

The Education Compass: Navigating Beyond the North of PISA, TIMSS, and PIRLS

Anh-Duc Hoang

ducha@skylineschool.edu.vn
Sky-Line School (Vietnam),
and RMIT University (Australia)

ABSTRACT: *Global learning measurement systems such as Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), and Programme for International Student Assessment (PISA) have been widely used to compare the educational performance of different countries. However, these systems have some limitations that need to be considered. After analyzing 1,265 research articles on such measurement agendas, this article highlights blind spots and blackspots surrounding large-scale metrics' development, measurement, and implications. This commentary article emphasizes that essential factors about culture and diversity have yet to be included in such metrics, leading to monochrome paintings of the global education landscape rather than mosaics. Overall, we advocate for a strategic shift in developing and adopting such global measurement agendas.*

KEYWORDS: Global Education Measurement; PISA; TIMSS; PIRLS; K-12 Education

→ Received 04/04/2024 → Revised manuscript received 18/06/2024 → Published 30/6/2024.

1. Introduction

During the first Conference on Education Data and Statistics that was held by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) on February 6, 2024, it was found that the achievements of 680 million children are neither counted nor acknowledged. The Assessment for Minimum Proficiency Level (AMPL) will be implemented by the UNESCO Institute for Statistics in order to fill this data gap (UNESCO, 2024). In addition, governments and policymakers all over the world should revisit the existing concerns that have not been mentioned in global large-scale learning assessments such as PISA, TIMSS, and PIRLS before attempting to address the gaps that have not yet been identified and that will be exposed by the upcoming AMPL exams. With the help of the existing body of literature, the purpose of this commentary article is to provide an overview of the most significant issues concerning these global metrics, as well as to suggest potential future discourses on the approaches to, and consequences for, specific educational settings.

Beginning in the middle of the twentieth century, the trend of globalization in education has become more apparent (Martens et al., 2014). This started with the movement of foreign students from developing countries to Western

countries to pursue their post-secondary degrees (Pham et al., 2022). After that, there was a greater diversification of international education as developing nations began to absorb the most effective practices from industrialized education systems regarding lessons at both strategic levels and classroom-related approaches (Hoang et al., 2020). Within such a scenario, international measuring agendas have been activated to capture the mobility of educational development between countries worldwide and to activate policy implications for policymakers at both the national and international levels. Trends in International Mathematics and Science Study (TIMSS) is a program launched in 1995 by the International Association for the Evaluation of Educational Achievement (IEA). This program is designed to measure the outcomes of mathematics curricula that are taught all over the world through each four-year cycle of evaluation (Webster & Fisher, 2000). Another initiative, the Progress in International Reading Literacy Study (PIRLS), was developed by the IEA in 2001, which compares the reading performance of kids in grade four (ages 9 to 10) from various countries every four years (Bracey, 2003). The Organization for Economic Cooperation and Development (OECD) launched the Programme for International Student Assessment (PISA) in

2000 (Prais, 2003). This assessment is an attempt to measure the ability of students who are 15 years old to apply the mathematics, science, and literacy skills and knowledge they have been taught in situations that are not found in a textbook. For three years, one of the key goals of the PISA is to provide further knowledge about the culture, economics, and policies connected to the development of education in the countries being measured. Overall, these metrics compare students' educational accomplishments through standardized examinations and provide implications for governments regarding the effective design and execution of educational policy to improve academic outcomes (Kell & Kell, 2014).

Both measurements have been adopted by 64~79 countries and territories worldwide and have collected valuable time-series data. Their reports always become the focus of policymakers, educational researchers, the media, and public audiences. However, there are also considerable criticisms about their limitations, which suggest that these measurements do not provide overall portraits of education systems but only focus on students' performance gaps based on specific standardized tests. Besides, ethical considerations surround the validity and reliability of data collection in such transnational evaluations. Within the context that the OECD is going to release new reports, *PISA 2024 Creative Thinking* (first launched in 2022) (OECD, 2022) and *PISA 2025 Learning in the Digital World* (OECD, 2023), which examine students' self-regulated learning and learning practices in the digital world, this paper summarizes key focuses of studies on these most well-established global metrics, yearning to provide readers with more comprehensive perspectives toward such measurements.

2. Methodological Approach

To capture the discourse on these global education assessments, the researcher searched within the Web of Science database, looking for research articles that include either one of the three keywords TIMSS, PIRLS, or PISA within their title, from 1995 to December 31, 2023. The researcher only included English publications that fell under the education, psychology, and social sciences categories when using the PRISMA flow (Figure 1), which was suggested by Moher et al. (2011) as a method for narrowing down research papers for bibliometric analysis for the purpose of the study. As a result, 1,432 results have been filtered into a corpus of 1,265 research articles. The full list of all articles can be found at Harvard Dataverse (Hoang, 2024). Table 1 summarizes the number of studies that cover each or multiple global large-scale measurements. Overall, only four articles cover all three metrics, and 37 articles cover two out of three metrics. PISA is the latest measurement but has received the highest scholarly attention (875 articles). Even though TIMSS was introduced five years before PISA, this measurement received attention from 306 articles, which is 34.97 percent of PISA. Especially PIRLS, the metric focused on literacy studies, received only 43 research papers. It would be beneficial to conduct additional research on PIRLS in order to get a comprehensive understanding of the issues surrounding global large-scale educational evaluations.

3. Discussion

The over-focus on measuring a limited set of skills and knowledge

In the first place, one element associated with all of these global metrics is the undue focus

Table 1. Number of Studies that cover each or multiple global large-scale education measurements from 1995 to 2023

	Single metric	And TIMSS	And PIRLS	All three metrics
TIMSS (Since 1995)	306	-	-	
PIRLS (Since 2001)	43	7	-	4
PISA (Since 2000)	875	26	4	

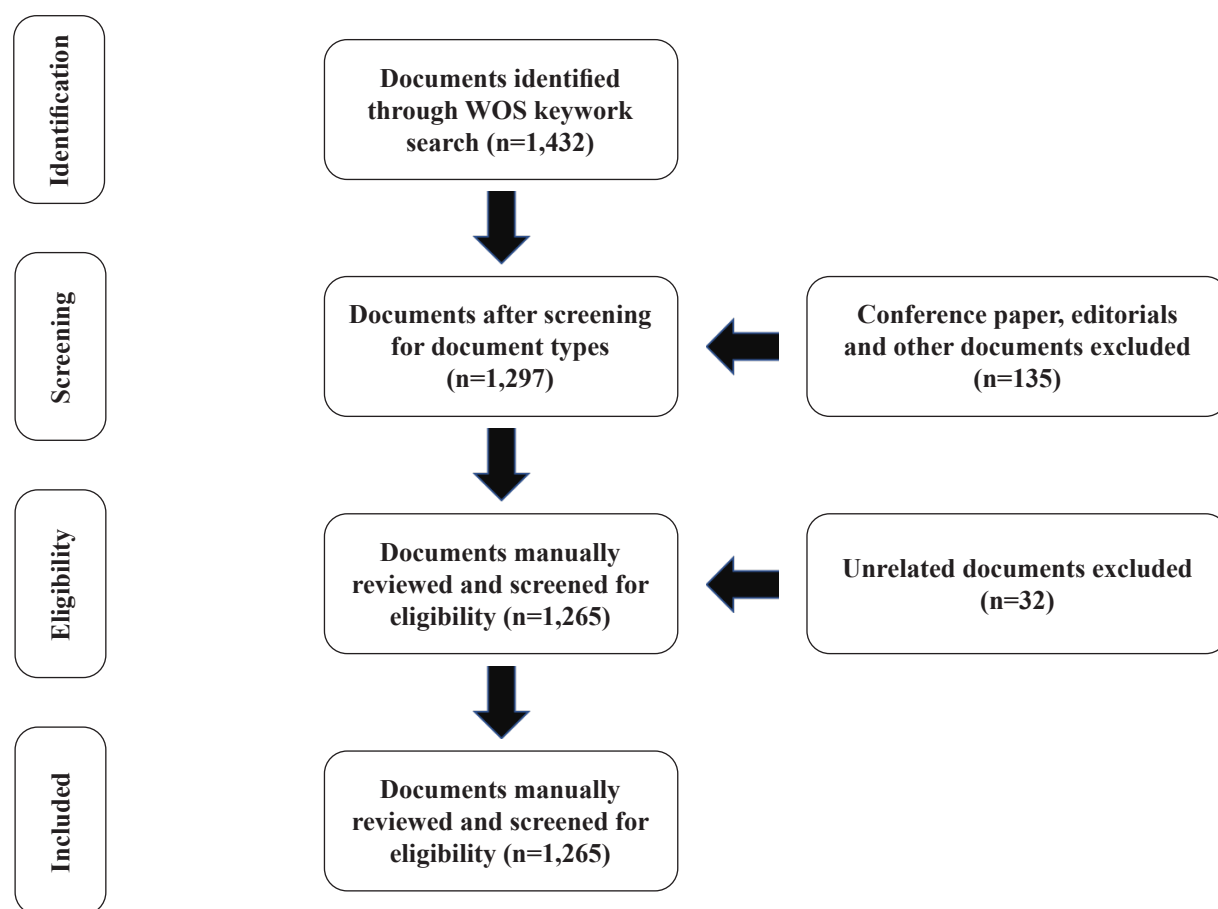


Figure 1: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram detailing the procedure to identify the corpus of 1,265 articles.

placed on standardized assessments. These examinations are designed to evaluate a particular collection of skills and information. Instead of assessing students’ “knowledge and skill for life,” PISA evaluates “knowledge and skill in assessment situations,” according to Dohn (2007), who suggested that this lack of measurement is problematic. Additionally, Dohn argued that this narrow focus on standardized assessments fails to capture the full range of students’ abilities and potential. This limited scope may not accurately reflect students’ readiness for real-world challenges and opportunities. The fact that this is the case indicates that such evaluations might not be able to capture the whole spectrum of skills children possess and might hide the students’ overall development. To prioritize narrow academic achievements over the cultivation of the breadth of 21st-century skills such as creativity, critical thinking, and non-cognitive abilities such as social-emotional intelligence,

collaboration, communication, adaptability, and civic engagement is a risky strategy in today’s modern world, which places a high value on creative problem-solving and problem-solving in general. It is important for educators and policymakers to recognize the limitations of traditional evaluations and consider alternative methods that can better assess a student’s holistic development. By fostering a well-rounded skill set in students, we can better prepare them for success in an ever-evolving global society.

The research on these three major worldwide evaluations is summarized in Figure 2. The research used co-occurrence keywords from the dataset of 1,265 research papers retrieved from the WoS database. The minimum threshold for the research studies was ten. Most studies investigating PISA and TIMSS are, without a doubt, the most prevalent. In addition to the nodes’ sizes, the nodes’ colors revealed several new topics (in yellow) compared to those visited

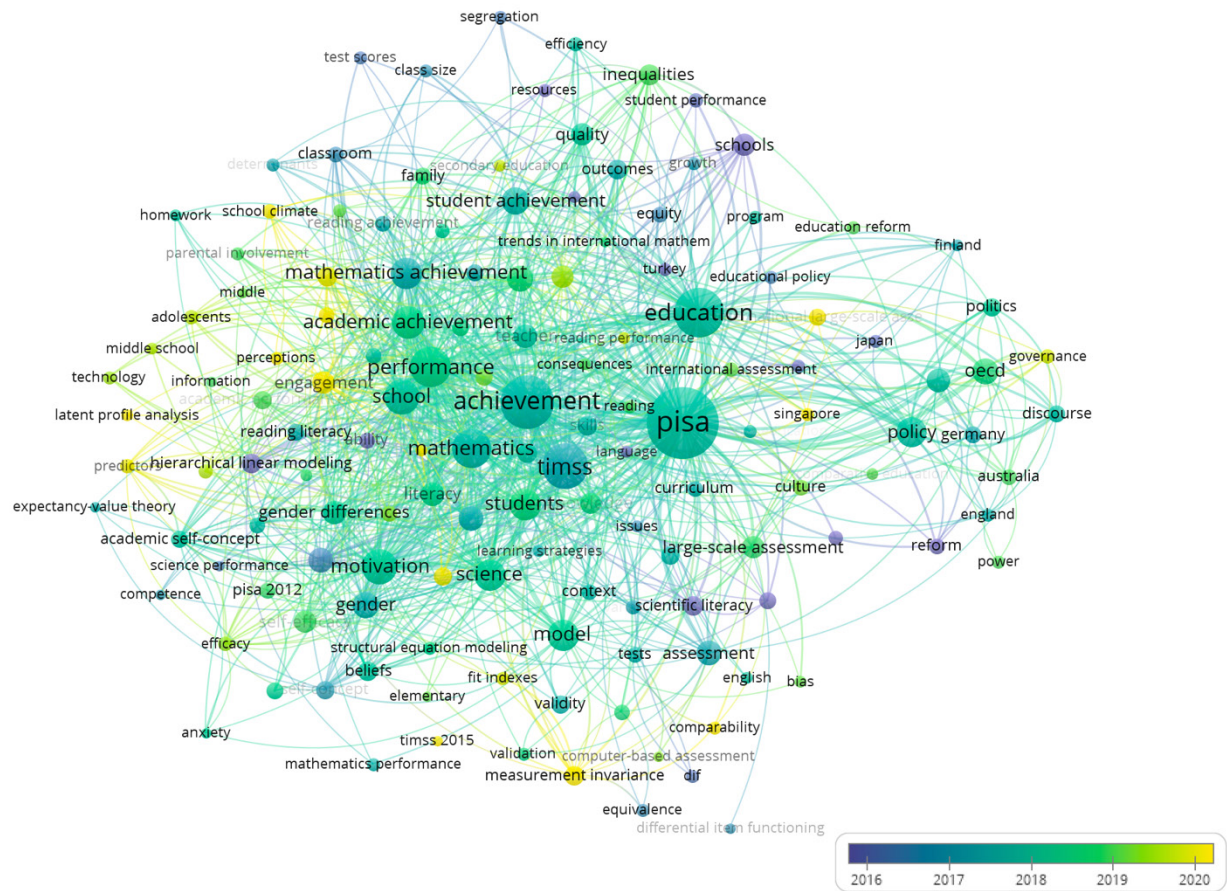


Figure 2. Co-occurrence of keywords on TIMSS, PIRLS, and PISA from 1995 to 2023.

frequently (in purple and green). However, there is still a long way to go until the novel concerns of student participation, perception, school atmosphere, and measurement invariance are fully addressed. Further comparisons regarding the total amount of research that focuses on TIMSS, PIRLS, and PISA are presented in the thematic map (Figure 3) that is presented here. Achievement is the motor theme (located in the top right quadrant) that drives those studies. However, as regards achievement, the current major global measurements focus on cognitive skills and academic performance, rather than other important aspects such as emotional intelligence and creativity. In addition, socioeconomic status is also a notable motor theme (located on the left side of the motor theme quadrant). However, aside from socioeconomic status, there are many other factors that can greatly impact a person's success and well-being. The emerging or decline theme (the bottom left quadrant) includes small nodes about policy discourses and local contexts.

This phenomenon suggests that studies on global measurement scales like TIMSS, PIRLS, and PISA need more attention to these specific details in order to provide a more comprehensive understanding of educational outcomes.

Furthermore, the complex and ever-changing nature of education is oversimplified because it only depends on the student's achievement of a narrow set of skills and information (Gravel & Robin, 2022). When it comes to the school level, these global measures were unable to adequately present a variety of criteria, including the quality of the teachers, the curriculum, the resources, and the level of family involvement, among others. When it comes to the national level, the excessive emphasis placed on the performance of specific grade level students on limited standardized exams not only hampers the impacts of particular efforts on the wide range of reforms in curriculum, preservice teacher education, teacher development, resource mobilization, and other areas, but it also inhibits the overall

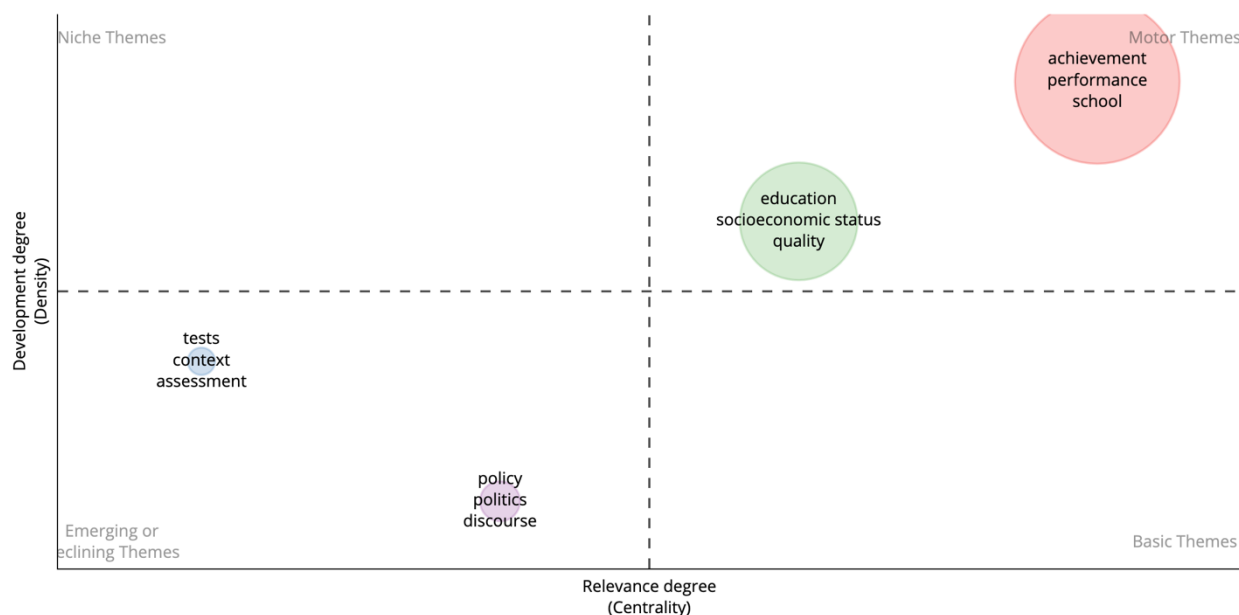


Figure 3. Thematic Development of Research Focus on TIMSS, PIRLS, and PISA, from 1995 to 2023.

effectiveness of these efforts. The only things included in publications such as PISA are broad statistics and superficial descriptive analyses of national reforms. What is often overlooked are the nuances and complexities of educational systems that cannot be captured by standardized test scores alone. This narrow focus on test results may lead to misguided policies and missed opportunities for comprehensive improvements in education. It does not, therefore, accurately reflect the effects of the laws and regulations that are in place in each of the participating countries (Rochex, 2006). PISA's forthcoming second report on creativity and first on learning practices in the digital environment should thus include additional research and discussions on the robustness of the results concerning micro-innovations and students' rounded backgrounds and development. Using data from such large-scale measurements, the most crucial question is: how can each government and school trust and use their action plans, and at what level can they apply this trust and utilization?

Despite the above incompleteness, there were also issues over the integrity and validity of the PISA report, particularly about the sampling procedure used across countries. Because of this, doubts arise regarding whether or not the

sampled group is genuinely representative of each country's educational system and student demographics. In addition, there is an additional worry regarding the genuine desire for a few typical numbers that can summarize each country's diverse demographics and the progress that has been made in development, by way of illustration, Schuelka (2013) it pointed out that the measurement and reporting procedures of TIMSS, PIRLS, and PISA failed to take into account the challenges faced by students with impairments. The situation of students who are immigrants is yet another example that has been passed over. As an illustration, Chinese immigrant children in New Zealand and Australia earn math scores more comparable to those of kids in Shanghai than their peers who are not immigrants in Australia and New Zealand (Feniger & Lefstein, 2014).

The possibility of bias in assessment tools is another facet that should be investigated in addition to the techniques already applied to such global education measures. Is there a cultural bias in their approach, favoring specific educational systems or countries over others? For example, do students from Asian backgrounds have to address concerns associated with the Western context in the same

section of questions asked concerning the Asian context as students from Western backgrounds must answer? Additionally, due to the nature of such global standardized examinations, they may mistakenly favor many different cultures and language situations, resulting in skewed results and consequences. These are important questions to consider in order to ensure the fairness and accuracy of assessment tools in global education. By addressing potential biases, we can strive for more equitable and inclusive evaluation methods that truly reflect the diverse experiences and backgrounds of all students. In addition, it would be fascinating to explore how other nations interpret and put into practice the findings of the PISA assessments. Do they utilize them as a foundation for making decisions regarding policy? When it comes to the methods and approaches that are used in schools, how do these decisions affect them? Understanding how different countries incorporate assessment tools like PISA results into their educational systems can provide valuable insights into the impact of standardized testing on policy-making and instructional practices. Thus, beyond the current discourses of these global measurement scales, there is a huge demand for comparative studies that examine the (in)effectiveness of international or transnational approaches when they are applied to other countries. Such insights can help inform the future educational strategies of countries from various backgrounds, as well as improve global education standards.

The monochrome portrays a country's education via its linkages with socioeconomic disparities.

In addition to putting excessive emphasis on the student's ability to acquire particular skill sets, these global education metrics also emphasized the connection between students' educational accomplishments and their socioeconomic standing. According to the findings of factorial thematic analysis of keywords derived from 1,265 research concerning the three most crucial global education metrics—TIMSS, PIRLS, and PISA—the results are presented in Figure 4. The primary focus of those measurements is indicated by the first branch on the left, which is

the association between academic achievement and socioeconomic position. A person's socioeconomic level may be the most crucial factor in determining whether or not they will be successful for the rest of their lives. Nevertheless, these global education measurements and other global metrics frequently struggle to reflect the complexities of socioeconomic differences between subgroups within a nation and between nations. This may be the case both inside and between nations, and especially, such general comparisons often ignore many perspectives on developing countries (Vuong, 2018). As a result of the emphasis placed on such average scores, new blind spots of educational injustices and inequalities may be created, which will impede efforts to address the underlying reasons of educational issues in specific locations. When seen from one perspective, these methods have the potential to guarantee specific degrees of reliability and validity, which in turn guarantees generalizability. However, as regards each school, school district, or province, a better understanding of their progress should be given higher priority than the yearning to compare themselves with other regions, which might have similar socioeconomic status but not the same culture and structures, as well as other operational and historical concerns. Therefore, while creating new implications from these global educational measures, policymakers, educational researchers, the media, and members of the public audience from each country need to exercise caution, especially around the release events of these global metrics (Baroutsis & Lingard, 2022).

Some studies focus on institutional impacts at the national and school levels on policy, rules, and quality (e.g. Raitano & Vona, 2013; Engel, 2015; Sato, 2017). These studies fall under the second branch of factorial thematic analysis. In addition, many studies study the connections between students' academic achievement and the role models that teachers provide, as well as the students' attitudes, levels of motivation, and levels of engagement (e.g. Glassow et al., 2021; and addressing the teaching contexts may be crucial for furthering this line of inquiry. International

large-scale assessments are well positioned to undertake such questions due to their systematic sampling of students, schools, and education systems. However, researchers are frequently prohibited from answering such questions due to measurement invariance related issues. This study uses the traditional multiple group confirmatory factor analysis (MGCFAs Østbø & Zachrisson, 2022; Vik et al., 2022). Despite this, the contributions of the components above have been obscured within these three global measures due to the excessive emphasis placed on students' academic performance through the utilization of recommended standardized examinations. In addition, these three global metrics are time-series data that quantify kids' accomplishments at a particular age after circles of three, four, or five years have passed. In this way, it can reflect the effects of national policies, rules, and investments on students on the verge of entering a specific grade level. On the other hand, the continuing improvement of students after multiple years is the most relevant time-series data that policymakers need to pay attention to. However, the time series data to tackle students' development progress have been ignored, even though these measurements primarily focus on the links between socioeconomic gaps and achievement. By filling this data gap,

policymakers and educational researchers can bring further insights into the effects of elements directly associated with the effectiveness of teaching and learning.

The unnecessary pressures and unmeaningful competition between countries

These global measures not only created unforeseen repercussions on schools, districts, and countries worldwide, but they also caused an excessive concentration on the accomplishment of children in limited skill sets and their links with socioeconomic status. In comparison to the goals of classrooms and the education system, the objective of global measurements such as PISA is still significantly different. Labaree (2014) even proposed an absolute statement: "PISA measures what no one teaches". The outcomes of inadequate conceptualization from these global measures are still used by many countries to modify their educational plans (Gaber et al., 2012). This is although there are issues that contradict each other. Regarding the situation in England, for example, Jerrim (2011) concluded that the deteriorating trend in PISA for this nation did not coincide with their results in TIMSS. This was since both methodologies needed an adequate data structure. In light of this, which perspectives ought to serve as the foundation upon which policymakers should

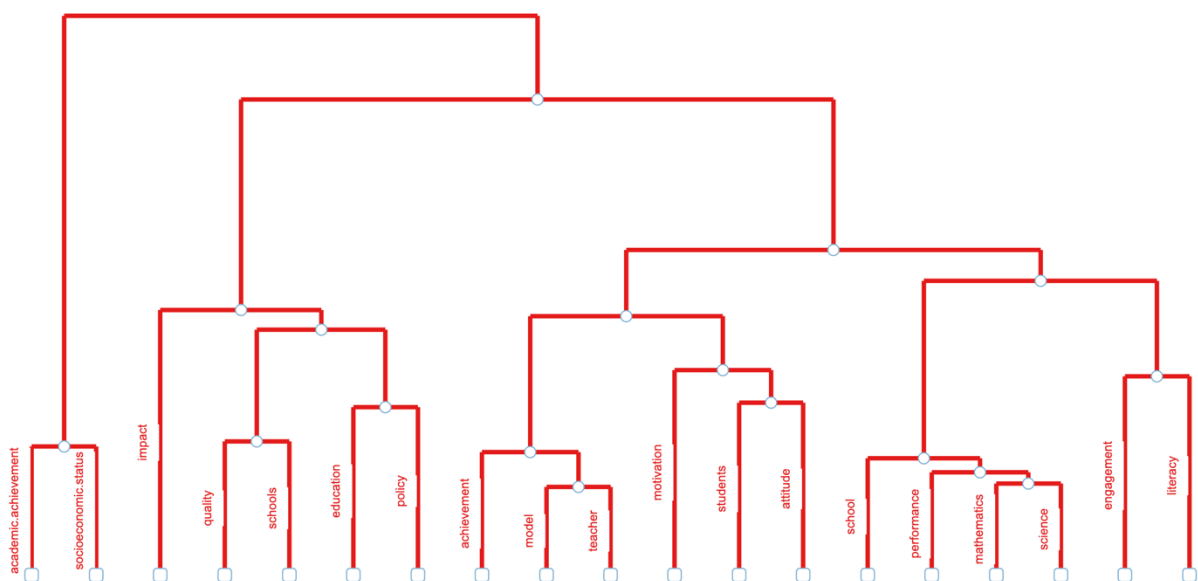


Figure 4. Factorial Thematic Analysis of Keywords from 1,265 Studies on TIMSS, PIRLS, and PISA from 1995 to 2023, extracted from the Web of Science Database

base their efforts to alter their educational policies and action plans? Another example is Portugal, where there is a significant disparity between the actual population and the sample used for the PISA study (Freitas et al., 2016). The Ministry of Education in Japan proudly owns a remarkable and successful endeavor. Concerning Japan, the media's interpretation of the PISA 2003 findings also caused pressure on policymakers (Takayama, 2008). However, the Japan Ministry of Education also leveraged this debate to redirect the public to follow their implementation of the *Yutori* (A Japanese word, which means low-pressure) curricular policy. Nevertheless, Japan is a noteworthy exception to this rule.

These systems have the potential to foster a culture of competition between countries, which can result in a limited concentration on test results rather than a comprehensive vision of education (see Muench et al., 2023; Vasalampi et al., 2023). This is even though such contentious discourses are still present across countries, for example, Canada (Stack, 2006), Swiss (Fredriksson et al., 2009), Italy (Checchi, D., & Verzillo, 2017), Turkey (Oz, 2021), Hongkong (Chen, 2022), Norway (Elstad, 2023). Because these systems can be used to rank countries according to their test scores, they can contribute to forming a hierarchy and a sense of supremacy among nations (Rowley et al., 2019). This may hurt the self-esteem of students and teachers in countries that are more poorly ranked (Hwang et al., 2017). In addition, countries that promote these rankings risk resorting to "teaching to the test," which can distort students' educational experience and drive teachers away from a more comprehensive and holistic approach to teaching.

Regarding the teaching and learning perspective, education is about providing students with the assistance they need to continuously develop themselves, regardless of whether the students are taking gigantic or modest steps. Educators and educational institutions can reduce the performance and opportunity gaps if they focus on their progress. From the classroom's perspective, individual education plans (IEP), which aim to make

students more resilient and develop them to become better versions of themselves daily, have become increasingly popular among countries. However, at the macro level, the worldwide large-scale education assessments, on the other hand, placed an excessive amount of emphasis on socioeconomic-based comparisons, neglected to present diverse efforts from a variety of countries, cultures, and subgroups (Burdett, 2013). They have also ignored the opportunities to investigate and elevate micro innovations, which are essential for improving education, without making significant financial investments.

4. Conclusion

In conclusion, although global measuring systems like TIMSS, PIRLS, and PISA have the potential to offer vital information regarding educational achievement, it is essential to keep in mind that these systems have some limitations that must be taken into consideration. The limitations discussed earlier serve as a reminder that, even though these assessments offer significant insights, a one-size-fits-all strategy could not solve everyone's educational problems because of the wide variety of educational environments around the world. Moving forward, it is vital to acknowledge these limits to contribute to developing an approach to education reform that is more nuanced, equitable, and effective on a local and global scale.

Utilizing these systems with other methods of evaluating educational quality and innovation is paramount. Additionally, avoiding excessively relying on test scores as the single instrument for determining educational achievement is essential. It is necessary to investigate alternate assessment techniques in addition to standardized testing to provide a more comprehensive perspective on the capabilities of teachers and students. Students can exhibit their abilities and knowledge in authentic contexts through performance-based assessments such as portfolios or projects. This contrasts the traditional method of assessing students through universally applicable tests. These kinds of evaluations have the potential to more accurately capture individual capabilities and give students from a wide range of

backgrounds the opportunity to exhibit their one-of-a-kind abilities. Implementing these forms of formative assessments in worldwide large-scale measurement, on the other hand, is extremely expensive according to current standards.

Instead of placing an excessive amount of emphasis on socioeconomic indicators, there ought to be a more significant number of prisms surrounding the achievement of students. One of the potential unintended consequences of a homogenized global agenda is that it may inhibit local creativity and distinctive approaches to education. Scholars are faced with the issue of proposing lean and adaptive dynamic models that can quantify the success of schools or regions. This is because of the complexity of socioeconomic disparities, cultural differences between regions and countries, policies and regulations, organizational structures, and practices. Such powerful agendas could embrace the culture of lifelong learning, elevate micro innovations, and highlight best practices in the classroom as an alternative to continuing to sustain the unnecessary pressures put on nations and territories based on their performance on standardized tests.

It is projected that the PISA will shortly release its second report on the creative talents of students in the year 2024. Additionally, it is anticipated that the PISA will release its first report on self-regulated learning and learning practices in the digital era in the year 2025. Is it possible that these subsequent reports will provide thorough depictions of the talents possessed by the students, or will they only serve as a perfunctory complement to the sacrifices

that have already been made in previous reports? What other inclusive dimensions, such as differences in culture, accessibility, family education, teacher-student interaction, and so on, will be reflected in these powerful reports, and how will these domains be reflected? What strategies will be utilized by public audiences, policymakers, media outlets, academics, and other stakeholders in order to successfully navigate the conversations and repercussions that the studies have generated? There is a high probability that this will result in an ongoing conversation that will place additional responsibilities on the shoulders of educational institutions, schools, teachers, students, and even individuals who are parents. When it comes to the ‘Yutori’ case, would the Japanese Ministry of Education be successful once more, just as they were in the past? PISA is an example of a worldwide large-scale assessment that utilizes expanded prisms that stretch beyond the perspectives of limited standardized tests and socioeconomic position. It is of the utmost significance that all stakeholders be given with such evaluations. Agendas that are more intelligent and inclusive need to be used to guide the actions and directions that education systems and schools located all over the world will take in the future. It is time for us to have this conversation.

Disclosure and Declarations

This research received no funding and does not have any conflict of interest. RMIT University provides access to the Web of Science database. Data analyses have been conducted through open-source software: VOS Viewer version 1.6.19 and R programming.

References

- Baroutsis, A., & Lingard, B. (2022). A methodological approach to the analysis of PISA microblogs: social media during the release of the PISA 2015 results. *Journal of Education Policy*, 37(6), 904–924. <https://doi.org/10.1080/02680939.2021.1937706>
- Bracey, G. W. (2003). PIRLS before the Press. *Phi Delta Kappan*, 84(10), 795–795. <https://doi.org/10.1177/003172170308401017>
- Burdett, N. (2013). The misuse of international studies in UK education. *SecED*.
- Checchi, D., & Verzillo, S. (2017). The Role of PISA in Regional and Private/Public Debates in Italy. In L. Volante (Ed.), *Routledge Research in Education Policy and Politics*, (pp. 127–148). Routledge.
- Chen, X. (2022). The effects of individual- and class-level achievement on attitudes towards mathematics: An analysis of Hong Kong students using TIMSS 2019. *Studies in Educational Evaluation*, 72, 101113. <https://doi.org/10.1016/j.stueduc.2021.101113>
- Dohn, N. B. (2007). Knowledge and Skills for PISA? Assessing the Assessment. *Journal of Philosophy of Education*, 41(1), 1–16. <https://doi.org/10.1016/j.stueduc.2021.101113>

org/10.1111/j.1467-9752.2007.00542.x

- Elstad, E. (2023). *The Evolution of the Extended Comprehensive School Model and the Modern Profession-Oriented Teacher Education After World War II* (pp. 35–72). https://doi.org/10.1007/978-3-031-26051-3_3
- 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>
- Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: an ironic investigation. *Journal of Education Policy, 29*(6), 845–855. <https://doi.org/10.1080/02680939.2014.892156>
- Fredriksson, U., Holzer, T., McCluskey-Cavin, H., & Taube, K. (2009). Strengths and Weaknesses in the Swedish and Swiss Education Systems: A Comparative Analysis Based on PISA Data. *European Educational Research Journal, 8*(1), 54–68. <https://doi.org/10.2304/eej.2009.8.1.54>
- Freitas, P., Nunes, L. C., Balcão Reis, A., Seabra, C., & Ferro, A. (2016). Correcting for sample problems in PISA and the improvement in Portuguese students' performance. *Assessment in Education: Principles, Policy & Practice, 23*(4), 456–472. <https://doi.org/10.1080/0969594X.2015.1105784>
- 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>
- Glassow, L. N., Rolfe, V., & Hansen, K. Y. (2021). Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: an alignment optimization approach. *Educational Assessment, Evaluation and Accountability, 33*(1), 105–137. <https://doi.org/10.1007/s11092-020-09348-2>
- Gravel, N., & Robin, L. (2022). Evaluating Education Systems Through Inequalities Between Families. *Dialogues Economiques*. <https://www.dialogueseconomiques.fr/en/article/evaluating-education-systems-through-inequalities-between-families>
- Hoang, A.-D. (2024). *Studies on TIMSS, PIRLS and PISA from 1995 to 2023*. Harvard Dataverse. <https://doi.org/10.7910/DVN/YXKVUR>
- Hoang, A.-D., Pham, H.-H., Nguyen, Y.-C., Nguyen, L.-K.-N., Vuong, Q.-H., Dam, M. Q., Tran, T., & Nguyen, T.-T. (2020). Introducing a tool to gauge curriculum quality under Sustainable Development Goal 4: The case of primary schools in Vietnam. *International Review of Education, 66*(4), 457–485. <https://doi.org/10.1007/s11159-020-09850-1>
- Hwang, J., Runnalls, C., Bhansali, S., Navaandamba, K., & Choi, K. M. (2017). “Can I do well in mathematics reasoning?” Comparing US and Finnish students' attitude and reasoning via TIMSS 2011. *Educational Research and Evaluation, 23*(7–8), 328–348. <https://doi.org/10.1080/13803611.2018.1500293>
- Jerrim, J. (2011). *England's “plummeting” PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline* (DoQSS Working Papers).
- Kell, M., & Kell, P. (2014). Global Testing: PISA, TIMSS and PIRLS. In *Literacy and Language in East Asia* (pp. 33–49). Springer Singapore. https://doi.org/10.1007/978-981-4451-30-7_5
- Labaree, D. F. (2014). Let's Measure What No One Teaches: PISA, NCLB, and the Shrinking Aims of Education. *Teachers College Record: The Voice of Scholarship in Education, 116*(9), 1–14. <https://doi.org/10.1177/016146811411600905>
- Martens, K., Knodel, P., & Windzio, M. (Eds.). (2014). *Internationalization of Education Policy*. Palgrave Macmillan UK. <https://doi.org/10.1057/9781137401694>
- Moher, D., Altman, D. G., Liberati, A., & Tetzlaff, J. (2011). PRISMA Statement. *Epidemiology, 22*(1), 128. <https://doi.org/10.1097/EDE.0b013e3181fe7825>
- Muench, R., Wieczorek, O., & Dressler, J. (2023). Equity lost: Sweden and Finland in the struggle for PISA scores. *European Educational Research Journal, 22*(3), 413–432. <https://doi.org/10.1177/147490412111069240>
- OECD. (2022). *PISA Creative Thinking*. <https://www.oecd.org/pisa/innovation/creative-thinking/OECD>.
- OECD. (2023). *PISA Learning Digital World*.
- Østbø, I. U., & Zachrisson, H. D. (2022). Student Motivation and Parental Attitude as Mediators for SES Effects on Mathematics Achievement: Evidence from Norway in TIMSS 2015. *Scandinavian Journal of Educational Research, 66*(5), 808–823. <https://doi.org/10.1080/00313831.2021.1939138>
- Oz, E. (2021). Comparability of teachers' educational background items in TIMSS: a case from Turkey. *Large-Scale Assessments in Education, 9*(1), 4. <https://doi.org/10.1186/s40536-021-00097-2>
- Pham, H.-H., Hoang, A.-D., Lai, S.-L., Dong, T.-K.-T., Nghia, T. L. H., Ho, M.-T., & Vuong, Q.-H. (2022). International education as an export sector: an investigation of 49 Vietnamese universities and colleges using Bayesian analysis. *Globalisation, Societies and Education, 1*–19. <https://doi.org/10.1080/14767724.2022.2081536>
- Prais, S. J. (2003). Cautions on OECD'S Recent Educational Survey (PISA). *Oxford Review of Education, 29*(2), 139–163. <https://doi.org/10.1080/0305498032000080657>
- Raitano, M., & Vona, F. (2013). Peer heterogeneity, school tracking and students' performances: evidence from PISA 2006. *Applied Economics, 45*(32), 4516–4532. <https://doi.org/10.1080/00036846.2013.791020>
- Rochex, J.-Y. (2006). Chapter 5: Social, Methodological,

- and Theoretical Issues Regarding Assessment: Lessons From a Secondary Analysis of PISA 2000 Literacy Tests. *Review of Research in Education*, 30(1), 163–212. <https://doi.org/10.3102/0091732X030001163>
- Rowley, K. J., McNeill, S. M., Dufur, M. J., Edmunds, C., & Jarvis, J. A. (2019). Trends in International PISA Scores over Time: Which Countries Are Actually Improving? *Social Sciences*, 8(8), 231. <https://doi.org/10.3390/socsci8080231>
- Sato, H. (2017). *The Structure of PISA Penetration into Education Policy in Japan and Norway* (pp. 209–230). <https://doi.org/10.1108/S1479-367920160000031011>
- Schuelka, M. J. (2013). Excluding students with disabilities from the culture of achievement: the case of the TIMSS, PIRLS, and PISA. *Journal of Education Policy*, 28(2), 216–230. <https://doi.org/10.1080/02680939.2012.708789>
- Stack, M. (2006). Testing, Testing, Read All about It: Canadian Press Coverage of the PISA Results. *Canadian Journal of Education / Revue Canadienne de l'éducation*, 29(1), 49. <https://doi.org/10.2307/20054146>
- 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>
- UNESCO. (2024). *Over half a billion children missing from education data*. UNESCO Conference on Education Data and Statistics. <https://www.linkedin.com/pulse/over-half-billion-children-missing-from-education-data-unesco-rwmhe/>
- Vasalampi, K., Tolvanen, A., Torppa, M., Poikkeus, A.-M., Hankimaa, H., & Aunola, K. (2023). PISA reading achievement, literacy motivation, and school burnout predicting Adolescents' educational track and educational attainment. *Learning and Individual Differences*, 108, 102377. <https://doi.org/10.1016/j.lindif.2023.102377>
- Vik, F. N., Nilsen, T., & Øverby, N. C. (2022). Aspects of nutritional deficits and cognitive outcomes – Triangulation across time and subject domains among students and teachers in TIMSS. *International Journal of Educational Development*, 89, 102553. <https://doi.org/10.1016/j.ijedudev.2022.102553>
- Vuong, Q.-H. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour*, 2(1), 5–5. <https://doi.org/10.1038/s41562-017-0281-4>
- Webster, B. J., & Fisher, D. L. (2000). Accounting for Variation in Science and Mathematics Achievement: A Multilevel Analysis of Australian Data Third International Mathematics and Science Study (Timss). *School Effectiveness and School Improvement*, 11(3), 339–360. [https://doi.org/10.1076/0924-3453\(200009\)11:3;1-G;FT339](https://doi.org/10.1076/0924-3453(200009)11:3;1-G;FT339)