="list-style-type: none"> • OECD's work in education:
http://www.oecd.org/education/ • PISA:
https://www.oecd.org/pisa/ | <ul style="list-style-type: none"> • IEA
http://www.iea.nl/ • TIMSS and PIRLS
http://timssandpirls.bc.edu/ |
|--|--|



UiO • **Faculty of Educational Sciences**
University of Oslo

⁶ (Rutkowski & Rutkowski, 2010)

Appendix

Other Assessments and Organizations

Both the IEA and OECD have administered other surveys and assessments in education. For example, the OECD has also administered the Teaching and Learning International Survey (TALIS), which surveys lower-secondary teachers and their principals about their working conditions and learning environments. TALIS is administered every five years, beginning with the first cycle in 2008. Additionally, the OECD administered the Survey of Adult Skills (PIAAC), with the first cycle spanning 2008 – 2016. PIAAC measures three content domains: literacy, numeracy, and problem solving in technology-rich environments. PIAAC is primarily computer-based and is administered to a representative sample of adults, aged 16 to 65 in 33 countries. The second PIAAC cycle is slated for 2018-2023. IEA additionally conducts the International Civics and Citizenship Study (ICCS), with two cycles (2009 and 2016). ICCS is an achievement test and survey in 25 educational systems of eighth grade students on their knowledge and understanding of civics and citizenship as well as their beliefs, attitudes, and behaviors. The IEA also administers the International Computer and Information Literacy Study (ICILS), the first round of which was conducted in 2013 with 19 participating systems. ICILS 2013 asked how well students are prepared for life in the information age. Although not comprehensive, other IEA studies include the Teacher Education and Development Study in Mathematics (TEDS-M 2007-2008) and the Second Information Technology Study (SITES 2006).

Other international organizations such as the United Nations Educational, Scientific, and Cultural Organization (UNESCO) have also developed assessments (e.g. Literacy Assessment and Monitoring Programme) and have provided funding for regionally based assessments (e.g. the Latin American Association for the Assessment of the Quality of Education, and the Southern and Eastern African Consortium for Monitoring Educational Quality).

Making Use of Background Questionnaires

Each of the assessments that we cover in this brief include a number of questionnaires that are filled out by students, teachers, principals, and in the case of PIRLS and PISA, parents. Finally, given that the IEA studies are curriculum-based, a representative from each participating system completes a curriculum questionnaire. Across all three surveys students generally provide information about themselves, their home environment, their attitude and experiences around learning. Teachers and principals provide information about the school and teaching and parents provide information about themselves, their home, students learning, and their connection with the school. Countries also include questions that are of specific interest to their own context. So, for example, in TIMSS 2011 some country options that were used to better understand home wealth included a freezer in Tunisia, family car in Ukraine, piano in Sweden, and a private house maid in Qatar.

As mentioned before, test scores and tables of rankings are what often make headlines; however, the information collected by these questionnaires are extremely valuable in their own right. For example, because we have a representative sample of students in each study the questionnaires provide strong evidence on issues such as students' feelings toward the subjects being assessed, rates of bullying, attitude towards learning and school to name a few. Interested educators could also examine how students, other teachers, and parents view educational systems both nationally and internationally. In fact, the questionnaires are highly underutilized by both the policy and research community. Given that the questionnaires are publicly available, an interested teacher could administer some of the same questions to their class and compare those results against other students in their country, as well as from

students around the world. Similarly, teachers and principals could take the teacher and principal survey and see how they and their school compares nationally or internationally. It is important to note that these sorts of approaches should be used for informational purposes and generally should not be used for decision making, particularly high-stakes.

References

- Bracey, G. W. (2000). The TIMSS “Final Year” Study and Report: A Critique. *Educational Researcher*, 29(4), 4–10. <http://doi.org/10.2307/1176452>
- Ercikan, K. (2003). Are the English and French versions’ of the Third International Mathematics and Science Study administered in Canada comparable? Effects of adaptations. *International Journal of Educational Policy, Research and Practice*, 4, 55–75.
- Goldstein, H. (2013, December 13). Educational policy: What can PISA tell us? [Web log post]. Retrieved from <http://hgeduc.blogspot.de/>
- Guardian. (2014, May 6). OECD and Pisa tests are damaging education worldwide. *The Guardian*. Retrieved from <http://www.theguardian.com>
- Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, 79(2), 210–231. <http://doi.org/10.1007/s11336-013-9347-z>
- Lockheed, M., Prokic-Breuer, T., & Shadrova, A. (2015). *The Experience of Middle-Income Countries Participating in PISA 2000-2015*. Washington DC: World Bank Publications. Retrieved from <http://dx.doi.org/10.1787/9789264246195-en>
- McGaw, B. (2008). The role of the OECD in international comparative studies of achievement. *Assessment in Education: Principles, Policy & Practice*, 15(3), 223–243. <http://doi.org/10.1080/09695940802417384>
- Rutkowski, L. (2011). The impact of missing background data on subpopulation estimation. *Journal of Educational Measurement*, 48(3), 293–312. <http://doi.org/10.1111/j.1745-3984.2011.00144.x>
- Rutkowski, L., & Rutkowski, D. (2010). Getting it “better”: The importance of improving background questionnaires in International Large-Scale Assessment. *Journal of Curriculum Studies*, 42(3), 411–430. <http://doi.org/10.1080/00220272.2010.487546>
- Schleicher, A. (2013, December 10). Are the Chinese cheating in PISA or are we cheating ourselves? [Web log post]. Retrieved from <http://oecdeducationtoday.blogspot.de/2013/12/are-chinese-cheating-in-pisa-or-are-we.html>
- Sjøberg, S. (2012). PISA: Politics, fundamental problems and intriguing results. *La Revue, Recherches En Education*, 14, 1–21.
- Wagemaker, H. (2008). Choices and trade-offs: reply to McGaw. *Assessment in Education: Principles, Policy & Practice*, 15(3), 267–278.