

# **An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies**

Gustavo E. Fischman and Amelia Marcetti Topper, with the collaboration of Iveta Silova, Jessica L. Holloway, and Janna Goebel

This project received financial support from *Open Society Foundation* to complete this report. We also want to acknowledge Dr. David C. Berliner, Dr. Pasi Sahlberg and Dr. Stephanie McBride-Schreiner for their thoughtful comments, which helped us refine our arguments and analysis.

All mistakes are our own.

The views expressed herein are those of the authors and do not necessarily reflect the views of CASGE.

**An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies** by Gustavo E. Fischman & Amelia Marcetti Topper with the collaboration of Iveta Silova, Jessica L. Holloway, & Janna Goebel

<http://dx.doi.org/10.14507/casge2.2017>

CASGE working papers are circulated for discussion and comment purposes. They have not been peer-reviewed.

copyright notice



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/). Readers are free to copy, display, and distribute content that appear in *CASGE* as long as the work is attributed to the author(s) and *CASGE*, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. All other uses must be approved by the author(s) or *CASGE*. By submitting a manuscript, authors agree to transfer without charge the following rights to *CASGE* upon acceptance of the manuscript: first worldwide serial publication rights and the right for *CASGE* to grant permissions as its editors judge appropriate for the redistribution of the content, its abstract, and metadata associated with the working paper in professional indexing and reference services. Any revenues from such redistribution are used solely to support the continued publication and distribution of working papers.



## Executive Summary

The use of cross-national comparative assessments to measure educational access and student learning is a relatively recent global phenomenon (Kirsch, Lennon, von Davier, Gonzalez, & Yamamoto, 2013; United Nations Educational, Cultural and Scientific Organization [UNESCO], 2016). Since the UNESCO Institute for Statistics' (UIS) 12-country pilot assessment in 1958, international large-scale assessments (ILSAs) have grown in number, student populations included, content covered, and national participation (Kirsch et al., 2013). Some of the most well-known and influential ILSAs include the Organisation for Economic Cooperation and Development's (OECD) *Program for International Student Assessment* (PISA), and the International Association for the Evaluation of Educational Achievement's (IEA) *Progress in International Reading Literacy Study* (PIRLS) and *Trends in International Mathematics and Science Study* (TIMSS).

To obtain a better, comparative understanding of student learning, stakeholder organizations - such as UNESCO - have started to explore a more ambitious goal of developing global learning metrics (GLMs) in relationship to existing ILSAs. The use of GLMs – intended to measure “access *plus* learning” (Learning Metrics Task Force [LMTF], 2013, p. 10, emphasis in the original; see also Robinson, 2011) - represents a strategic shift away from ILSAs traditional focus on documenting access to education using school enrollment data. For example, the Learning Metrics Task Force - a joint effort of UIS and the Brookings Institution's Center for Universal Education created for the purpose of developing GLMs - has recommended measuring student learning across seven domains ranging from typical school subjects (e.g., numeracy and mathematics, science and technology) to overall well-being and civic participation (e.g., arts and culture, social and emotional well-being, physical well-being).

Despite the intense and often controversial nature of the conversations around the purpose of ILSAs and, relatedly, the development of GLMs, there is little research examining whether and to what extent ministries of education, national policymakers, and other national political and social actors value these types of international measures of student attainment. Nor is there a robust set of literature exploring to what extent, or how, these educational stakeholders have integrated ILSAs or GLMs into their work at the national level. To answer these questions, we conducted an exploratory review of the research and policy literature on ILSAs/GLMs and administered two surveys to ILSA/GLM experts, policymakers, and educators to understand whether, to what extent, and how have ILSAs/GLMs influenced education policymaking at the national level.

Our review of the literature confirms that ILSAs/GLMs, with their multiple and ambiguous uses, increasingly function as *solutions in search for the right problem* (see Lewis, 2016) – that is, they appear to be used by governments as tools to legitimate existing or new educational reforms. Our survey results pointed to a growing perception among researchers, policymakers, educators, and other stakeholders that ILSAs/GLMs are having an effect on national educational policies, with over one-third (38%) of all survey respondents stating that ILSAs/GLMs were generally misused in national policy contexts. However, while the ILSA/GLM literature indicates that ILSAs/GLMs are having some influence (Breakspear, 2012), there is little evidence that any positive or negative *causal* relationship exists between ILSA/GLM participation and the implementation of education reforms (Baird et al., 2016; Rutkowski & Delandshere, 2016). Consistent with these findings, our survey respondents were divided over whether, based on their professional and personal experiences, ILSAs/GLMs were positive contributors or hindrances to national education reform efforts.

Perhaps the most significant change associated with the use of ILSAs/GLMs in the literature we reviewed is the way in which new conditions for educational comparison at the national, regional and global level have been made possible. ILSAs/GLMs provide governments and education stakeholders with new modes of educational comparison that purportedly allow for the assessment of educational achievement both within (e.g., cities, states, regions) and between countries (Streitholt, Rosén, & Bos, 2013). Although not explicitly discussed in the literature, it was also evident that participating in ILSAs/GLMs developed a new educational market that required specialized personnel to implement, monitor, and assess these rather complex survey efforts and the resulting data, thereby helping to drive the development of a new class of consultants, monitors, evaluators, and researchers at the national, regional and global levels (Gove & Cvelich, 2014; Greger, 2012; Lewis & Lingard, 2015).

## Table of Contents

Introduction.....	7
Methodology.....	9
Review of the ILSA/GLM Literature .....	9
Survey of ILSA/GLM Stakeholders .....	11
Research Findings .....	12
Review of the Literature: ILSAs/GLMs as Tools of Legitimation .....	12
Stakeholder Surveys: ILSAs/GLMs Perceived as Instrumental to Policy Reform .....	16
Discussion.....	23
Acknowledgements .....	26
Abbreviations and Acronyms.....	26
Glossary.....	27
References.....	28
Appendix.....	34

## Introduction

The use of cross-national comparative assessments to measure educational access and student learning is a relatively recent global phenomenon (Kirsch, Lennon, von Davier, Gonzalez, & Yamamoto, 2013; United Nations Educational, Cultural and Scientific Organization [UNESCO], 2016). Since the UNESCO Institute for Education's (UIS) 12-country pilot assessment in 1958, international large-scale assessments (ILSAs) have grown in number, student populations included, content covered, and national participation (Kirsch et al., 2013). ILSAs appear to be playing an increasingly vital role in driving, supporting, and shaping educational policy and reform efforts promoted by governments at the national and, in several countries, sub-national levels (Breakspear, 2012; Hopkins, Pennock, Ritzen, Ahtaridou, & Zimmer, 2008).

Some of the most well-known and influential ILSAs include the Organization for Economic Cooperation and Development's (OECD) *Program for International Student Assessment (PISA)* and *World Education Indicators' Survey of Primary Schools (WEI-SPS)*, and the International Association for the Evaluation of Educational Achievement's (IEA) *Progress in International Reading Literacy Study (PIRLS)*, *Trends in International Mathematics and Science Study (TIMSS)*, and *International Civic and Citizenship Education Study (ICCS)*. There are also a number of ILSAs that are region specific: *Latin American Laboratory for Assessment of Quality in Education (LLECE)*, *Southern and Eastern Africa Consortium for Monitoring Educational Quality (SAQMEC)*, *Program for the Analysis of Educational Systems of the Conference of Ministers of Education of French-Speaking Countries (PASEC)*, India's *Annual Status of Education Report (ASER)* survey, and Research Triangle Institute's open-source *Early Grade Reading Assessment (EGRA)* and *Early Grade Mathematics Assessment (EGMA)* surveys for low-income countries.<sup>1</sup> Taken together, these ILSAs have been administered on variable schedules in almost every country around the globe,<sup>2</sup> with many national governments participating in more than one ILSA at a time<sup>3</sup> and more participating in ILSAs generally over the past two decades.<sup>4 5</sup>

To obtain a more comprehensive and comparative understanding of student learning, organizations that administer ILSAs and other stakeholder organizations have started to explore a more ambitious goal of developing global learning metrics (GLMs) in relationship to existing ILSAs. GLMs provide standardized learning outcomes, typically in the areas of literacy and numeracy, with which countries can benchmark their progress over time. Some of the organizations that have been actively engaged in GLM development include: UNESCO's UIS division and SAQMEC network; IEA; OECD; the World Bank; CONFEMEN; the ASER Center<sup>6</sup>; and the Global Partnership for Education (GPE).<sup>7</sup>

---

<sup>1</sup> EGRA/EGMA were developed with support from the U.S. Agency for International Development (USAID) and the World Bank.

<sup>2</sup> PISA - every 3 years; PIRLS - every 5 years; TIMSS - every 4 years; LLECE - no set frequency; ASER - annual (household based); EGRA/EGMA - no set frequency/varies by country

<sup>3</sup> For example, the United States participates in PISA, PIRLS, and TIMSS, while the Philippines has only participated in the 1999 and 2003 TIMSS administrations (IEA, 2017; OECD, n.d.).

<sup>4</sup> Forty-three countries participated in PISA 2000, for instance, compared with 76 countries in PISA 2015 (OECD, n.d.).

<sup>5</sup> It should be noted that IEA studies have been primarily used for research purposes by national experts, whereas the OECD has an aggressive role as a policy advisor and interpreter of its own findings.

<sup>6</sup> As part of ASER's administration of the *Annual Status of Education Report (ASER)* survey in India.

OECD, through its sponsorship of PISA, is perhaps the most prominent organization participating in GLM development today, as PISA is currently administered in 72 countries across the globe (OECD, n.d.).

In general, existing GLM efforts tend to focus on students meeting certain agreed-to milestones, such as reading proficiency by grade 3. Although recommendations vary across initiatives and working groups, there are efforts to forge partnerships across stakeholders to develop a more standardized list of GLMs (e.g., Winthrop, Anderson, & Cruzalegui, 2015). For example, the Learning Metrics Task Force (LMTF) - a joint effort of UIS and the Brookings Institution's Center for Universal Education created for this express purpose - has recommended measuring student learning across seven domains ranging from typical school subjects (e.g., numeracy and mathematics, science and technology) to overall well-being and civic participation (e.g., arts and culture, social and emotional well-being, physical well being).

The use of GLMs – described as measures of “access *plus* learning” (LMTF, 2013, p. 10, emphasis in the original; see also Perlman Robinson, 2011) – represents a strategic shift away from national and international organizations' traditional focus on documenting school enrollment. This shift in focus was cemented following the approval of the United Nation's (UN) Education 2030 Framework for Action in May 2015 and the launching of their aspirational Sustainable Development Goals in September 2015. These included achieving quality education, defined as ensuring “inclusive and equitable quality education” as well as the promotion of “lifelong learning opportunities for all” (UN, 2016).

While the need for universal access to education is widely supported, how learning should be measured and the need to establish GLMs are hotly debated topics.<sup>8</sup> Those in favor of GLMs see them as a way to improve children's educational development (LMTF, 2013; Schleicher, 2009; Winthrop & Simons, 2013), while GLM critics say these outcomes are often too narrow to provide a useful understanding of a nation's educational progress (Ary, Jacobs, Sorensen, & Razavich, 2006; Education International, 2011). Moreover, others voice a broader concern about systems of accountability and the commercialization of public education (Gorur, 2016; Verger, 2008; Verger et al., 2012). Despite the intense and often controversial nature of the conversations around the purpose of ILSAs and, relatedly, the development of GLMs, there is little research examining to what extent ministries of education, national policymakers, and other national political and social actors value these types of international measures of student attainment. Nor is there an established set of literature exploring to what extent, or how, these educational stakeholders integrate ILSAs or GLMs into their work at the national level.<sup>9</sup>

To answer these questions, we conducted an exploratory review of the research and policy literature on ILSAs/GLMs and conducted two surveys of ILSA/GLM stakeholders to understand how these assessments and metrics have influenced primary and secondary education policy at the national level. More specifically, we sought to understand:

---

<sup>7</sup> Through GPE's work with UIS and GPE partner countries.

<sup>8</sup> To date there are no consensual or accepted standards of proficiency and no agreed tests to ensure that participating countries' metrics and measures are comparable to each other and over time. The SDG indicator framework recommended three specific points of measure to track global progress on learning outcomes: “Percentage of children/young people in i) grades 2/3; ii) at the end of primary; and iii) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics.”

<sup>9</sup> It is important to acknowledge that our review didn't produce much literature focused on resistance to ILSAs.



- 1) To what extent have ILSAs/GLMs influenced national primary/secondary education policymaking?
- 2) What changes in national primary/secondary education policies and practices have been made in countries where there is evidence of ILSA/GLM influence?

The ultimate purpose of our review is to provide the international community with a better understanding of the ways in which - if at all - ILSAs/GLMs influence the work of educational stakeholders at the national level (e.g., “PISA Shock” and the like). Our work also serves as a basis for future research on the effect of ILSAs/GLMs on educational decision-making and reforms, including a more comprehensive systematic review of the global literature.

## Methodology

As described above, we conducted an exploratory review of the ILSA/GLM literature and administered two surveys on the perceptions of ILSA/GLM use among researchers, policymakers, union representatives, and educators. The following sections detail our methodological approach to collecting and analyzing these data sources.

### Review of the ILSA/GLM Literature

**Data Collection.** We collected relevant materials published between 2000 and 2016 in English, Spanish, and Portuguese. Given that the majority of primary/secondary-focused ILSAs began after 2000 (see Glossary), our review captured a substantial portion of the ILSA/GLM literature available to date.

We searched five well-known electronic databases (Google Scholar, EBSCO-Education Full Text, ERIC, JSTOR, and SciELO) using a combination of key search terms (e.g., international large scale assessments, global learning metrics, global educational reform model, and policy impact, policy use, policy influence). We were most interested in obtaining official reports and policy briefs produced by national and international organizations (e.g., OECD, UNESCO), and scholarly articles published in peer-reviewed journals that offered analysis based on empirical research.<sup>10</sup> In other words, we did not include commentaries and opinion-based perspectives, even if they were published in peer-reviewed journals. We also included relevant book chapters, but not entire books, given the limited timeframe of our study. While we scanned relevant newspaper articles, our goal was to focus on empirical publications instead of relying on second-hand reporting, and the newspaper articles found did not directly document the influence of ILSAs/GLMs on educational policy.

While there are several ILSAs and GLM efforts that focus specifically on tertiary schooling and adult learners (e.g., IEA's *Teacher Education and Development Study in Mathematics* [TEDS-M], OECD's *Assessment of Higher Education Learning Outcomes* [AHELO] and Programme for the International

---

<sup>10</sup> The list of journals reviewed include, but are not limited to: *American Journal of Evaluation*, *British Journal of Sociology of Education*, *Cadernos de Educação*, *Comparative Education*, *Comparative Education Review*, *Compare*, *Current Issues in Comparative Education*, *Discourse*, *Educação e Pesquisa*, *Education Evaluation and Policy Analysis*, *Educational Research and Evaluation*, *Ensaio: Avaliação e Políticas Públicas em Educação*, *European Educational Research Journal*, *European Education*, *International Journal of Education Development*, *Journal of Education Policy*, *Research in Comparative and International Education*, *Revista Brasileira de Educação*, *Revista de Educación (España)*; *Revista Iberoamericana de Educación*, *REICE*, *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, *Revista Española de Educación Comparada*, *Revista de Investigación en Educación*.

Assessment of Adult Competencies [PIAAC], UNESCO's *Global Report on Adult Learning and Education* [GRALE]), we limited this initial exploratory study to ILSAs that target primary and secondary school student populations. The next stage of our analysis will involve expanding our research to include all ILSAs regardless of student population.

**Literature Sample.** From our initial scan, we identified a total of 209 publications. Seventy-one of these publications specifically addressed how ILSAs/GLMs had influenced education policies at the national, or sub-national, level. Interestingly, almost all of these publications were published in English; with very few articles published in Spanish or Portuguese that specifically addressed the use of ILSAs/GLMs at the national level or the effect of ILSAs/GLMs on educational policy based on empirical research. A few articles in Spanish and/or Portuguese were translations of articles published in English.

Two-thirds (64%) of the final set of 71 publications were published in peer-review journals, with the most frequently cited journals being *Journal of Education Policy* (6), *Comparative Education* (5), *European Educational Research Journal* (5), *European Education* (4), and *Research in Comparative and International Education* (4). The majority of the articles (76%) were published after 2009, which speaks to the rising interest in documenting the policy influence of ILSAs/GLMs on national and subnational educational reform efforts. Our final set publications include policy profiles of 46 countries, as well as general commentary on European Union (EU) and OECD countries. Consistent with Lockheed (2015), high-income countries predominantly located in the Global North were the most studied cases in the literature, with Germany (10) being the most frequently profiled nation, followed by England/United Kingdom (8), France (7), the United States (7), and Canada (6).

The majority (90%) of the publications we reviewed focused on national, or in several cases sub-national (Engel & Frizzell, 2015), responses to PISA. There were four studies that looked at the effect of TIMSS on educational reforms, three for EGRA, and two for PIRLS and SACMEQ. We did not find any publications examining the policy influence of other major ILSAs (i.e., ICCS, LAMP, LLECE, WEI-SPS). That is, while there were publications documenting the outcomes of these ILSAs, the literature we reviewed did not specifically address how these other major ILSAs had made an impact on national educational policy or reforms efforts. This finding may be associated with the global prominence of PISA and OECD's recommendations to countries aspiring to improve their PISA results, as well as the much publicized "PISA Shock" (e.g., Hopfenbeck et al., 2017; OECD, 2014; Waldow, 2009; Young, 2015) some countries have experienced.<sup>11</sup>

**Data Analysis.** After we finalized our article corpus, we began the review by selecting a random group of eight articles to calibrate our review process. We each read the eight articles and then met to discuss our findings. During this initial discussion, we modified our review table to reflect what we agreed were the most salient factors for our review (e.g., country context, policy influence, evidence). We then proceeded to use this review tool with the remaining articles.

**Limitations.** Our literature review was limited to English, Spanish, and Portuguese publications, and likely excludes relevant literature published in other languages, such as Tillmann, Dederig, Kneuper, Kuhlmann, and Nessel's (2008) German-language book examining the effect of PISA on German educational policy. While the timeframe of our literature scan overlaps with the administration of

---

<sup>11</sup> The term "PISA Shock" was coined following the release of the first PISA report in December 2001 to describe countries, most notably Germany, that performed much lower than expected.

many of the most prominent and globally implemented GLMs, there may be relevant literature published before 2000. Our decision to exclude books also likely overlooked some relevant publications, but our inclusion of relevant available book chapters helped mitigate this possible limitation.

### Survey of ILSA/GLM Stakeholders

**Survey Instruments.** To complement our review of the ILSA/GLM literature, we developed and administered two surveys to explore the perceptions about the use of ILSAs/GLMs among researchers, policymakers, union representatives, and educators (see Appendix). Both surveys were designed to gather perceptions about which ILSAs/GLMs are used, how they are used, and to what extent ILSAs/GLMs contribute to or hinder national and global reform efforts and consisted of several closed- and open-ended questions. The second survey, however, included additional demographic items and respondents' use and engagement with ILSAs/GLMs. The surveys were piloted with five ILSA/GLM experts.

**Data Collection.** Our surveys employed both convenience and snowball sampling methods. We administered both surveys between November and December 2016.

**“Expert” Survey Sample.** We designed the first survey to capture the perceptions of individuals who are internationally and regionally well-known for their engagement with ILSAs/GLMs, including academic and non-governmental researchers, and individuals who work for organizations that administer ILSAs or are involved in GLM development. Based on our work on ILSAs/GLMs, we identified a convenience sample of 50 “experts” who we knew had extensive experience researching or working with ILSAs/GLMs and emailed them a link to the survey. This survey included a request for recommendations of colleagues who may like to share their perspectives about the use and influence of ILSAs/GLMs. This snowball sample produced 34 additional individuals for a final expert survey response rate of 29% ( $n=24$ ).

**“Non-Expert” Survey Sample.** We designed the second survey to capture the perceptions of individuals who are less experienced with ILSAs/GLMs but are interested in the issue, such as educators and policymakers. Participants consisted of registrants of the Inaugural Symposium of the Comparative and International Education Society held November 10-11, 2016, at Arizona State University. This Symposium focused specifically on the possibility and desirability of ILSAs/GLMs, and was attended by a diverse international group of ILSA/GLM researchers, policymakers, and other interested stakeholders. The survey was emailed to the entire list of non-expert symposium registrants ( $n=132$ ) and had a response rate of 50% ( $n=66$ ).

**Survey Analysis.** We administered both surveys via Qualtrics, and analyzed the quantitative items using the Qualtrics survey software. We exported the open-ended English and Spanish survey responses, had two research team members code them separately in Microsoft Excel, and then compared these codes to ensure inter-rater reliability. There were five questions that were identically phrased in the “Expert” and “Non-Expert” surveys. The responses to these items were aggregated for analysis to compare to the results that were disaggregated by expert or non-expert participant status.

## Research Findings

The following sections present, first, our review of the ILSA/GLM literature and, second, our analysis of the two ILSA/GLM stakeholder surveys. Presenting the findings in this order grounds the survey data in the wider literature. We would like to note upfront that we have purposefully referred to the implementation and use of ILSAs/GLMs by *governments* and/or *government agencies* instead of the more general, and more common, references to the work of *nations* and *countries*, as the agency to contribute to – or hinder – reform efforts are driven by the political interests of governments and policymakers.

### Review of the Literature: ILSAs/GLMs as Tools of Legitimation

Based on the media attention ILSAs receive, particularly after the release of PISA results and international rankings (Baird et al., 2016; Meyer & Benavot, 2013), it appears as though the increased participation in ILSA-related projects and studies has had a profound effect on education and policy discourses globally (Breakspear, 2012; Hopfenbeck et al., 2017), but not necessarily in the ways that make direct *causal* policy linkages easy to identify.<sup>12</sup> Our review of the literature indicates that some governments appear to have implemented policy changes as a result of ILSAs scores, while others seem to have used these scores to justify and/or accelerate pre-existing school reforms. Either way, ILSAs are effectively being used to legitimize these various educational policy reform agendas. Moreover, governments' increased efforts to improve their rankings, or “rank up,” have contributed to cross-national policy borrowing and widespread policy convergence (Baird et al., 2011; Breakspear, 2014; Lawn & Grek, 2012; Verger, 2014). There has also been increasing representation of studies examining ILSA involvement and outcomes in academic journals (Dominguez, Vieira, & Vidal, 2012; Lenkeit, Chan, Hopfenbeck, & Baird, 2015), which has helped to contribute to the broad perception that ILSA participation has had a significant impact on educational policy worldwide. In this way, ILSAs – and by extension GLMs – are being wielded by participating national governments as tools of legitimation to promote or justify creating, maintaining, or changing educational policies and reform efforts.

One of the key challenges in determining to what extent there is conclusive evidence about the relationship between ILSA/GLM participation and the direction of national educational reforms is that some countries with similar assessment outcomes have seen different governmental responses to educational reform policies. Two prominent examples followed the release of the PISA results in 2001 and 2004, in which a group of countries experienced what has been called “PISA Shock” (OECD, 2014; Waldow, 2009; Young, 2015). Policymakers in countries such as Germany, Japan, and Denmark, having found themselves considerably and surprisingly out-ranked by other countries around the world and under intense media and public scrutiny, reacted quickly to develop and implement new large-scale education reform policies (Breakspear 2012, 2014; Bulle, 2011). The German government created a new set of PISA-aligned national standards that allowed them to track student performance over time and developed new support systems for disadvantaged and immigrant students who scored particularly low on PISA (Breakspear, 2014; Ertle, 2006), as did Danish government. The vigorous public and political debate following Japan's drop in rank between the 2000 and 2003 PISA assessments, however, drove its policymakers to invest in national assessments and raise the stakes

---

<sup>12</sup> GLMs are also related to the development of “Big Data” movements and the use of learning analytics in educational systems. For a critical perspective on the topic see O'Neil (2016). Breakspear (2012) found that “PISA results have had an influence on policy reform in the majority of the participating countries/economies” and “over 85 percent of policy makers, local government officials, academics and researchers report having a relatively high level of knowledge of PISA processes and impact” (as cited in Hopkins et al., 2008, p. 19).

attached to them (Breakspear, 2014; Takayama, 2008).

This problem was also well described in Baird et al.'s (2016) study of policy and media reactions to the 2009 and 2012 PISA results in Canada, China (Shanghai), England, France, Norway, and Switzerland. The authors found that governments use their PISA participation as a “spell” or “magic wand” to justify any policy reform regardless of their actual PISA results. Using a different and friendlier framing of ILSAs, Lockheed and Wagemaker's (2013) exploration of the usefulness of ILSAs as policy tools describe them as either “thermometers” to measure progress or “whips” to incite policy action. Other scholars understand ILSAs as tools of “knowledge regulation” (Carvalho, Costa, & Afonso, 2009) in the spaces of policy decision-making, political theater and discourse, and practical capacity building to serve as educational change agents and not just benchmarks of access or progress (see Carvalho, 2012; Carvalho & Costa, 2015; Lascoumes & Le Gales, 2007). Regardless of how the positioning of ILSAs has been described in the literature, it is evident that educational and political authorities and political stakeholders in each country are using ILSAs to legitimize educational policies in ways that reflect their political contexts.

Throughout the literature we reviewed, we found evidence that national participation in ILSAs/GLMs and their resulting rankings have been used to legitimize and accelerate pre-existing or pre-planned educational reform efforts. For example, according to Baird et al. (2011), the French government exaggerated France's comparatively poor performances on TIMSS, PIRLS, and PISA to justify the systemic governmental reforms already underway. These reforms included a refocusing on fundamental skills through the revision of the primary school curricula with a particular focus on literacy and science, and decentralizing funding so that schools have control over their budgets. According to Dobbins and Martens (2012), France's PISA results, which were preceded by the country's poor performance on the 1997 OECD's International Adult Literacy Survey (IALS), resulted in public upset that provided government officials with the opportunity to quickly move forward with the intended structural reforms aimed at improving student performance, accountability, and school and teacher autonomy (see also Pons, 2016). This last reform effort – increasing pedagogical and financial autonomy for schools and teachers – reflects the French government's policy borrowing of what they perceive to be key drivers behind Finland's “miracle” PISA ranking (Dobbins & Martens, 2012).

Likewise, Switzerland's participation in PISA 2000 and the IEA PIRLS and TIMSS surveys contributed to their government's acceleration of pre-existing efforts to make the curricula and standards more unified across the country, as well develop better systems for monitoring educational progress within and across regions (Baird et al., 2011; Bieber & Martens, 2011). The UK's New Labour government has also used the country's PISA results to promote a narrative of declining educational standards that then justified reforms to General Certificate of Secondary Education (GCSE) targets (Sellar & Lingard, 2013). In Japan and Portugal, policymakers have used PISA as a subtler rhetorical ally to reframe and guide the proposed governmental reforms. Japanese policymakers have re-interpreted Finland's success through their preferred to justify curricular reforms and the implementation of standardized testing (Takayama, 2008, 2010). While the influence of PISA on educational reforms in Portugal is harder to disentangle from other factors (e.g., European Union education agenda), policymakers' concerns over Portugal's PISA results have coincided with the implementation of national assessments in Portuguese and math and a re-positioning of curricula in terms of competencies (Teodoro & Estrela, 2010).

In terms of the kinds of reforms governments have implemented, many have focused on developing

or strengthening their national assessment programs and frameworks as result of PISA participation. Engel (2015), for example, found that the Spanish government has increasingly used PISA as “a reference point for measuring priority areas of Spanish education policy” (p. 110). Norwegian policymakers, having also experienced “PISA shock,” implemented a national quality assessment and school evaluation system, as well as national curriculum and assessment projects (Baird et al., 2011). While there are other national governments that implemented new testing and evaluation systems during this time period, such as Romania, Sweden, Macedonia, Malaysia, Norway and Israel (Baird et al., 2011; Choi & Jerrim, 2016; Elley, 2002; Engel, 2015; Feinger et al., 2012; Forestier & Crossley, 2014), PISA cannot be isolated as the definitive cause as the global educational climate has been one of increased measurement and student evaluation and tracking.

Relatedly, some national governments have introduced new standards and/or modified their national curricula. The Shanghai government, for example, made substantive changes to their curriculum, such as shifting from pedagogies that focused on “transmission of content” to focusing more on “real-life” problems (Sellar & Lingard, 2013). Policymakers in Macedonia, Malaysia, Portugal, Romania, Spain, and Switzerland also developed new national standards and/or curricula (Baird et al., 2011; Choi & Jerrim, 2016; Elley, 2002; Engel, 2015; Teodoro & Estrela, 2010). Spanish policymakers, for instance, used PISA, PIRLS, and TIMSS to justify their simplification of the national curriculum by focusing on competencies, foreign languages, and information technology and communications (Choi & Jerrim, 2016; see also Gil, Beltran, & Redondo, 2016). Following Macedonia’s first TIMSS-R participation in 1999 – their first ever participation in an international assessment – policymakers worked to better align the math and science curricula with the assessment, revise national assessments, and implement in-service teacher training (Elley, 2002).

Another common reaction to ILSAs/GLMs was to propose modifying teacher quality policies, particularly in the areas of teacher training, evaluation, new teacher development, and teaching requirements. This was found to be the case in Spain, Estonia, Poland, Brazil, Colombia, Japan, Portugal, and Israel (Afonso & Costa, 2009; Choi & Jerrim, 2016; Engel, 2015). Paine, Bloemeke, and Aydarova (2016) argued that changes to teacher-related policies are common because this is one area that policymakers can easily influence. They posit that the most efficient way to influence student achievement outputs (on ILSAs) is by focusing on inputs. While complex factors, such as social class, might be the greatest predictor of student achievement, Paine and colleagues argued that focusing on “teacher quality” was “within reach” of the policymakers, resulting in various accountability reforms that tightened the control of teachers and teacher preparation programs.<sup>13</sup> It must be noted that such policy changes are difficult, if not impossible, to explicitly (and solely) attribute to ILSA/GLM participation. As Afonso and Costa (2009) argued, “PISA adapts to the contexts, is appropriated by the governments and is re-politicized, in line with a pre-existing agenda” (p. 61).

Some scholars have pointed to the discourse created by the media’s coverage of ILSA results as the justification used governments for their educational reform efforts (Baroutsis & Lingard, 2016; Sellar & Lingard, 2013; Takayama, 2008). Arguably, and as indicated above, ILSAs/GLMs have had a greater rhetorical effect than a direct policy one (Baird et al., 2011; Paine et al., 2016; Takayama, 2010; Teodoro & Estrela, 2010). In this sense, ILSAs/GLMs have come to serve as a policy rationale that

---

<sup>13</sup> We want to highlight that focusing on “teacher quality” was not just within reach, and an easy target, it is millions of dollars cheaper than to change the public services attending poor neighborhoods, housing policies, providing health care, addressing unemployment and other social policies that could reduce social inequality, which is highly correlated to higher ILSAs scores. See Berliner (2017), Powers, Fischman, & Berliner (2016), and Sahlberg (2015).

helps to legitimize policy reform efforts. Berényi & Neumann (2009) reported that during the 2002 Hungarian election, the liberal party increasingly referenced PISA results as a reason to support their education agenda (see also Bajomi, Berényi, Neumann, & Vida, 2009). Hong Kong's success on PISA has drawn the attention of the UK's authorities, who also looked to Singapore, Sweden, and Finland for best practices (Forestier & Crossley, 2014). The UK's then Secretary of State of Education, Michael Gove, frequently referenced PISA and high-scoring countries to justify his education reform efforts (Bittlingmayer, Boutiuc, Heinemann, & Kotthoff, 2016; Forestier & Crossley, 2014; see also Baird et al., 2011). The use of comparisons and quasi-educational rankings was not uncommon across nations, as many countries experienced a discursive shift when their media began to report on ILSA results. Several countries with educational systems as different as Argentina, Australia, France, Germany, Japan, Portugal, Spain, Turkey, and the USA have experienced similar rhetorical effects, (Afonso & Costa, 2009; Baroutsis & Lingard, 2016; Froese-Germain, 2010; Gorur, 2011; Rautalin & Alasuutari, 2008; Takayama, 2008; Vega Gil, Beltrán, & Redondo, 2016). These countries have also implemented various education policies that have been associated with an increasingly global connectedness and competition (made possible by ILSAs participation), such as Australia's introduction of national teaching standards (Breakspear, 2014; Figazzolo, 2009; Gorur, 2011; Sellars & Lingard, 2013), the USA's use of student achievement scores to measure teacher quality (Bieber & Martens, 2011; Engel & Frizzell, 2015; Figazzolo, 2009; Martens & Neimann, 2013; Plisko, 2013; Sellar & Lingard, 2013), and Spain's narrowing of their curriculum to focus on tested subjects (Choi & Jerrim, 2016) and competencies (Gil, 2016).

Our review found evidence that ILSA/GLMs have also been used as legitimation tools at the local or subnational levels. Morgan (2015) found a quick subnational adoption across Canada of PISA-friendly reforms. Similarly, Engel and Frizzell (2015) document how Canada's provinces have used PISA scores to validate how their existing policies have contributed to high student performance (e.g., Alberta) or to validate their reforms to combat low student performance (e.g., Saskatchewan). Although there has been limited documented evidence of the influence of ILSA/GLMs on national education reform efforts in the USA, likely due to the country's historical disinterest in ILSAs/GLMs (Bieber & Martens, 2011), Engel and Frizzell (2015) found that some policymakers at the state level have begun using their PISA results as evidence of the effectiveness of their educational systems. Massachusetts, for example, recently cited their 2012 PISA results as proof of their national and international competitiveness, and Florida incorporated their PISA results into their federal Race to the Top grant proposal (Engel & Frizzell, 2015). Interestingly, while the connection between ILSAs/GLMs and the USA's educational reform efforts at the national level has been largely unobserved, Martens & Niemann (2013) document the parallels between PISA's stated goals and the goals of the Race to the Top program, as articulated in the American Recovery and Reinvestment Act of 2009, and the reauthorized education act of March 2010, 'A Blueprint for Reforms.' Froese-Germain (2010) suggests that PISA is being used as a subtle ally in the USA to justify educational reforms aimed at increasing efficiency and competitiveness (and likely in other countries as well).

According to Breakspear (2014), six countries have started to set new goals (as identified by Breakspear; Table 1) for international ranking and more outward-looking policy- and practice-borrowing (e.g., looking to high-ranking countries for "best practices"). Table 1 illustrates the PISA-based goals identified by Breakspear (2014) for a number of OECD countries. with very little actual practical impact.<sup>14</sup>

---

<sup>14</sup> It should be noted that only a small number of national representatives (one or two) from each country were interviewed as part of the original study (Breakspear, 2012) and at a particular point in time. That said,

**Table 1. PISA-based education goals**

Country	Goal
Australia	To be ranked in top 5 PISA countries by 2025
Brazil	To reach OECD average PISA performance by 2021, and all schools and regions are rated on the Index of Development of Basic Education, which is benchmarked to PISA
Denmark	To be ranked in top 5 PISA countries
Mexico	To reach 435 in math and reading by 2012
Thailand	To reach OECD average PISA performance by 2021
Wales	To be ranked in top 20 PISA countries by 2015

Source: Breakspear, 2014

An outgrowth of ILSAs/GLMs increasing influence on the debates about education reform is their contribution to the creation of national agencies of educational evaluation and the related development of the technical capacity to apply, monitor, and interpret the results (Gove & Cvelich, 2011; Greger, 2012; Lewis & Lingard, 2015; Strietholt & Scherer, 2017). Gove and Cvelich's (2011) report on EGRA outcomes in various participating countries indicating the scaling-up efforts in low-income countries that historically had limited evaluation capacity. Similarly, prior to PISA Germany and the Czech Republic did not implement national testing, and the resulting "shock" prompted the governments of both countries to develop ways to more formally assess student progress (Greger, 2012).

Above and beyond setting target goals for PISA scores or rankings, and despite obvious contextual differences, ILSAs/GLMs provide information about the competencies of any given country's students, and, ultimately provide indications about what could be considered national standards about the quality of their education systems. Moreover, Carvalho and Costa's (2015) examination of the political reception of PISA in six European countries suggests that PISA in particular "provides optimism about the possibility of reform and creates confidence in national policy actors" and legitimizes "policy problems and solutions with the blessing of putative universal, independent, expert knowledge" (p. 644). This shift in conditions of possibility have influenced education systems in various ways around the world, including the generation of a robust educational market for ILSA/GLMs, including the development of an entire sector dedicated to implement, and process ILSAs (e.g., IEA, OECD, Pearson, UNESCO, etc.) that included the development of assessment capacities, such as national assessment units and data collection systems, and the emergence of a new professional sector of ILSA consultants.

### **Stakeholder Surveys: ILSAs/GLMs Perceived as Instrumental to Policy Reform**

A combined total of 90 people responded to the surveys, including 24 experts directly working with ILSAs/GLMs, 40 researchers and policymakers, and 13 graduate students pursuing research on ILSAs/GLMs, as well as other education stakeholders (teachers, university administrators, and foundations/donors personal). Survey A specifically targeted ILSA/GLM experts who all had experience working on ILSAs/GLMs in their own and often in other national contexts. The majority (72%) of non-expert survey respondents (e.g., educators and policymakers) stated that they also have experience working on ILSAs/GLMs with educational stakeholders (e.g., politicians, policymakers,

---

numerous governments have made similar types of goals, even though in reality these kinds of goals, which are intended to focus educational reform efforts, serve more of a political function



unions, teachers, media, etc.) in different national contexts, with 23% ( $n=15$ ) reporting that they work frequently, 49% ( $n=32$ ) reporting that they work occasionally, and 28% ( $n=18$ ) reporting that they never with stakeholders ( $n=65$ ). The respondents' to both surveys combined experiences with ILSAs/GLMs covered a variety of geographic regions, ranging from Africa to Latin America, Asia, Western and Eastern Europe, Middle East, North America, and Australia.

When all 90 respondents were asked about their experiences working with ILSAs/GLMs in different national contexts, 58% ( $n=52$ ) responded that education stakeholders are primarily concerned with international student achievement studies such as PISA (40%), TIMSS (24%), PIRLS (12%), and PIAAC (9%). Respondents also mentioned other metrics that may play a role in national education policy contexts, including USAID's EGRA, Save the Children's IDELA (International Development Learning Assessment), UNESCO's TERCE (Third Regional Comparative and Explanatory Study), as well as national, state, and municipal student assessments (e.g., Brazilian SAEB and ENEM).<sup>15</sup> The following sections detail respondents' perspectives on the use and impact of ILSAs/GLMs in national policy contexts, and ILSAs/GLMs as contributions or hindrances to policy reform efforts.

***ILSAs/GLMs in National Policy Contexts.*** While the existing literature may be inconclusive about the causal relationship between ILSA/GLM results and their use in national education policy contexts, both the expert and non-expert surveys generated a number of perspectives on the topic. In particular, the vast majority of all respondents (96%) agreed that ILSAs/GLMs are used in national education contexts (Table 1). However, the respondents' opinions were split regarding how the metrics are used in different policymaking contexts. In particular, 58% of all respondents outlined different instrumental ways in which ILSAs/GLMs are used in national education contexts, while 38% of respondents offered criticisms of the misuse of ILSAs/GLMs. Only 4% of all respondents stated that metrics are not used at all in national policymaking processes. Focusing on the instrumental use of ILSAs/GLMs in different national contexts, the respondents explained that metrics are generally used to set goals, targets, and benchmarks for national education reforms, especially in terms of education equity and quality agenda, as well as to monitor the implementation of these reforms. As one researcher explained, for example: "It is not possible to talk about improving education without a measure." Three respondents noted that ILSAs/GLMs are also used to align curriculum and national student assessment instruments to international/global standards. In addition, survey respondents stated that ILSAs/GLMs are commonly used as a "policy trigger/lever" to justify existing reforms, validate and legitimize policy changes, or demonstrate the need for funding, especially in international development contexts, which is consistent with our review of the ILSA/GLM literature. The percentage of respondents who discussed ILSAs/GLMs in terms of their instrumental value was higher among non-expert than expert respondents (63% and 52%, respectively). In other words, a larger majority of respondents with an interest in ILSAs/GLMs but with, perhaps, less in-depth work experience with these metrics responded favorably that ILSAs/GLMs are used in instrumental ways.

More than one-third (38%) of all respondents thought that ILSAs/GLMs are generally misused in national policy contexts, with the higher percentage of expert survey respondents discussing misuse of metrics compared to non-expert survey respondents (43% and 33%, respectively). While acknowledging the circulation of metrics locally and globally, these experts and non-experts respondents explained that policymakers are aware of ILSAs/GLMs, but do not use them

---

<sup>15</sup> ENEM is the National Exam of Upper Secondary Education that high school students take in Brazil to be considered for university admission. SAEB is the National System of Basic Education Assessment aiming to evaluate the quality of instruction.

meaningfully for several reasons. According to survey respondents, some policymakers have strong reservations about using ILSAs/GLMs because they believe that such metrics are, as described by one expert respondent, “imperfect snapshots that often distort more than illuminate.” Second, ILSAs/GLMs are too broad and abstract to be used in in any meaningful ways. Third, respondents noted that policymakers might simply lack the skills of analyzing and interpreting ILSA/GLM data to make it useful for their national contexts.

Although policymakers may look at comparative data generated by ILSAs/GLMs, the survey responses indicate that in general policymakers do not use ILSA/GLM data for in meaningful ways. Rather than guiding reforms, one respondent suggested that ILSAs/GLMs lead to the construction of the narratives of “crisis” and “need,” as widely documented in our review of the literature. These narratives skew national educational reforms in particular directions, often increasing teacher and student accountability and a focus on learning competencies. Also consistent with our review of the literature, four of the expert respondents noted that policymakers in some countries are more likely to make political or symbolic references to ILSAs/GLMs to invoke a *name, shame, and blame* strategy or compare their country’s standing in global rankings, as reflected in this respondent’s comment: “Policymakers do not use metrics, they are just concerned about the rankings.” Just as the ILSA/GLM literature indicates that these assessments and metrics are being used as tools of policy legitimation, the majority of our expert and non-expert survey respondents are seeing this actualized at the national level.

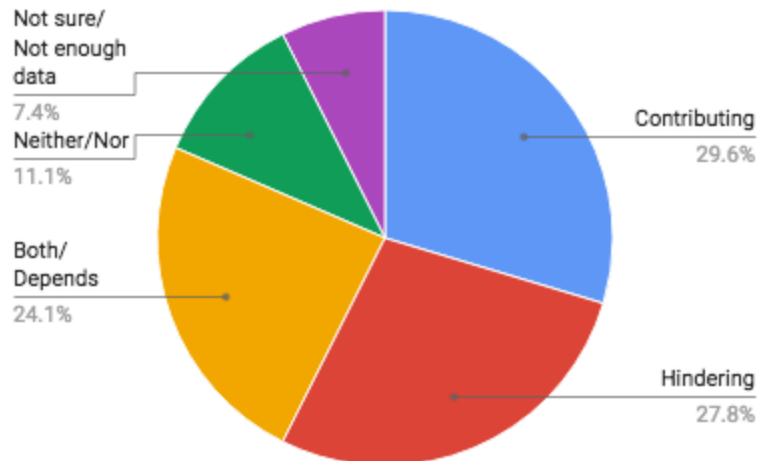
**Table 2. The use of ILSAs/GLMs in national education policy contexts**

Use of ILSAs/GLMs	Expert Survey	General Survey	Combined Responses
<b>Used for Instrumental Purposes</b> <ul style="list-style-type: none"> <li>● To set goals, targets, and benchmarks for reforms and measure reform progress</li> <li>● Validate policy change or justify existing reforms</li> <li>● Demonstrate the need for funding</li> <li>● To adapt and/or develop national testing</li> </ul>	52%	63%	58%
<b>Misused</b> <ul style="list-style-type: none"> <li>● Policymakers have little understanding of ILSAs/GLMs</li> <li>● ILSAs/GLMs are decontextualized and therefore not helpful</li> <li>● ILSAs/GLMs are too broad and abstract to be used meaningfully</li> <li>● For “name, shame, and blame” purposes</li> <li>● For ceremonial effects</li> </ul>	43%	33%	38%
<b>Not Used</b>	5%	4%	4%
<b>Number of Respondents</b>	21	24	45

**Perceived Impact of ILSAs/GLMs on National Policymaking.** Although the vast majority of all respondents (96%) indicated that ILSAs/GLMs are being used in national policymaking contexts, their opinions were split regarding the question of whether ILSAs/GLMs helped or hindered national education reform processes (Figure 1). In particular, 30% of all respondents believed that ILSAs/GLMs contributed to national reforms, while 28% of all respondents thought that they hindered national education reform processes ( $n=54$ ). Approximately one-quarter (24%) of all respondents thought that ILSAs/GLMs could both hinder and contribute to reform processes,

depending on the particular circumstances and the ways in which these metrics are used locally. Finally, 11% of the respondents thought that ILSAs/GLMs neither helped nor hindered national education reform processes and 7% stated that there was not enough information or evidence to determine the effect of ILSAs/GLMs on national policymaking processes.

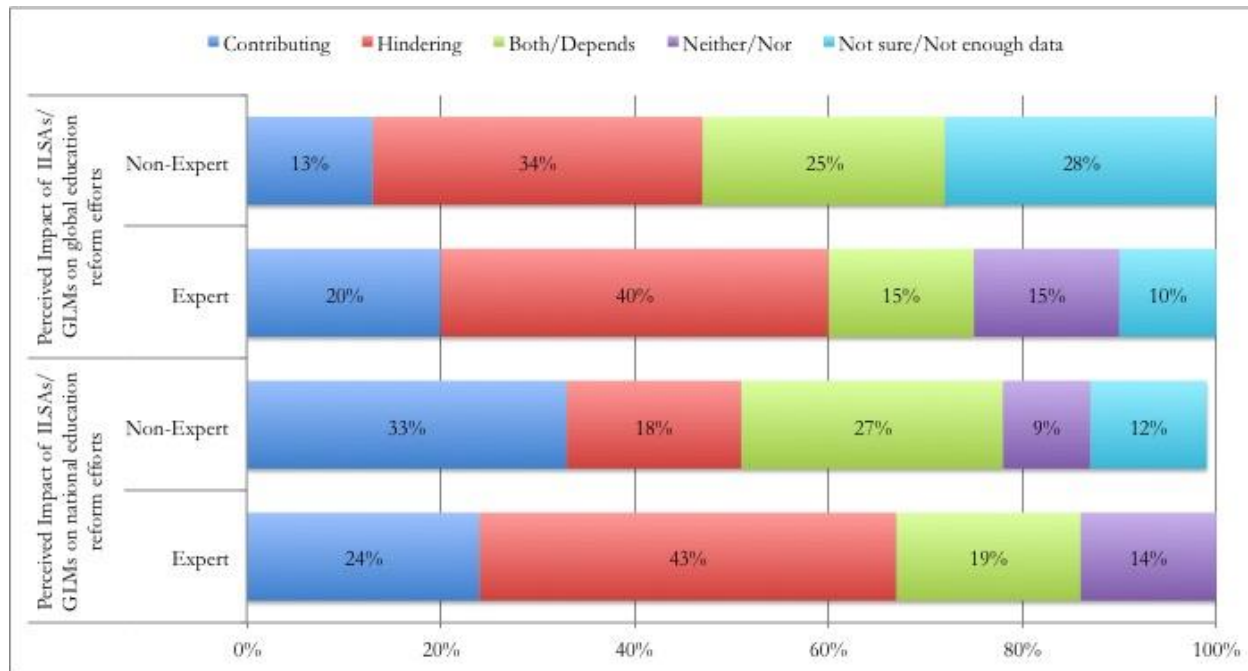
**Figure 1. Survey respondents' perceptions of the contribution and/or hindrance of ILSAs/GLMs to national education reform efforts**



Note: Includes responses from both the expert and non-expert surveys.

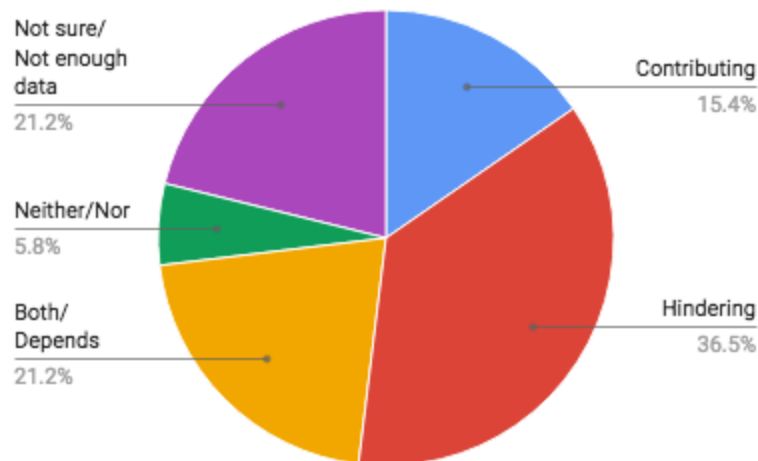
Again, respondents to the expert survey and non-expert survey differed in their perceptions regarding the contribution and hindrance of ILSAs/GLMs to national education reform efforts (Figure 2). The experts were more critical of ILSAs/GLMs, with 43% of expert survey respondents reporting that ILSAs/GLMs are hindering reform efforts compared with only 18% of non-experts. It is possible that experts' first-hand experience of working with ILSAs/GLMs in various national contexts informed their opinions regarding the complex ways in which metrics impact national policies.

**Figure 2. Survey respondents' perceived impact of ILSAs/GLMs on national and global education policies**



We saw a shift in respondents' perceptions of the contribution and hindrance of ILSAs/GLMs to global education reform efforts compared with their responses to ILSAs/GLMs role in national education reforms (Figure 3). In the context of global reforms, more than one-third (37%) of all respondents thought that ILSAs/GLMs hindered education policymaking ( $n=52$ ). Only 15% of all respondents believed that ILSAs/GLMs contributed to global reform efforts ( $n=52$ ). Comparing responses to the two surveys, 33% ( $n=33$ ) on non-expert survey respondents thought that ILSAs/GLMs are contributing to national reform efforts, but only 13% ( $n=32$ ) felt that they are contributing to global reform efforts. Similarly, only 18% ( $n=33$ ) of the non-experts reported that ILSAs/GLMs hinder national reform efforts while 34% ( $n=32$ ) responded that ILSAs/GLMs hinder reform efforts on a global scale. The expert survey respondents consistently reported that ILSAs/GLMs contributed or hindered both national and global reform efforts. For example, 24% ( $n=21$ ) of experts responded that ILSAs/GLMs contribute to national reform efforts and 20% ( $n=20$ ) of experts responded that ILSAs/GLMs contribute to global reform efforts. Similarly, there was little change in experts' perception of ILSAs/GLMs hindering national and global reform efforts. Of the experts surveyed, 43% ( $n=21$ ) felt that ILSAs/GLMs hinder national reform efforts while 40% ( $n=20$ ) responded that that ILSAs/GLMs hinder global reform efforts.

**Figure 3. Survey respondents' perceptions of the contribution and/or hindrance of ILSAs/GLMs to global reform efforts**



Note: Includes responses from both the expert and non-expert surveys.

***ILSAs/GLMs as Contributors to National/Global Education Reform Efforts.*** The respondents who thought that ILSAs/GLMs contribute to national education reform efforts (30%  $n=54$ ) explained that ILSAs/GLMs enable governments to align their educational systems to international/global standards, including the alignment of testing and examinations systems and curriculum. As one of the expert survey respondents explained, ILSAs/GLMs are effectively used as a reform “trigger” or “lever” to advance particular education reform agendas, but without providing specifics about those agendas. Not surprisingly, this is exactly the main point of critique by those policy/research experts who are concerned with the rise of ILSAs/GLMs worldwide (Bieber & Martens, 2011; Engel & Frizzell, 2015; Figazzolo, 2009; Martens & Neimann, 2013; Plisko, 2013; Sellar & Lingard, 2013), and consistent with our literature review findings.<sup>16</sup>

***ILSAs/GLMs as Hindrances to National/Global Education Reform Efforts.*** Several of the non-expert respondents (13%,  $n=24$ ) noted that ILSAs/GLMs are a “distraction” that perverts the purpose of education by reducing it to simply output measures and shifts the focus from student learning to receiving a high ILSA ranking. One expert survey respondent explained: “Global learning metrics are mainly hindering reform efforts because countries seem to become fixated on their league

<sup>16</sup> Some respondents echoed the arguments historically used - and widely critiqued - in comparative and international education to rationalize the use of standardized comparison (e.g., Lockheed & Wagemaker, 2013; Lascoumes & Le Gales, 2007; Schleicher, 2009; Schleicher & Shewbridge, 2004) suggesting that ILSAs/GLMs allow governments to identify problems in their education systems, search for “best practices” in other contexts, and learn from experts in top-performing countries about “what works” in their education systems. The assumption in this international focus on quality education is that best policies and practices can be successfully transferred to different national contexts.

table position rather than on a sound discussion of what education quality is.” Echoing the same point, another respondent noted:

Global learning metrics tend to stultify the conversation and thinking on what is important. I am not arguing that they lack value, but they should not dominate the conversations on educational effectiveness when there is so much “gaming” going on and so many other important dimensions of human development that can be affected by education.

More importantly, several experts noted that the reliance on ILSAs/GLMs tends to impose a neoliberal education reform agenda globally, focusing on policy performativity and outcomes rather than on the process of education and whether or not that process is just. While advancing a neoliberal reform agenda, ILSAs/GLMs simultaneously contribute to erasing alternative policy options. As one of the respondents stated, ILSAs/GLMs “prevent nations from developing a caring childhood and cooperative values and creativity.... [instead] childhoods globally are becoming increasingly less democratic or open, turning children into entrepreneurial adults.” In other words, ILSA/GLM critics emphasized that ILSAs/GLMs narrow down policy visions and reduce policy discussions to a predetermined set of policy goals and outcomes, which promotes neoliberal education reform agenda.<sup>17</sup>

***ILSAs/GLMs as Both Contributors and Hindrances.*** Importantly, almost one-quarter (24%,  $n=54$ ) of all respondents thought that ILSAs/GLMs can both hinder and contribute to the national education reform efforts at the same time. As one respondent remarked, ILSAs/GLMs may “be too flawed to make strong policy recommendations but they have some use in pointing out possible areas of strengths and weaknesses.” Similarly, other expert survey respondents noted that ILSAs/GLMs are useful in their benchmarking function but not as influencers of local policy, as articulated by one respondent: “an important external benchmark, but they do not offer much information about local needs or local educators’ values and perspectives for education reform.” Implicit in many of these responses is the critique of how ILSAs/GLMs are used at the national level, pointing out that the potential of ILSAs/GLMs is thwarted by the lack of capacity among national policymakers to use metrics in meaningful ways. In particular, several expert survey respondents explicitly stated one of the main reasons for the misuse of ILSAs/GLMs is the lack of “PISA-literacy” among many policymakers. In other words, these experts conclude that a key problem is the limited institutional technical capacity at the national and local levels rather than political momentum.

On a more critical note, some experts suggested the discussion of ILSAs/GLMs cannot be reduced to how to make and compare metrics. Rather, education stakeholders who use ILSAs/GLMs need to continuously consider the political dimensions because “the distribution of education is based on power.” Similarly, another expert survey respondent noted that one of the major effects of ILSAs/GLMs seems to be “a strengthening of the unequal distribution of educational power and prestige and a reordering of the global cultural map with the ascension of ‘Confucian’ values.” However, other survey respondents noted that ILSAs/GLMs could perhaps be useful if they were

---

<sup>17</sup> Well-known critics such as Gorur (2016), Grek (2009), Sellar & Lingard (2013) argue that ILSAs/GLMs narrow educational goals, hinder creativity, and stifle real discussions about education quality and equity in different national settings. In this context, policy discussions become reduced to simplistic sound bites, overlooking the importance of contextual factors and missing the “how” of education policy. In addition, they are culturally insensitive to many countries, expensive to administer, and results are not shared with the ultimate stakeholders, teachers and students.

used as one of many measures to understand education development. While such a diversification of assessments and metrics based on both the geopolitical and the educational value dimensions would not make a “neat and tidy” system of benchmarking, it could thus open possibilities for more contextually relevant and pedagogically innovative use of metrics in national and global education reform contexts.

## Discussion

ILSA/GLM participation has generated the perception of their having a profound affect on education policy, but primarily in ways that make it difficult to draw direct, causal relationships between the two (Paine et al., 2016). Our research confirms that some countries have the same ILSA results but implement different policies, and vice versa, even though we cannot definitively say that these policy convergences/reactions can be explicitly linked to ILSAs/GLMs participation and performance. Figure 4 offers a visual representation of the data presented in this paper, and the divergence between how the perceived use of ILSAs/GLMs and the less linear reality. Our research sought to document the extent to which the use of ILSAs/GLMs have influenced national primary/secondary education policymaking and the changes in education policies and practices, and as this figure shows we found that the way the education community thinks about the use of these assessments differs greatly in practice.

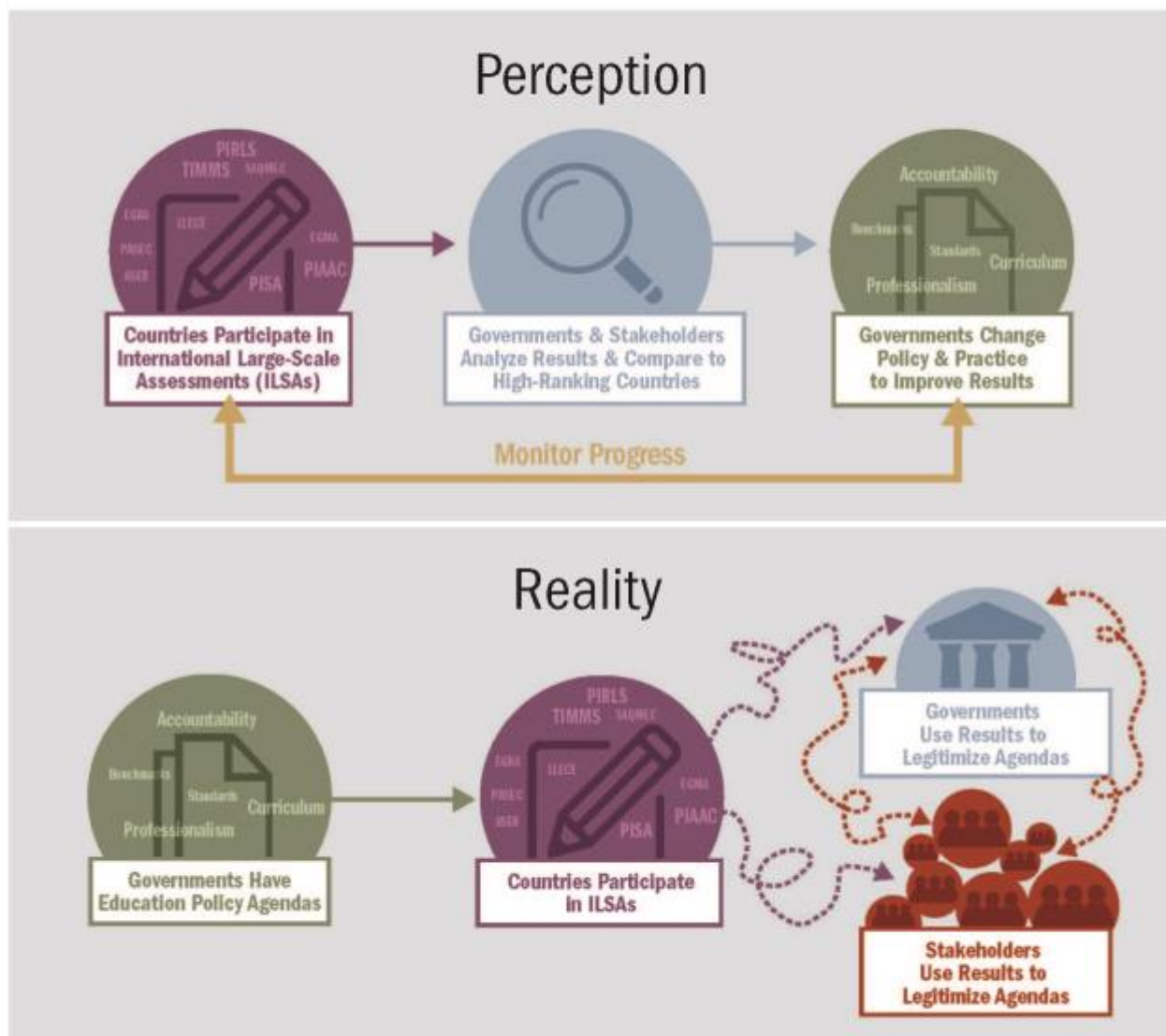
All participating countries have long histories of education policy reform that extend above and beyond ILSAs/GLMs participation. In the United States, for example, the government’s shift towards policies emphasizing teacher, school, and state accountability based on students’ performance on standardized tests similar to ILSAs has historically been attributed to the publication of *A Nation at Risk* in 1983, but the U.S. Department of Education’s bi-annual National Assessment of Education Progress (NAEP) preceded this report by 20 years. With the prominence of ILSAs over the past two decades, and PISA especially, national performance on these assessments have become the new indicator to justify reforms like the No Child Left Behind and Race to the Top legislation. When national education policy is viewed historically, it is a challenge to determine what reform or policy is attributable to what event. Every country that participates in ILSAs has a similar complicated story, and ILSA/GLM participation becomes one more string in the tangled web of preexisting and emergent policy movements. In sum, in spite of the strong rhetoric about influence, these policies are all contextually contingent and are perhaps best studied as detailed national case studies.<sup>18</sup>

---

<sup>18</sup> That said, there are a number of country-specific reports under the OECD’s Strong Performers and Successful Reformers series, such as for the United States, Japan, and Korea, which include detailed policy suggestions to these governments. Furthermore, many governments (Denmark, Sweden, Chile, Wales, Scotland, and so on) have invited the OECD to review their national education policies and suggest reforms to move ahead. It is therefore hard to say whether the impacts you identify here are due to PISA alone or as a result of combined OECD influence.

Figure 4. Perception and reality of the use of ILSAs

## International Large-Scale Assessments and Global Learning Metrics: Good Solutions in Search of the Right Problems?



*Design by Next Chapter Communications/Tangible Designs*

That said, our review of the literature and our survey findings indicate that ILSAs/GLMs have made it



possible for governments to imagine new models of educational policy and practice legitimization (Breakspear, 2014; Lawn & Grek, 2012; Verger, 2014). For Baird et al. (2011) ILSAs/GLMs have done less to distinguish policy differences and more to foster policy convergence. The extended use of international comparison and rankings, coupled with how mass media reports on ILSA/GLM results are a key dynamic in the national debates about educational systems (Haas & Fischman, 2010). Our review sees a clear trend that educational authorities at the national level are pressured to show improvements in ILSAs scores or GLM rankings, and governments have begun seeking strategies to move into high-ranking positions (Bieber & Martens, 2011; Engel, 2015; Meyer & Benavot, 2013). Despite contextual factors that inexorably effect how students perform on various ILSAs, international competition has fostered a sense of decontextualized, “what works” mentality (Lewis, 2016), where governments attempt to emulate the policies and practices of high-scoring countries (Engel, 2015; Lingard, Rawolle, & Taylor, 2005; Phillips & Ochs, 2003). However, as argued by Lewis and Lingard (2015), these “supposedly ‘objective’ and decontextualised PISA measures are rendered in ways that reflect the local political and cultural contexts in which the data are received, highlighting the contingent uptake of PISA knowledge within national schooling systems, despite its global circulation” (p. 627).

As such, the direct link between ILSA scores (or GLM rankings) and education policy change is a difficult, if not impossible, endeavor. But what we can acknowledge is that ILSAs/GLMs create a purportedly objective field of comparison that can, and do, fundamentally change the conditions and global contexts affecting educational policymaking. As a result, governments have new pressures to obtain a favorable ranking and, as a byproduct, to assure that their policies are well-suited for that purpose. Regardless, it is growing increasingly evident that the ILSAs/GLMs are having a profound influence on how stakeholders worldwide are making decisions on education policies and political agendas by serving to identify the results that fit previously established narratives of good educational reforms.

## Abbreviations and Acronyms

EFA	Education for All
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Mathematics Assessment
EU	European Union
GEFI	Global Education First Initiative
GMR	Global Monitoring Report
GPE	Global Partnership for Education
IEA	International Association for the Evaluation of Educational Achievement
ICCS	International Civic and Citizenship Education Study
LAMP	Literacy Assessment and Monitoring Programme
LMTF	Learning Metrics Task Force
MDG	Millennium Development Goal
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
STEP	Skills Towards Employability and Productivity
TIMSS	Trends in International Mathematics and Science Study
USAID	U.S. Agency for International Development
UIS	UNESCO Institute for Statistics
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Cultural and Scientific Organization
WEI-SPS	World Education Indicators' Survey of Primary Schools

## Glossary

<b>International Large Scale Assessments</b>	<b>Begin Date</b>	<b>Sponsoring Agency</b>	<b>Description</b>
Trends in International Mathematics and Science Study (TIMSS)	1995	International Association for the Evaluation of Educational Achievement (IAEEA)	TIMSS monitors trends in mathematics and science achievement every four years, at the fourth and eighth grades.
Program for International Student Assessment (PISA)	2000	Organisation for Economic Cooperation and Development (OECD)	PISA is a triennial international survey, which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students.
Progress in International Reading Literacy Study (PIRLS)	2001	International Association for the Evaluation of Educational Achievement (IAEEA)	PIRLS monitors trends in reading achievement at the fourth grade.
Literacy Assessment and Monitoring Programme (LAMP)	2003	UNESCO Institute for Statistics (UIS)	LAMP provides the diagnostic information required to monitor and improve literacy skills.
World Education Indicators' Survey of Primary Schools (WEI-SPS)	2005	OECD, UIS	WEI-SPS examines the main issues and inputs shaping primary schools.
Early Grade Reading Assessment (EGRA)	2006	RTI International with support from the U.S. Agency for International Development (USAID) and the World Bank	EGRA is an open-source assessment used to measure foundational reading skills.
Programme for the International Assessment of Adult Competencies (PIAAC)	2008	OECD	PIAAC measures the key cognitive and workplace skills needed for individuals to participate in society and for economies to prosper.
Early Grade Mathematics Assessment (EGMA)	2009	RTI International with support from the U.S. Agency for International Development (USAID) and the World Bank	EGMA is an open-source assessment used to measure foundational math skills.
Skills Towards Employability and Productivity (STEP)	2011	World Bank, Human Development Network (HDN)	STEP measures labor market skills in low and middle-income countries.

## References

1. Afonso, N., & Costa, E. (2009). A influência do Programme for International Student Assessment (PISA) na decisão política em Portugal: o caso das políticas educativas do XVII Governo Constitucional Português. *Sísifo. Revista de Ciências da Educação*, 53-64.
2. Aksit, N. (2007). Educational reform in Turkey. *International Journal of Educational Development*, 27(2), 129–137. <https://doi.org/10.1016/j.ijedudev.2006.07.011>
3. Ary, D., Jacobs, L., Sorensen, C., & Razavieh, A. (2006). *Introduction to research in education (8th Ed)*. Belmont, CA: Wadsworth CENGAGE Learning.
4. Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., & Daugherty, R. (2011). Policy effects of PISA. Report Commissioned by Pearson UK. Retrieved from [https://www.researchgate.net/profile/Talia\\_Isaacs/publication/262007629\\_Policy\\_effects\\_of\\_PISA\\_Pearson\\_UK/links/0046353b2b0f37aa20000000.pdf](https://www.researchgate.net/profile/Talia_Isaacs/publication/262007629_Policy_effects_of_PISA_Pearson_UK/links/0046353b2b0f37aa20000000.pdf)
5. Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G., & Yu, G. (2016). On the supranational spell of PISA in policy. *Educational Research*, 58(2), 121-138. <https://doi.org/10.1080/00131881.2016.1165410>
6. Bajomi, I., Berényi, E., Neumann, E., & Vida, J. (2009). The reception of PISA in Hungary. Project KNOWandPOL, WP12. Retrieved October 29, 2013, from <http://www.knowandpol.eu/>
7. Baroutsis, A., & Lingard, B. (2016). Counting and comparing school performance: An analysis of media coverage of PISA in Australia, 2000-2014. *Journal of Education Policy*, 1-18. <http://dx.doi.org/10.1080/02680939.2016.1252856>
8. Becker, F. (2014). Large-scale assessments and quality in basic education. In Napier, D. B. (Ed.), *Qualities of Education in a Globalized World* (pp. 63-81). New York, NY: Springer.
9. Berényi, E., & Neumann, E. (2009). Grappling with PISA. Reception and translation in the Hungarian policy discourse. *Sísifo*, 10, 41–52.
10. Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education*, 46(1), 101–116.
11. Bittlingmayer, U. H., Boutiuc, A. F., Heinemann, L., & Kotthoff, H-G. (2016). Jumping to conclusions- The PISA knee-jerk: Some remarks on the current economic-educational discourse. *European Education*, 48(4). <http://dx.doi.org/10.1080/10564934.2016.1249188>
12. Breakspear, S. (2012). *The Policy Impact of PISA* (OECD Education Working Papers No. 71). <https://doi.org/10.1787/5k9fdqfqr28-en>
13. Breakspear, S. (2014). *How does PISA shape education policy making? Why how we measure learning determines what counts in education*. (Seminar Series Paper No. No. 240). East Melbourne, Australia: Centre for Strategic Education.
14. Bulle, N. (2011). Comparing OECD educational models through the prism of PISA. *Comparative Education*, 47(4), 503–521.
15. Carnoy, M., Khavenson, T., & Ivanova, A. (2015). Using TIMSS and PISA results to inform educational policy: a study of Russia and its neighbours. *Compare: A Journal of Comparative and International Education*, 45(2), 248–271. <https://doi.org/10.1080/03057925.2013.855002>
16. Carvalho, L. M. (2012). The fabrications and travels of a knowledge-policy instrument. *European Educational Research Journal*, 11(2), 172-188.
17. Carvalho, L. M., Costa, E., & Afonso, N. (2009). PISA—Fabrication, circulation and use in six European Countries: Integrative report. Project KNOWandPOL, WP17. Retrieved from: [http://www.knowandpol.eu/IMG/pdf/o31.pisa.fabrication\\_circulation\\_use.pdf](http://www.knowandpol.eu/IMG/pdf/o31.pisa.fabrication_circulation_use.pdf)
18. Carvalho, L. M., & Costa, E. (2015). Seeing education with one's own eyes and through PISA lenses: considerations of the reception of PISA in European countries. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 638–646. <https://doi.org/10.1080/01596306.2013.871449>
19. Choi, Á., & Jerrim, J. (2016). The use (and misuse) of PISA in guiding policy reform: the case of Spain.

- Comparative Education*, 52(2), 230–245. <https://doi.org/10.1080/03050068.2016.1142739>
20. Dobbins, M., & Martens, K. (2012). Towards an education approach à la *finlandaise*? French education policy after PISA. *Journal of Education Policy*, 27(1), 23–43. <https://doi.org/10.1080/02680939.2011.622413>
  21. Domínguez, M., Vieira, M.-J., & Vidal, J. (2012). The impact of the programme for international student assessment on academic journals. *Assessment in Education: Principles, Policy & Practice*, 19(4), 393–409. doi:10.1080/0969594X.2012.659175
  22. Drager, J. (2012, July). *Accountability as a driver for reform: The “PISA shock” of 2001 - a spotlight on the case of Germany*. PowerPoint Presentation, Cambridge, Massachusetts.
  23. Dubek, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, 40, 315–322. <https://doi.org/10.1016/j.ijedudev.2014.11.004>
  24. Education International. (2011). Policy paper on education: Building the future through quality education.
  25. Elley, W. B. (2002). *Evaluating the impact of TIMSS-R (1999) in low-and middle income countries*. An Independent Report on the Value of World Bank Support for an International Survey of Achievement in Mathematics and Science. Unpublished. Retrieved from [http://pub.iea.nl/fileadmin/user\\_upload/Publications/Electronic\\_versions/Elley\\_Impact\\_TIMSS-R.pdf](http://pub.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/Elley_Impact_TIMSS-R.pdf)
  26. Engel, L. C. (2015). Steering the National: Exploring the Education Policy Uses of PISA in Spain. *European Education*, 47(2), 100–116. <https://doi.org/10.1080/10564934.2015.1033913>
  27. Engel, L. C., & Frizzell, M. O. (2015). Competitive comparison and PISA bragging rights: sub-national uses of the OECD’s PISA in Canada and the USA. *Discourse: Studies in the Cultural Politics of Education*, 36(5), 665–682. <https://doi.org/10.1080/01596306.2015.1017446>
  28. Ertl, H. (2006). Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32(5), 619–634. <https://doi.org/10.1080/03054980600976320>
  29. Feniger, Y., Livneh, I., & Yogev, A. (2012). Globalisation and the politics of international tests: the case of Israel. *Comparative Education*, 48(3), 323–335. <https://doi.org/10.1080/03050068.2011.622539>
  30. Figazzolo, L. (2009). Impact of PISA 2006 on the education policy debate. *Education International*. Retrieved from <http://download.ei-ie.org/docs/IRISDocuments/Research%20Website%20Documents/2009-00036-01-E.pdf>
  31. Forestier, K., & Crossley, M. (2015). International education policy transfer – borrowing both ways: the Hong Kong and England experience. *Compare: A Journal of Comparative and International Education*, 45(5), 664–685. <https://doi.org/10.1080/03057925.2014.928508>
  32. Froese-Germain, B. (2010). The OECD, PISA and the Impacts on Educational Policy. *Canadian Teachers’ Federation (NJ1)*. Retrieved from <http://eric.ed.gov/?id=ED532562>
  33. Gaber, S., Cankar, G., Umek, L. M., & Tašner, V. (2012). The danger of inadequate conceptualisation in PISA for education policy. *Compare: A Journal of Comparative and International Education*, 42(4), 647–663. <https://doi.org/10.1080/03057925.2012.658275>
  34. Gil, L. V., Beltran, J. C. H., & Redondo, E. G. (2016). PISA as a political tool in Spain: Assessment instrument academic discourse and political reform. *European Education*, 48, 89–102. DOI: 10.1080/10564934.2016.1185685
  35. Gilmore, A. (2005). The impact of PIRLS (2001) and TIMSS (2003) in low-and middle-income countries. *Amsterdam: International Association for the Evaluation of Educational Achievement (IEA)*. Retrieved from <https://pdfs.semanticscholar.org/b40a/7b8f2170056448fc705f83064a3f90cadec4.pdf>
  36. Gorur, R. (2011). Policy as Assemblage. *European Educational Research Journal*, 10(4), 611–622. <https://doi.org/10.2304/eej.2011.10.4.611>
  37. Gorur, R. (2016). Seeing like PISA: A cautionary tale about the performativity of international assessments. *European Educational Research Journal*, 15(5), 598–616.
  38. Gove, A., & Cvelich, P. (2011). Early reading: Igniting education for all. A report by the early grade learning community of practice. RTI International. Retrieved from <http://eric.ed.gov/?id=ED520290>
  39. Gove, A., Habib, S., Piper, B., & Ralaingita, W. (2013). Classroom-up Policy Change: early reading and

- math assessments at work. *Research in Comparative and International Education*, 8(3), 373.  
<https://doi.org/10.2304/rcie.2013.8.3.373>
40. Greger, D. (2012). When PISA does not matter? The case of the Czech Republic and Germany. *Human Affairs*, 22(1). <https://doi.org/10.2478/s13374-012-0004-5>
  41. Grek, S. (2009). Governing by numbers: the PISA “effect” in Europe. *Journal of Education Policy*, 24(1), 23–37. <https://doi.org/10.1080/02680930802412669>
  42. Gür, B. S., Çelik, Z., & Özoğlu, M. (2012). Policy options for Turkey: a critique of the interpretation and utilization of PISA results in Turkey. *Journal of Education Policy*, 27(1), 1–21.  
<https://doi.org/10.1080/02680939.2011.595509>
  43. Haas, E., & Fischman, G. (2010). Nostalgia, entrepreneurship, and redemption: Understanding prototypes in higher education. *American Educational Research Journal*, 47(3), 532–562.
  44. Hopfenbeck, T. N., Lenkeit, J., El Masri, Y. et al. (2017). Lessons learned from PISA: a systematic review of peer-reviewed articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*. DOI: 10.1080/00313831.2016.1258726
  45. Hopkins, D., Pennock, D., Ritzen, J., Ahtaridou, E., & Zimmer, K. (2008). External evaluation of the policy impact of PISA. OECD doc. EDU/PISA/GB(2008)35/REV1. Paris, France: OECD.
  46. IEA. (2017). TIMSS & PIRLS [web site]. Retrieved from <http://timssandpirls.bc.edu/>
  47. Kirsch, I., Lennon, M., von Davier, M., Gonzalez, E., & Yamamoto, K. (2013). On the growing importance of international large-scale assessments. In von Davier, M., Gonzalez, E., Kirsch, I., Yamamoto, K. (Eds.), *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research* (pp 1-11). New York, NY: Springer.
  48. Korsnakova, P. (2013). Large-scale Comparative Studies of Educational Achievement: an impact case. *European Educational Research Journal*, 12(2), 251. <https://doi.org/10.2304/eeerj.2013.12.2.251>
  49. Lascoumes, P., & Le Gales, P. (2007). Introduction: Understanding public policy through its instruments? From the nature of instruments to the sociology of public policy instrumentation. *Governance*, 20(1), 1-21. doi:10.1111/j.1468-0491.2007.00342.x
  50. Lawn, M., & Grek, S. (2012, May). *Europeanizing education: Governing a new policy space*. Symposium Books Ltd.
  51. Learning Metrics Task Force [LMTF]. (2013). *Toward universal learning: Recommendations from the Learning Metrics Task Force*. Montreal and Washington, D. C.: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.
  52. Lenkeit, J., Chan, J., Hopfenbeck, T. N., Baird, J.-A. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, 16, 102–115.  
doi:10.1016/j.edurev.2015.10.002
  53. Lewis, S. (2016): Governing schooling through ‘what works’: the OECD’s PISA for Schools, *Journal of Education Policy*. DOI: 10.1080/02680939.2016.1252855
  54. Lewis, S., & Lingard, B. (2015). The multiple effects of international large-scale assessment on education policy and research. *Discourse*, 36(5). <http://dx.doi.org/10.1080/01596306.2015.1039765>
  55. Lingard, B., Rawolle, S., & Taylor 1, S. (2005). Globalizing policy sociology in education: working with Bourdieu. *Journal of education policy*, 20(6), 759-777.
  56. Lockheed, M. (2015, March). *Learning from PISA*. PowerPoint Presentation presented at the PISA for Development, International Advisory Group Meeting, Paris, France.
  57. Lockheed, M., T. Prokic-Bruer, & Shadrova, A. (2015). *The experience of middle-income countries participating in PISA 2000-2015*. The World Bank. Washington, D.C.: OECD Publishing, Paris.  
<http://dx.doi.org/10.1787/9789264246195-en>
  58. Lockheed, M. E., & Wagemaker, H. (2013). International Large-Scale Assessments: thermometers, whips or useful policy tools? *Research in Comparative and International Education*, 8(3), 296.  
<https://doi.org/10.2304/rcie.2013.8.3.296>
  59. Makuwa, D. K., & Maarse, J. (2013). The Impact of Large-Scale International Assessments: a case study of how the Ministry of Education in Namibia used SACMEQ assessments to improve learning outcomes. *Research in Comparative and International Education*, 8(3), 349.  
<https://doi.org/10.2304/rcie.2013.8.3.349>
  60. Martens, K., & Niemann, D. (2013). When Do Numbers Count? The Differential Impact of the PISA

- Rating and Ranking on Education Policy in Germany and the US. *German Politics*, 22(3), 314–332. <https://doi.org/10.1080/09644008.2013.794455>
61. Meyer, H.-D., & Benavot, A. (2013). PISA, power and policy the emergence of global educational governance. Oxford: Oxford Studies in Comparative Education, Symposium Books.
  62. Morgan, C. (2016). Tracing the sub-national effect of the OECD PISA: Integration into Canada's decentralized education system. *Global Social Policy*, 16(1), 47-67.
  63. OECD. (n.d.). Programme for International Student Assessment [web page]. Retrieved from <http://www.oecd.org/pisa/>
  64. OECD. (2011). Germany: Once weak international standing prompts strong nationwide reforms for rapid improvement. In *Lessons from PISA for the United States*. Paris, France: OECD Publishing. <http://dx.doi.org/10.1787/9789264096660-10-en>
  65. OECD. (2013). Policy implications of school management and practices. In *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices*. Paris, France: OECD Publishing. <http://dx.doi.org/10.1787/9789264201156-10-en>
  66. OECD. (2014). *Lessons from PISA for Korea*. OECD Publishing. <https://doi.org/10.1787/9789264190672-en>
  67. OECD. (2016, December 6). How does PISA help shape education reform? [Video file]. Retrieved from <https://www.youtube.com/watch?v=-xpOn0OzXEw>  
<https://www.youtube.com/watch?v=-xpOn0OzXEw>
  68. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown Publishing Group.
  69. Paine, L., Bloemke, S., & Aydarova, O. (2016). Teachers and teaching in the context of globalization. *Handbook of Research on Teaching*, 717-786.
  70. Perlman Robinson, J. (2011). A global compact on learning: Taking action on education in developing countries. Center for Universal Education at The Brookings Institution, Washington, DC.
  71. Phillips, D., & Ochs, K. (2003). Processes of policy borrowing in education: Some explanatory and analytical devices. *Comparative education*, 39(4), 451-461.
  72. Plisko, V. (2013). Participation in international large-scale assessments from a US perspective. *Research in Comparative and International Education*, 8, 326–334. doi:10.2304/rcie.2013.8.3.326
  73. Pons, X. (2016). Tracing the French policy PISA debate: A policy configuration approach. *European Educational Research Journal*, 15(5), 580–597.
  74. Powers, J. M., Fischman, G. E., & Berliner, D.C. (2016). Making the visible invisible: Willful ignorance of poverty and social inequalities in the research-policy nexus. *Review of Research in Education* 40(1), 744-776. DOI 10.3102/0091732X16663703
  75. Rautalin, M., & Alasuutari, P. (2009). The uses of the national PISA results by Finnish officials in central government. *Journal of Education Policy*, 24(5), 539-556.
  76. Rutkowski, D., & Delandshere, G. (2016). Causal inferences with large scale assessment data: using a validity framework. *Large-scale Assessments in Education*, 4(1), 6.
  77. Sahlberg, P. (2015). *Finnish Lessons 2.0*. New York, NY: Teachers College Press.
  78. Schleicher, A. (2009). Securing quality and equity in education: Lessons from PISA. *Prospects*, 39(3), 251-263.
  79. Schleicher, A., & Shewbridge, C. (2004). *What Makes School Systems Perform?: Seeing School Systems Through the Prism of PISA*. OECD Publishing.
  80. Sellar, S., & Lingard, B. (2013). Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. *Comparative Education*, 49(4), 464–485. <https://doi.org/10.1080/03050068.2013.770943>
  81. Shamatov, D. (2014). Education quality in Kyrgyzstan and the programme for international student assessment (PISA). In Napier, D. B. (Ed.), *Qualities of Education in a Globalized World* (pp. 43-62). New York, NY: Springer.
  82. Strietholt, R., Rosén, M., & Bos, W. (2013). A correction model for differences in the sample compositions: The degree of comparability as a function of age and schooling? *Large-Scale Assessments in Education*, 1(1), 1–20. doi:10.1186/2196-0739-1-1
  83. Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis

- debate. *Comparative Education*, 44(4), 387–407. <https://doi.org/10.1080/03050060802481413>
84. Takayama, K. (2010). Politics of externalization in reflexive times: Reinventing Japanese education reform discourses through “Finnish PISA success”. *Comparative Education Review*, 54(1).
  85. Teodoro, A., & Estrela, E. (2010). Curriculum policy in Portugal (1995–2007): global agendas and regional and national reconfigurations. *Journal of Curriculum Studies*, 42(5), 621–647.
  86. Tillmann, K. J., Dederig, K., Kneuper, D., Kuhlmann, C., & Nessel, I. (2008). *PISA als bildungspolitisches Ereignis: Fallstudien in vier Bundesländern*. Springer-Verlag.
  87. United Nations. (2016, June 3). Progress towards the sustainable development goals. Report to the Secretary-General, Economic and Social Council. New York, NY: Author. Retrieved from <http://unstats.un.org/sdgs/files/report/2016/secretary-general-sdg-report-2016--EN.pdf>
  88. UNESCO. (2016). Education 2030: Incheon declaration. Retrieved from: <http://en.unesco.org/world-education-forum-2015/incheon-declaration>
  89. Vega Gil, L., Beltrán, J. C. H., & Redondo, E. G. (2016). PISA as a Political Tool in Spain: Assessment Instrument, Academic Discourse and Political Reform. *European Education*, 48(2), 89–103. <https://doi.org/10.1080/10564934.2016.1185685>
  90. Verger, A. (2008). Measuring educational liberalisation. A global analysis of GATS. *Globalisation, Societies and Education*, 6(1), 13–31.
  91. Verger, A. (2014). Why do policy-makers adopt global education policies? Toward a research framework on the varying role of ideas in education reform. *Current Issues in Comparative Education*, 16(2), 14–29.
  92. Verger, A., Novelli, M., & Altinyelken, H. K. (2012). *Global education policy and international development: New agendas, issues and policies*. New York, NY: Bloomsbury Publishing.
  93. Vida, E. N.-J. (2009). The Reception of PISA in Hungary. Retrieved from [http://oitk.tatk.elte.hu/sites/default/files/files/o31\\_pisa\\_hungary.pdf](http://oitk.tatk.elte.hu/sites/default/files/files/o31_pisa_hungary.pdf)
  94. Waldow, F. (2009). What PISA Did and Did Not Do: Germany after the “PISA-shock.” *European Educational Research Journal*, 8(3), 476–483. <https://doi.org/10.2304/eej.2009.8.3.476>
  95. Winthrop, R., Anderson, K., & Cruzalegui, I. (2015). A review of policy debates around learning in the post-2015 education and development agenda. *International Journal of Educational Development*, 40, 297–307.
  96. Winthrop, R., & Simons, K. A. (2013). Can International Large-Scale Assessments Inform a Global Learning Goal? Insights from the Learning Metrics Task Force. *Research in Comparative and International Education*, 8(3), 279–295.
  97. Yemini, M., & Gordon, N. (2015). Media representations of national and international standardized testing in the Israeli education system. *Discourse: Studies in the Cultural Politics of Education*, 1–15. <https://doi.org/10.1080/01596306.2015.1105786>
  98. Young, H. (2015, November 25). What can we learn from the great German school turnaround? *The Guardian*.



## Appendix

### Survey A: ILSA/GLM Experts

1. Thinking about your experiences working with educational stakeholders in different national contexts (politicians, policymakers, unions, teachers, media, etc.), which of the following ILSAs/GLMs are they concerned about? (check all that apply)
  - LAMP
  - PIAAC
  - PIRLS
  - PISA
  - STEP
  - TIMSS
  - WEI-SPS
  - Other
2. From your experience, how do the stakeholders you work with use ILSAs/GLMs in national education reform efforts?
3. Based on your experience, to what extent do you think ILSAs/GLMs are contributing to or hindering national education reform efforts?
4. Based on your experience, to what extent do you think ILSAs/GLMs are contributing to or hindering global education reform efforts?
5. What country and/or regions does the majority of your work on ILSAs/GLMs focus on?
6. Could you recommend any other colleagues who may like to share their perspectives about the use and influence of ILSAs/GLMs? Please list any names below.

### Survey B: ILSA/GLM Educators/Policymakers

1. What is your primary role in working with ILSAs/GLMs? (check one)
  - Researcher
  - Policymaker
  - Union representative
  - Student
  - Teacher
  - Media
  - Other
2. How often do you work with educational stakeholders (politicians, policymakers, unions, teachers, media, etc.) in different national contexts on ILSAs/GLMs?
  - Frequently
  - Sometimes
  - Never
3. Thinking about your experiences working with educational stakeholders in different national contexts (politicians, policymakers, unions, teachers, media, etc.), which of the following ILSAs/GLMs are they concerned about? (check all that apply)
  - LAMP

- PIAAC
  - PIRLS
  - PISA
  - STEP
  - TIMSS
  - WEI-SPS
  - Other
4. From your experience, how do the stakeholders you work with use ILSAs/GLMs in national education reform efforts?
  5. Based on your experience, to what extent do you think ILSAs/GLMs are contributing to or hindering national education reform efforts?
  6. Based on your experience, to what extent do you think ILSAs/GLMs are contributing to or hindering global education reform efforts?
  7. How have you used ILSAs/GLMs in your own work?
  8. What country and/or regions does the majority of your work on ILSAs/GLMs focus on?

**The Center for Advanced Studies in Global Education**

P.O. Box 871611, Tempe, AZ 85287-1611

CASGE@asu.edu (480)

727-5346

education.asu.edu/casge



@asucasge



ARIZONA STATE UNIVERSITY

