

## Pakistan's First-Ever Participation in an International Large-Scale Assessment (TIMSS): Critique and Implications

Sohail Ahmad  
Aga Khan University

Sherwin Rodrigues  
Aga Khan University

Sadia Muzaffar Bhutta  
Aga Khan University

### Abstract

*The International Large-Scale Assessment (ILSA) is a rapidly growing field in education which has gained considerable attention of stakeholders across the globe. Historically, ILSA emerged from the developed context which has drifted to the developing contexts in a short span of time due to globalisation. Pakistan has participated in ILSA (i.e., TIMSS) for the first time in the 2019 cycle. In light of the global critique on ILSA, this paper presents a critical analysis on Pakistan's participation in TIMSS by raising questions embedded in contextual realities. The discussion adds to the understanding of ILSA in terms of historical developments, theoretical underpinnings towards participation in ILSA, and the general as well as context-specific critique on ILSA. The paper ends with the argument in favour of strengthening national LSA instead of relying only on ILSA.*

**Keywords:** International large-scale assessment, TIMSS, students' performance, Pakistan.

### Introduction

The International Large-Scale Assessment (ILSA) is a rapidly growing field in education which has gained due attention from stakeholders including policy makers, practitioners, researchers, politicians, media, and the general public

(Addey et al., 2017; Baroutsis & Lingard, 2021). Some well-known ILSAs are – but not limited to: Trends in International Mathematics and Science Study (TIMSS); Progress in International Reading Literacy Study (PIRLS); and Programme for International Student Assessment (PISA). ILSA was initiated in high income economies; however, within a short span of time it drifted to the middle and low-income economies (Kamens & Benavot, 2011). This rapid expansion of ILSA and excessive responses from researchers, confirms its influential role in educational developments both globally and nationally. Generally, literature on ILSA comprises two categories. First, the proponents of ILSA advocate the benefits of ILSA in terms of standardisation of education, accountability in education, evidence-based policies, increasing public expenditure on education, and learning from best-performing countries (e.g., Addey et al., 2017; Cox & Meckes, 2016; Ho, 2016). Secondly, several studies critique ILSA by highlighting numerous psychometric and methodological issues, misuse of ILSA results (inappropriate interpretation), governance by numbers and ignoring contextual realities while drawing conclusions (e.g., Emler et al., 2019; Gorur, 2017; Komatsu & Rappleye, 2017; Komatsu & Rappleye, 2021). Arguably, despite such limitations, it is hard to deny the influence of ILSA in the arena of international education, especially during globalisation.

With this backdrop, the current study aims to present a critical analysis on Pakistan's first-ever participation in ILSA (i.e., TIMSS). The paper begins with a brief introduction to historical developments and theoretical approaches to and benefits of participation in ILSA, followed by a general critique. After generic discussion on ILSA, the paper presents a critical analysis of Pakistan's participation in TIMSS 2019, by raising questions in light of the contextual realities and potential of ILSA. In conclusion, some viable options to maximise the benefits of ILSA are proposed.

### **Historical Development of ILSA: TIMSS and PISA**

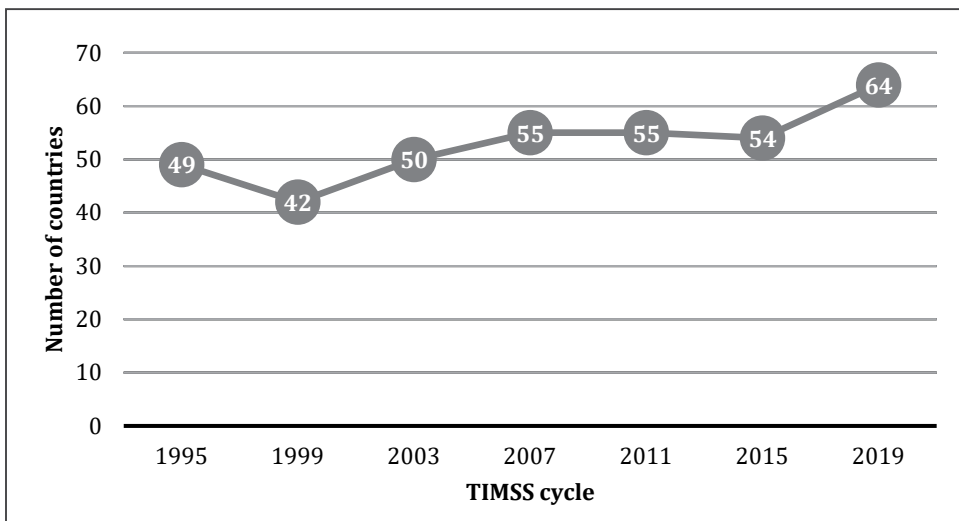
ILSA entails “Studies in which both achievement of certain age/grade in one or more subjects is compared across education systems and effects of contextual factors at system, school, classroom and students' level on achievement” (Bos, 2002, p.2). The emergence of ILSA can be traced to the late 1950s when a group of researchers at UNESCO discussed conducting assessment of students' achievement in several countries (Hernández-Torrano & Courtney, 2021). This discussion resulted in the establishment of the International Association for the Evaluation of Education

Achievement (IEA), which conducted the First International Mathematics Study (FMIS) in 1964 to compare the performance of students among 10 participating countries (Rutkowski et al., 2014). In the 1990s, TIMSS and PISA were initiated by the International Association for the Evaluation of Educational Achievement (IEA) and Organisation for Economic Co-operation and Development (OECD), respectively.

TIMSS was initiated by the IEA in 1995 with the aim to provide international comparative evidence of grades four and eight students' achievements in Mathematics and Science. As depicted in figure 1, the number of countries participating in TIMSS has an upward trend where the highest participation is visible in the most recent cycle.

**Figure 1**

*Trend in countries participation in TIMSS (Source, IEA, 1995)*

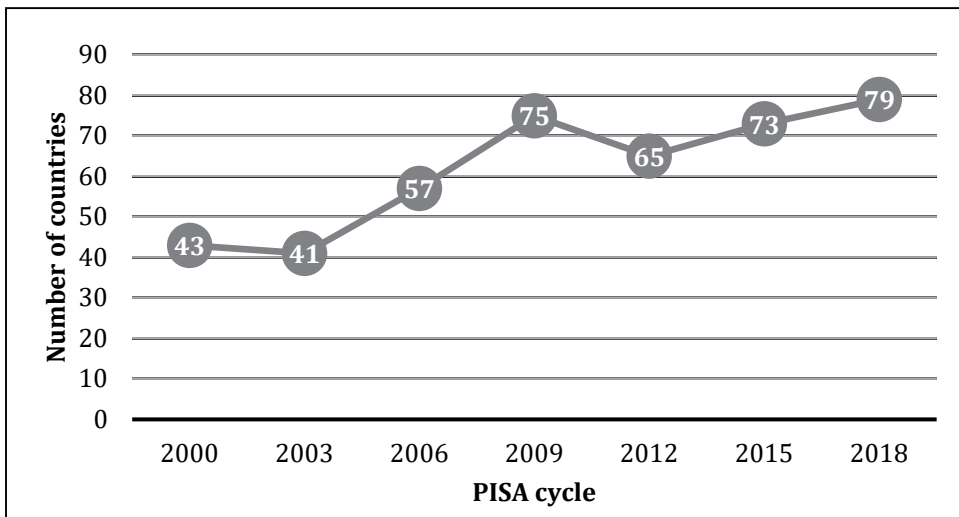


TIMSS evaluate students' achievement in the two target subjects in terms of knowledge, concepts, processes, skills, and attitudes in a cycle every four years (Mohammadpour & Shekarchizadeh, 2015; Mullis et al., 2020). Further, it compares the results of students' performance across participating countries with the intention to bring improvement in Science and Mathematics education globally (Atar et al., 2021; Elliott et al., 2018; Sabaha & Hammouri, 2010).

Similarly, in 1999 the OECD launched PISA with the aim to assess 15-year-old children's competencies in a three-year cycle, in different domains such as reading, mathematical and scientific literacy, and global competency (OECD, 2019). Since its inception, a consistent increase in participating countries is visible, as shown in figure 2. Some countries withdrew in the 2012 cycle; however, PISA succeeded in following an upward trajectory in the later cycles. The highest participation can be observed in the most recent cycle. Arguably, the growing participation of countries could be due to the influential role of PISA in promoting global educational reforms and accountability, and globalisation (Sellar & Lingard, 2014).

## Figure 2

*Trend of the number of participating countries in PISA*



Both TIMSS and PISA assess students' achievement along with the associated factors and compare results across participating countries. TIMSS and PISA are instrumental for cross-country comparisons in students' performance, teacher education and educational reforms (OECD, 2019; Mullis et al., 2020; Tonga et al., 2022). In addition, the periodic nature of the two assessments is also useful in making within-country comparisons longitudinally. As discussed, though TIMSS and PISA have recently emerged, both have observed continuous advancement in terms of gaining attention (e.g., from politicians, policymakers, practitioners, researchers, and global educational developments), methodologically (e.g., advance statistical analysis and publication), and technological integration

(e.g., synchronized administration, artificial intelligence in test administration).

### **What Motivates Countries to Partake in ILSA?**

There are numerous motivating factors that contribute to the choice of participating in ILSA. Apparent justifications for participating in any ILSA include, but are not limited to, evidence-based policy formulation, competing with the globalised world, bringing accountability through standardised results, improving curriculum, teaching in the light of ILSA results, and capacity building of local staff working on assessment (Addey & Sellar, 2018). The factors that motivate countries to participate in ILSA can be broadly explained in a four-dimensional framework which includes: i) the rationale choice model; ii) the policy diffusion model, iii) the macro-dissatisfaction perspective; and, iv) the financial aid model (Hernández-Torrano & Courtney, 2021 ; Kijima, 2010).

In the rationale choice model, countries participate in ILSA with a motive to make evidence-based policies and support their decisions politically. More specifically, the government of the participating country attempts to use ILSA results as a reference point to devise new and/or revise the existing education policies and meet their political interests. The government of Turkey, for example, legitimized their decision of curriculum revision using the results of PISA 2003 and 2006 (Gür et al., 2012). Similarly, governments in other countries, such as France and Portugal, also used ILSA results to justify their decisions of educational reforms and legitimize the development of new reforms (Afono & Costa, 2009; Lundahl & Serder, 2020; Pons, 2011). Apart from internal benefits, countries also aim to demonstrate their visibility in the international arena and sustain their reputation through better performance in ILSA.

In the policy diffusion model, countries participate in ILSA to explore and borrow effective educational practices with the intention of uplifting the quality of education. Based on ILSA results, countries can borrow best practices from those demonstrating higher performance. This borrowing and lending of practices can be between developed and developing countries, or a developed country may adapt practices from others (Chung, 2016; Mohamed & Morris, 2021; Pettersson et al., 2017). It is argued that secondary analysis of ILSA data is more beneficial for the countries pertaining to their participation with the Rationale choice or Policy diffusion models (Torney-Purta & Amadeo, 2013), as it would provide more evidence

for justifying their decisions of policy revision or borrowing. However, Hernández-Torrano and Courtney (2021) state that secondary analysis can be beneficial for all participating countries, regardless of their initial motivational orientation, as both TIMSS and PISA initial results could not provide a comprehensive picture of any education system. Having said that, secondary analysis may be less influential in terms of evoking political and media interest as compared to results dissemination of primary data (Wiseman, 2013). The reasons could be that secondary analyses are usually conducted later, and may not capture the attention of policy makers (Torney-Purta & Amadeo, 2013).

Countries with macro-dissatisfaction motivational orientation, participate in ILSA to capture attention of the international community towards educational crises and to propel focus towards resolutions. Moreover, in the financial aid model, motivation of countries may be linked with receiving more financial aid from international donors for uplifting quality and equity issues in education. It has been found that countries that participate in ILSA tend to receive 37% more foreign aid than those that do not (Kijima, 2010). Besides, it has been reported that low or middle-income countries are more likely to participate in ILSA when they are sponsored by donor organisations. Thus finance seems to be a key decisive factor for participating in ILSA for most of the low and middle-income countries (Lockheed et al., 2015).

It is worthwhile to note that different countries may have different orientations towards participating in ILSA. The rising influence of globalisation on education is rooted in almost all the four models. What motivated Pakistan to appear in the TIMSS 2019 cycle? In the latter section, speculations of Pakistan's participation in light of these four models will be made.

### **Critiquing ILSA: Benefits and Limitations**

Arguably, the borrowing and lending of education policies and practices of the best performing countries can be attributed to the strength of ILSA in the contemporary education policy discourse. The primary focus of ILSA (e.g., TIMSS and PISA) is to measure and compare educational success and access in order to make informed decisions nationally and internationally. Learning outcomes are gaining increased attention in almost all global education agendas. Sustainable Development Goal (SDG) – 4, for example, calls for robust efforts on students'

learning outcomes as half of the targets (for SDG 4) highlight learning skills and outcomes. It is argued that ILSA can be instrumental in determining and monitoring the global educational target (e.g., Education for All, SDG 4- Quality Education) where evidence of students' learning outcomes can be used for determining the targets, devising strategies, evaluating progress, and ensuring accountability (Addey & Seller, 2019). Additionally, wider media attention that ILSA receives, specifically after releasing the results, has a profound influence on the education discourse globally (Breakspear, 2012; Hopfenbeck et al., 2018). However, a serious caution should be noted while making any decision in the context of global reforms because these results are not necessarily appropriate for all the countries in the world, as the current participation rate is almost one-third of the total countries.

At the national level, ILSA results serve as a baseline for legitimizing many educational reforms. Studies have reported changes being initiated because of ILSA results. For example, the 'PISA shock' led many countries to revisit their education policies and practices in order to improve their position in the international ranking (e.g., Afono & Costa, 2009; Lundahl & Serder, 2020; Ringarp, 2016; Pons, 2011). Moreover, it is argued that the cross-national comparison of students' performance through a standardised approach allows countries to share and learn from the experiences of each other in order to improve the learning outcomes of students (Cresswell et al., 2015). In addition, ILSA can also be instrumental for countries to plan targeted interventions using results. Table 1 presents examples of targeted interventions by countries based on ILSA results (e.g., Afonso & Costa 2009; Bialecki et al., 2017; Choi & Jerrim 2016; Paine et al., 2016).

**Table 1***Resources Invested on the Basis of Cross-National Assessments**Montoya & Coleman (2018)*

<b>Area of resource investment</b>	<b>Examples of actions taken by countries</b>
Teachers, training, and professional development	In-service professional development programmes for teachers and leaders
	Technology integration in classroom teaching
	Attracting bright candidates to teaching profession through incentives
	Improving teachers' pedagogical skills and teaching literacy
	Increased incentives for in-service teachers
	Increasing budget for education for class size reduction and infrastructure development
Education funding	Initiative to strengthen literacy development
	Funding programmes to promote reading and literacy
	Donors helping to stimulate a policy response
	Interventions based on the findings, which are also used to influence policy dialogue and action
Education materials and time resources	An increase in classroom instruction time
	Recruitment of more teachers
	Use of assessment to inform curriculum and instruction
	Strategies for quality teaching and assessment
	Influencing the national educational programme
	Policy and practice influence

Apart from the above, there are several other benefits of ILSA for participating countries as well as researchers around the world, which includes: curriculum reforms; utilizing strategies; enhancing access and equity; providing access to large-scale representative data at regional and international level; international comparison; and indicators to monitor and evaluate educational processes (Tobin et al., 2016; UNESCO, 2018). Furthermore, researchers conducting secondary



analysis on the publicly available ILSA data have made a remarkable addition to publications internationally (Hernández-Torrano & Courtney, 2021).

The benefits of ILSA depend on the credibility and validity of data and results and how the same are appropriately used in policy development (Goldstein & Moss, 2014). It is argued that ILSA (e.g., TIMSS and PISA) follow robust protocols for tools development, data collection, and analysis to generate reliable results (Leung, 2014). That said, numerous studies have noted limitations in such ILSA collecting data across a wide range of geopolitical, socioeconomic, cultural, and linguistic countries (e.g., Emler et al., 2019; Komatsu & Rappleye, 2017; Komatsu & Rappleye, 2021). Therefore, one needs to be cognizant of these limitations prior to making any response to ILSA results. The primary focus of ILSA is on learning outcomes to gauge the quality of education of the participating countries. However, there are many other aspects (e.g., teaching quality, resources) that can also play a significant role in defining or ranking the quality of education. Furthermore, ILSA employs a cross-sectional survey across the participating countries which have different cultures and languages. Since ILSA translates tests into various languages, there is always a possibility of cross-cultural translation error which may influence students' responses in the test or questionnaire (e.g., He & Kubacka, 2015; Solano-Flores et al., 2013). Additionally, the cultural differences can also influence students' responses on test items, as some items may have different meanings in various cultures.

Moreover, another limitation is of the 'exclusive' population and sampling. The definition of the target population is very specific, hardly provides space for others. For example, the target population of TIMSS is all grades 4 and 8 school-going children in the participating country. Many countries might have different types of education systems, but TIMSS focuses only on formal schools. Similarly, criticism has been directed at the sampling frame used for ILSA. Usually, the sampling frame cannot cover all eligible individuals in ILSA. Many eligible individuals are excluded because of small school size or schools located in far flung areas which adds to increased costs. In some cases, the target population is excluded because of representing a minority language group or distinct education system (e.g., Madrassahs) which is usually referred to 'reduced coverage'. Furthermore, another limitation of ILSA is related to a likely inappropriate interpretation of its results where interpretations mainly involve establishing cause and effect. For example,

studies have sought to identify the impact of ILSAs on policy as well as economic growth (e.g., Breakspear, 2012; Fischman et al., 2019; Komatsu & Rappleye, 2017) but establishing a causal relationship between ILSA data and educational reforms is problematic because policy processes are complex and rarely driven by a single causal factor (Paine et al., 2016).

### **Pakistan's Participation in TIMSS: Raising Questions**

As discussed above, Pakistan for the first time participated in the TIMSS 2019 cycle. The results of grade 4 students' performance in both Science and Mathematics are a serious cause of concern for the stakeholders. However, before making any decision based on these results, one needs to be cognizant of the context as well as findings from other sources. In the following section, we have attempted to raise and respond to questions specific to TIMSS, but the discussion has implications for other general large-scale assessment studies as well.

### **Is the national score of TIMSS truly representative of the country's performance?**

There is a growing critique on the lack of inclusiveness in the population and sampling of ILSA studies. TIMSS target population is all grades 4 and 8 students of the participating country. Since Pakistan has participated only for grade 4 the target population is all school going grade 4 students. Understandably, the national score must be the representation of the target population. However, a deeper look into the sampling and contextual realities poses many questions related to appropriate representation. The education system of Pakistan is very complex as various types of institutions are operating at the same time which include public, private and religious schools (Madrassahs). While developing a sampling frame, TIMSS recruited students from two strata (i.e., public, and private) while completely ignoring religious schools. One may question if religious schools meet the definition of the target population for TIMSS or not. Currently, many madrassahs have initiated an integrated education system where they provide formal schooling along with religious education to children. Interestingly, this trend is growing rapidly where madrassahs have started the integration of formal schooling with religious subjects. Approximately, there are more than 280 madrassahs only in Punjab that provide an integrated curriculum (Bhatt, 2020). Nevertheless, students in some of these madrassahs appear in the international Cambridge O and A level examination. The

TIMSS sample from Pakistan does not represent students from Madrassah schools which limits generalisation of the results to the target population. Nevertheless, simply recruiting the sample from private schools without providing demographic information also exposes the limitation of the result of TIMSS. Region wise sampling of private schools is more skewed towards the province of Punjab which is already performing better in the national LSA. In other words, of the total 32 selected private schools, an overwhelming majority (n=26) were from Punjab as shown in table 2.

**Table 2**

*Sample representation from Pakistan in TIMSS 2019*

<b>Explicit Strata</b>	<b>Total Sample schools</b>
Public – Smaller Regions	20
Public - Khyber Pakhtunkhwa	24
Public - Punjab	50
Public - Sindh	22
Private - Punjab	26
Private - All Other Regions	8

Within the private school system there are three further categories – elite, mediocre and mushroom street schools. There are huge differences in terms of students' socioeconomic status, curriculum, quality of education and resources, among these three types of schools. Apparently, almost all national studies revealed better performance of students studying in private schools; however, comparison within private schools reveals no difference between mushroom street private schools and public schools (Bhutta & Rizvi, 2022). This differences in performance may also have implications for TIMSS results. Therefore, a clear description of the sample would help to understand whether the national scale mean score is truly representative of the target population. This would help in making well-informed interventions and reforms. Despite the fact that the current score of Pakistan is not encouraging, appropriate representation would have unfolded the true picture (in favour or against) of the current gross national scores. Should we expect more pathetic conditions or hope for better outcomes in true representation?

### **Are the TIMSS results comprehensive to understand contextual realities?**

The results of (Mullis et al., 2020) revealed that Pakistan stands second from the bottom, in terms of students' performance in science and mathematics, among the 64 participating countries in TIMSS. Within country comparison by subjects reveals that students' performance in mathematics is relatively better than science in grade 4. What would be the possible reasons for these low scores? Since TIMSS collected data on various aspects, a deeper insight would help to understand the pattern or raise questions. Students' self-reported data about liking/disliking target subjects revealed that an overwhelming majority of the students 'like learning science and mathematics'. There is general agreement in the education fraternity that students' attitude towards science and mathematics significantly contributes to their performance (e.g., Berger et al., 2020). In the case of Pakistan, despite the strong attitude, students' scores are comparatively low in both science and mathematics. This discrepancy raises questions about the validity and reliability of self-reported data particularly in the context of Pakistan. Is there a problem with self-reported data? Or there may be certain other contextual factors that would be more dominant in influencing students' performance in the target subjects. Since TIMSS claims for a standardised approach to data collection and analysis, a visible limitation of TIMSS (i.e., lack of capacity to capture contextual uniqueness) can be found in this case.

Taking instructional time as another example, TIMSS reports that Pakistan is among the few top countries where more instructional time is given to science and mathematics. Though TIMSS mentioned limited responses on this data, the results still need to be critically examined. Generally, research suggests that more instructional time would lead to better learning outcomes; however, according to TIMSS results it seems to be invalid. Despite more instructional time reported by Pakistan, the students' scores are very low. Here a question arises that how did the teachers and students utilise the instructional time? Since, due to limitations of ILSA – lack of contextual and deeper insights –TIMSS could not add further into this explanation. But national LSA can help in understanding instructional time. According to a recent nationwide study, the 'active' instructional time is 23 minutes and 24 minutes per day for both science and mathematics, respectively (Bhutta & Rizvi, 2022). In other words, active instructional time per year is almost 90 hours for both science and mathematics in Pakistan, which is far below what has been reported by TIMSS (i.e., 157 and 139 hours/year for mathematics and science, respectively).

### Does ILSA contribute to the quality enhancement of students' learning?

There is a consistent argument in the literature that LSA, whether international or national, has a greater potential of promoting accountability in education, making informed policy decisions, and improving the quality and access to education (Emler et al., 2019). In the context of Pakistan, the outcomes of LSA efforts hardly resonate with the above-mentioned line of argument. In Pakistan, the first national LSA study was conducted in 2005 by the National Education Assessment System (NEAS). To date, NEAS has conducted five studies comprising a national representative sample of almost 15,000 students of grade 4 and 8 across Pakistan. A summary of the results is presented in table 3.

**Table 3**

*Students' Score in National LSA Conducted by NEAS*

Year	Grade 4		Grade 8	
	Mathematics	Science	Mathematics	Science
2005	421	-	-	-
2006 / 2007	404	467	457	-
2008	369	-	-	467
2014	-	433	461	-
2016	484	-	-	478

Though the results as presented in table 2 show an upward trend of improvement in students' score, the country could not reach the minimum standard of proficiency of 500 mean scores in the last sixteen years. Theoretically, in light of these results the government should have introduced targeted interventions in order to raise the quality of students' learning (Fischman et al., 2019). One may argue that the outcomes of any education reform take time and the upward trend is manifestation of the journey towards improvement. However, two recent studies, one International – TIMSS and the other national – Bhutta & Rizvi (2022), confirm almost little or no improvement in students' scores as compared to the first National study in 2005. According to TIMSS 2019 results, the mean scale score of Pakistan in mathematics (M=328) and science (M=290) was far below what the NEAS results have reported. Here a question may arise whether the quality of students' learning is improving or declining in the country? Perhaps, others would attempt to

identify issues with TIMSS tests in order to justify the upward trend in the national scores reported by NEAS. However, TIMSS results can also be validated by another nationwide study conducted by independent researchers (Bhutta & Rizvi, 2022). This study recruited more than 15,000 elementary grade students from 153 public and private schools across six regions of Pakistan. The researchers used Standardized Achievement Tests (SATs) for both science and mathematics along with a classroom observation scale to examine the quality of teaching inside the classroom. The results revealed that only one percent of the students scored more than 80% in these tests whereas the score of an overwhelming majority of students was below 33%. The study also reported the 'weak' quality of pedagogical practices in both the science and mathematics classroom where teaching in the classroom is primarily teacher-centred and mainly encouraging rote-based learning (Bhutta & Rizvi, 2022). The results of TIMSS and the national study confirm the low performance of students in the core school subjects i.e., science and mathematics. The discussion also raises questions about the outcomes of the huge investment of public money on LSA in the country. It is too early to say whether these results would convince the key stakeholders to plan targeted intervention or not. However, the above discussion contradicts popular opinion that LSA has the potential to bring improvement in the quality of education in any context.

### **Does ILSA contribute to educational reforms?**

Generally, it is argued that countries use ILSA results for introducing educational reforms including curriculum revision (e.g., Chung, 2016; Gur et al., 2012; Lundahl & Serder, 2020). There is minimal evidence in the context of Pakistan where the government introduced educational reforms based on TIMSS 2019 results. One of the massive education reforms of the curriculum – Single National Curriculum (SNC) – was initiated in 2018, a year before the announcement of TIMSS results. The reform of SNC was embedded in the manifesto of the currently governing political party. The ministry of education hardly rationalised SNC reform in light of TIMSS or NEAS results. On the other hand, efforts have been made by curriculum developers to link the SNC content with the TIMSS framework. Similarly, teacher training manuals were also designed before the TIMSS result and training has commenced.

Moreover, after the 18<sup>th</sup> amendment in the Constitution of Pakistan, education is now a provincial subject. Some of the provinces have recently developed five-

year action plans; however, they hardly used TIMSS results as a point of reference for targeted policy goals and interventions. One of the possible reasons could be that TIMSS could not provide within Pakistan comparison. This does not mean that provinces have limited evidence related to students' learning outcomes. Over the last decade all the provinces have witnessed a major focus on the LSA assessment culture in the country. For example, apart from NEAS, many of the provinces are also conducting provincial LSAs while other independent studies have also been conducted (e.g., ASER, 2018; Bhutta & Rizvi, 2022). This shows fragmented efforts for assessing students' performance which may have negative consequences on students as well. Despite these extensive LSA studies, Pakistan's performance is not improving which may be due to the reasons that our policy and reforms are not well-informed by the results of LSA.

### **Conclusion and Recommendations**

ILSA will remain an enduring and powerful feature of the global educational landscape for the foreseeable future. What we currently need are productive, creative and generative ways of dealing with ILSAs (Komatsua & Rappleye, 2021).

Due to the ongoing influence of globalisation, it would be challenging for countries to detach from ILSA. Despite its limitations there is a need to negotiate viable options to get the benefits from an investment in ILSA's participation. This paper follows the same line of argument that ILSA results will always have certain limitations which may not provide a comprehensive understanding of the results in context. Therefore, national LSA studies need to be continued to find contextually rich explanations for ILSA results which would help in devising effective educational reforms and monitoring progress. Nevertheless, national LSA should be robust enough to capture diverse factors related to students' performance.

The above discussion provides some critical insights into TIMSS results in the context of Pakistan. The questions posed, related to TIMSS Pakistan, and the discussion is consistent with the conclusion drawn by Fischman et al. (2019), which demonstrates contradictions between the perceived promises of LSA and its real contribution to education reforms. For instance, ILSA is conceived to be used as an instrument for monitoring educational progress: i) Countries participate in ILSA. ii) The government and stakeholders analyse the results and compare these with other countries. iii) The government changes policies and practices with the intention to

improve education by introducing reforms in curriculum, teaching and learning, and strengthening accountability. However, reality contradicts the perceptions where countries have a predetermined education policy agenda, participate in ILSA and use the results to legitimise their agenda. In order to reap the maximum benefits of ILSA, there is a pressing need for the government to focus on evidence-based yet sustainable educational reforms. In addition, the ultimate aim of participation in ILSA is to bring improvement in educational practices, therefore, the decision of participation must be well conceived in terms of cost-benefits analyses with persistent commitment for improvement. Furthermore, the results of ILSA should not be treated only as a 'diagnostic' source but it should be followed by proper 'treatment' through targeted interventions in education. Nevertheless, the first ever participation of Pakistan in TIMSS should be taken as a baseline for the newly introduced educational reforms (e.g., Single National Curriculum) in order to examine their efficacy.

### References

- Addey, C., & Sellar, S. (2019). Rationales for (non) participation in international large-scale learning assessments. *Education Research and Foresight : UNESCO Working paper*.
- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare: A Journal of Comparative and International Education*, 47(3), 434-452. <https://doi.org/10.1080/03057925.2017.1301399>
- Afonso, N., & Costa, E. (2009). The influence of the Programme for International Student Assessment (PISA) on policy decision in Portugal: The education policies of the 17th Portuguese Constitutional Government. *Sísifo. Educational Sciences Journal*, 10, pp. 53-64. <http://sisifo.fpce.ul.pt>
- ASER. (2018). Annual status of education report (ASER) Pakistan national (urban). *ASER Pakistan*.
- Atar, B., Atalay, K. & Kibrislioglu, U. (2021). Comparability of TIMSS 2015 mathematics test scores across country subgroups. *The Journal of Experimental Education*, 1-19. <https://doi.org/10.1080/00220973.2021.1913978>
- Baroutsis, A. & Lingard, B., (2021). A methodological approach to the analysis of PISA microblogs: Social media during the release of the PISA 2015 results, *Journal of Education Policy*, DOI: 10.1080/02680939.2021.1937706
- Berger, N., Mackenzie, E., & Holmes, K. (2020). Positive attitudes towards mathematics



- and science are mutually beneficial for student achievement: A latent profile analysis of TIMSS 2015. *The Australian Educational Researcher*, 47(3), 409-444. <https://doi.org/10.1007/s13384-020-00379-8>
- Bhutt, K. (2020, February 18). *Pakistani madrassas: Enduring through challenges*. <https://www.aa.com.tr/en/asia-pacific/pakistani-madrassas-enduring-through-challenges/1737480>
- Bhutta, S. M., & Rizvi, N. F. (2022). Assessing teachers' pedagogical practices and students' learning outcomes in science and mathematics across primary and secondary school level: A nationwide study. Institute of Educational Development. Agha Khan University. Karachi: Pakistan.
- Białecki, I., Jakubowski, M., & Wiśniewski, J. (2017). Education policy in Poland: The impact of PISA (and other international studies). *European Journal of Education*, 52(2), 167-174.
- Bos, K. (2002). Benefits and limitations of large-scale international comparative achievement studies. The Case of IEA's TIMSS Study. Enschede: OCTO, University of Twente. ISBN: 9036517397
- Breakspear, S. (2012). The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance, OECD Education Working Papers, No. 71, OECD Publishing, Paris, <https://doi.org/10.1787/5k9fdfqffr28-en>
- 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.
- Chung, J. (2016). The (mis)use of the Finnish teacher education model: 'Policy-based evidence-making'? *Educational Research*, 58(2), 207-219. <https://doi.org/10.1080/00131881.2016.1167485>
- Cox, C., & Meckes, L. (2016). International large-scale assessment studies and educational policy-making in Chile: Contexts and dimensions of influence. *Research Papers in Education*, 31(5), 502-515. <https://doi.org/10.1080/02671522.2016.1225349>
- Cresswell, J., Schwantner, U., & Waters, C. (2015). A review of international large-scale assessments in education: Assessing component skills and collecting contextual data. PISA: The World Bank. Washington, D.C. OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264248373-en>
- Elliot, J., Stankov, L., Lee, J., & Beckmann, J. F. (2018). What did PISA and TIMSS ever do for us? The potential of large-scale datasets for understanding and improving educational practice. *Comparative Education*, 1-23. <https://doi.org/10.1080/03050068.2018.1545386>

- Emler, T. E., Zhao, Y., Deng, J., Yin, D., & Wang, Y. (2019). Side effects of large-scale assessments in education. *ECNU Review of Education*, 2(3), 279-296. <https://doi.org/10.1177/2096531119878964>
- Fischman, G. E., Topper, A. M., Silova, I., Goebel, J., & Holloway, J. L. (2019). Examining the influence of international large-scale assessments on national education policies. *Journal of Education Policy*, 34(4), 470–499. <https://doi.org/10.1080/2680939.2018.1460493>
- Goldstein, H. & Moss, G. (2014). Knowledge and numbers in education, *Comparative Education*, 50:3, 259-265. <https://doi.org/10.1080/14681366.2014.926138>
- Gorur, R. (2017). Towards productive critique of large-scale comparisons in education. *Critical Studies in Education*, 58(3), 341–355. <https://doi.org/10.1080/17508487.2017.1327876>
- Gur, 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>
- He, J., & Kubacka, K. (2015). Data comparability in the teaching and learning international survey (TALIS) 2008 and 2013. OECD Education Working Paper No 124. Paris: OECD Publishing. <https://doi.org/10.1787/19939019>
- Hernández-Torrano, D., & Courtney, M. G. R. (2021). Modern international large-scale assessment in education: An integrative review and mapping of the literature. *Large-Scale Assessments in Education*, 9(1), 17. <https://doi.org/10.1186/s40536-021-00109-1>
- Ho, E. C. H. (2016). The use of large-scale assessment (PISA): insights for policy and practice in the case of Hong Kong. *Research Papers in Education*, 31, 5516–5528.
- Hopfenbeck, T. N., Lenkeit, J., El-Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333-353.
- Kamens, D. H., & Benavot, A. (2011). National, regional, and international learning assessments: Trends among developing countries, 1960–2009. *Globalisation, Societies and Education*, 9(2), 285-300. <https://doi.org/10.1080/14767724.2011.577337>
- Kijima, R. (2010). Why participate? Cross-National assessments and foreign aid to education. In Wiseman, A.W. (Ed.) *The Impact of International Achievement Studies on National Education Policymaking (International Perspectives on Education and Society, Vol. 13)*, Emerald Group Publishing Limited, Bingley, pp.

- 35-61. [https://doi.org/10.1108/S1479-3679\(2010\)0000013005](https://doi.org/10.1108/S1479-3679(2010)0000013005)
- Komatsu, H., & Rappleye, J. (2017). A new global policy regime founded on invalid statistics. Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, 53(2), 166–191. <https://doi.org/10.1080/03050068.2017.1300008>
- Komatsu, H., & Rappleye, J. (2021). Rearticulating PISA. *Globalisation, Societies and Education*, 19(2), 245–258. <https://doi.org/10.1080/14767724.2021.1878014>
- Leung, F. K. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579-605.
- Lockheed, M., Prokic-Bruer, T., & Shadrova, A. (2015). The experience of middle-income countries participating in PISA 2000-2015 (PISA series). Washington, DC & Paris: The World Bank & OECD Publishing. doi: <http://dx.doi.org/10.1787/9789264246195-en>
- Lundahl, C., & Serder, M. (2020). Is PISA more important to school reforms than educational research? The selective use of authoritative references in media and in parliamentary debates. *Nordic Journal of Studies in Educational Policy*, 6(3), 193-206. <https://doi.org/10.1080/20020317.2020.1831306>
- Mohamed, M., & Morris, P. (2021). Buying, selling, and outsourcing educational reform: The global education industry and ‘policy borrowing’ in the Gulf. *Compare: A Journal of Comparative and International Education*, 51(2), 181-201. <https://doi.org/10.1080/03057925.2019.1607255>
- Mohammadpour, E., & Shekarchizadeh, A. (2015). Mathematics achievement in high- and low- achieving secondary schools. *Educational Psychology*, 35(6), 689-713. <https://doi.org/10.1080/01443410.2013.864753>
- Montoya, S., & Coleman, D. (2018, April 25). A sound investment: The benefits of large-scale learning assessments. *UNESCO UIS*. <https://uis.unesco.org/en/blog/sound-investment-benefits-large-scale-learning-assessments>
- Mullis, I. V., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). TIMSS 2019 international results in mathematics and science. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/international-results>.
- OECD. (2019). PISA 2018 Assessment and Analytical Framework. Paris: OECD. <https://doi.org/10.1787/b25efab8-en>
- Paine, L., Blömeke, S., & Aydarova, O. (2016). Teachers and teaching in the context of globalization. *Handbook of Research on Teaching*, 717-786.
- Pettersson, D., Prøitz, T. S., & Forsberg, E. (2017). From role models to nations in need of advice: Norway and Sweden under the OECD’s magnifying glass. *Journal of*

- Education Policy*, 32(6), 721-744. <https://doi.org/10.1080/02680939.2017.1301557>
- Pons, X. (2011). What do we really learn from PISA? The sociology of its reception in three European countries (2001–2008) 1. *European Journal of Education*, 46(4), 540-548.
- Ringarp, J. (2016). PISA lends legitimacy: A study of education policy changes in Germany and Sweden after 2000. *European Educational Research Journal*, 15(4), 447-461. <https://doi.org/10.1177/1474904116630754>
- Rutkowski, L. (2014). *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*.
- Sabaha, S., & Hammouri, H. (2010). Does subject matter matter? Estimating the impact of instructional practices and resources on students' achievement in science and mathematics: Findings from TIMSS 2007. *Evaluation and Research in Education*, 23(4), 287-299. <https://doi.org/10.1080/09500790.2010.509782>
- Solano-Flores, G., Contreras-Niño, L. Á., & Backhoff, E. (2013). The measurement of translation error in PISA-2006 items: An application of the theory of test translation error. In *Research on PISA* (pp. 71-85). Springer, Dordrecht. [https://doi.org/10.1007/978-94-007-4458-5\\_5](https://doi.org/10.1007/978-94-007-4458-5_5)
- Sellar, S. & Lingard, B. (2014). The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*.
- Tobin, M., Nugroho, D., & Lietz, P. (2016). Large-scale assessments of students' learning and education policy: Synthesising evidence across world regions. *Research Papers in Education*, 31(5), 578–594. <https://doi.org/10.1080/02671522.2016.1225353>
- Tonga, F. E., Eryiğit, S., Yalçın, F. A., & Erden, F. T. (2022). Professional development of teachers in PISA achiever countries: Finland, Estonia, Japan, Singapore and China. *Professional Development in Education*, 48(1), 88-104.
- Torney-Purta, J., & Amadeo, J.A. (2013). International Large-Scale Assessments: Challenges in reporting and potentials for secondary analysis. *Research in Comparative and International Education*, 8(3), 248–258. <https://doi.org/10.2304/rcie.2013.8.3.248>
- UNESCO, (2018). Impact of large-scale assessments. <http://uis.unesco.org/sites/default/files/documents/impact-large-scale-assessments-2018-en.pdf>
- Wiseman, A. W. (2013). The impact of international achievement studies on national education policymaking. *International Perspectives on Education and Society*, Vol. 13. Emerald Group Publishing, Bingley. [https://doi.org/10.1108/S1479-3679\(2010\)00](https://doi.org/10.1108/S1479-3679(2010)00)