

# MATHEMATICAL REASONING AND PROBLEM-SOLVING IN PISA 2022 – HOW DO PERFORMANCE PROFILES VARY ACROSS COUNTRIES?

Jenni Ingram, Gabriel Lee, and Jamie Stiff

University of Oxford, UK

*In PISA 2022, a new process subdomain was introduced focused on mathematical reasoning. This process was seen as the core of the problem-solving process that typifies PISA mathematics assessments. The results of PISA 2022 suggest that students in some countries have relative strengths specifically in mathematical reasoning, relative to the other problem-solving processes, while in other countries, this is an area of relative weakness. In this paper, we explore whether distinctive country profiles can be identified based on relative differences in performance on the four subdomain processes using Latent Profile Analysis. The profiles identified offer further support for considering the role of cultural and language contexts when comparing performance in international education studies.*

## INTRODUCTION

International large-scale assessment (ILSA) studies in education, such as PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study), are highly influential drivers of policy and curriculum reform. In December 2023, the results of PISA 2022 were announced, which has been followed by many countries and education systems looking (selectively) towards the curricula, policies and practices of the higher-performing countries, using the PISA results as an external justification for reform (Johansson & Strietholt, 2019; Rojano & Solares-Rojas, 2018). For many of these higher-performing countries and education systems, these results are also being used as a justification for the success of recent reforms. Yet this process of turning to ILSAs for justification usually does “not draw on the detailed analytical insights that might be drawn from the PISA data” (Lingard, 2017, p.1). Throughout the remainder of this paper, the term country is used to describe all countries and education systems that participated in PISA 2022.

Around the world, mathematics curricula have become increasingly similar both in terms of the content areas included and the role of problem-solving and mathematical reasoning (Kadijevich et al., 2023; Valero, 2023). This homogenisation has been attributed to the growing globalisation of education policy and the influence of ILSAs, particularly ones that provide a ‘ranking’ of performance in mathematics and other key curriculum areas (Takayama, 2008). Of the two largest ILSAs focused on school mathematics, PISA assesses mathematical literacy with a focus on problem-solving in different task contexts, in contrast to TIMSS, where the focus is largely on mathematical content. The influence of both of these on mathematics curricula around

the world is visible in the increasing inclusion of particular mathematical content (e.g. statistics) and mathematical processes such as problem-solving as a major objective. That is, mathematics curricula around the world are including more and more of what is assessed in these ILSAs (Stacey et al, 2015).

In 2022 a new component, mathematical reasoning, was added to the mathematics framework for the PISA (OECD, 2023a). Described as being a core aspect of mathematical literacy, the PISA assessment framework “highlights the centrality of mathematical reasoning both to the problem-solving cycle and to mathematical literacy in general” (p. 23). This change aligns with recent research advocating that the main goal of a mathematics curriculum needs to be understanding which is reached through reasoning, with problem-solving as a means to develop this reasoning (Olivares et al., 2019). This new emphasis offers an opportunity to examine variations in students' strengths and weaknesses in the processes involved in problem-solving and mathematical reasoning in different country contexts.

One argument explaining the higher rankings of East Asian countries that has been seen in recent cycles of PISA is the tight focus of the national curricula in these countries on mathematics and science (and languages) rather than a more general and broader school curriculum, which are precisely the skills and content tested in TIMSS and PISA (Deng & Gopinathan, 2016). These arguments are often based on studies that show that students in these countries perform relatively low in measures of critical thinking and creativity (Lim, 2010) and also focus on the content of these curricula rather than the aims and objectives.

Problem-solving and reasoning, in contrast, are often considered in the aims, objectives and intended outcomes of a national curriculum. For example, Singapore was the highest-performing country in mathematics in PISA 2022, with an average score significantly higher than any other country (OECD, 2023b). The mathematics curriculum in Singapore has had problem-solving as the “primary aim” since 1990 (Fan & Zhu, 2007), with reasoning as part of the ‘processes’ component of the curricula. In Singapore, reasoning is the focus of the majority of the learning outcomes in the secondary mathematics curriculum (Serçe & Acar, 2021).

The results reporting the average subdomain scores published by the OECD show that some countries scored significantly higher or significantly lower in some subdomain areas than in others (<https://www.oecd.org/pisa/>). The research described above suggests that these relative strengths and weaknesses can be partly explained by differences in curricular focus across the different countries, particularly in relation to the mathematical content subdomains. The new subdomain of mathematical reasoning in PISA 2022 also offers an opportunity to examine these differences in relation to the mathematical processes.

This paper focuses on two research questions; What profiles in mathematics process performance in PISA 2022 can be identified? How do outcomes on the process subdomains differ across these profiles?

## METHODS

PISA uses a multi-stage adaptive test design meaning that participating students are required to only answer a subset of the items used. The mathematics and mathematics subdomains scores are calculated using plausible values for all students which have been scaled using multi-dimensional models. The mathematics scale and each of the process and content subscales are scaled together, meaning that analyses can consider relationships between the performances on each of the subdomain scores.

In this paper, we focus on the four process subdomains in mathematics; mathematical reasoning (or ‘reasoning’), formulating situations mathematically (or ‘formulate’), employing mathematical concepts, facts, and procedures (or ‘employing’) and interpreting, applying and evaluating mathematical outcomes (or ‘interpreting’). Data were drawn from the 76 countries for which the subdomain scores are available. The average subdomain scores were calculated in a way that takes into account PISA’s complex sample design, including the use of student weights and plausible values.

The subdomain scores correlate highly with the overall mathematics scale, with higher-performing countries tending to perform highly on all scales, and lower-performing countries tending to perform lower. To address these general differences in performance, relative scores were obtained for each subdomain by subtracting the mean of the mathematical reasoning subdomain score from each of the process subdomain mean scores in each country. The relative scores therefore reflect the relative strengths or weaknesses in each subdomain to the mathematical reasoning mean score. Note, the overall mathematics score is not necessarily equal to the mean of the four process subdomain scores.

In addition, a homogeneity score was calculated for each country, which is the sum of the absolute values of the four subdomain scores minus the average of these four subdomain scores. A homogeneity score close to zero indicates that there was little variation between performance in the four subdomains, while a large value shows more pronounced strengths and weaknesses in particular areas. The average homogeneity score across all participating countries was 11.9 with a standard deviation of 5.0.

Latent profile analysis (LPA) was used to identify groups of countries with distinct performance profiles across the mathematics process subdomains. Countries are classified into groups based on membership probabilities estimated in the LPA model. The LPA variables were the four relative subdomain scores. Initially, the potential of three or four classes were examined based on a visual analysis of the relative subdomain scores, and the findings of a similar cross-country analysis of mathematics performance over time using TIMSS data (Johansson & Strietholt, 2019). Using the *mclust* package in R (Scrucca et al., 2023), the best fitting profile model was identified initially by comparing 3-group and 4-group models using the Bayesian Information Criterion and then using a Bootstrap Likelihood Ratio Test ( $p < 0.001$ ) to identify the optimal number of profiles. The final model was a 3-group spherical varying volume model.

**RESULTS**

The Latent Profile Analysis identified 3 groups. The first group included 28 countries, while the second group included 27 countries, and the third group included 21 countries. The indicators for these 3 groups are shown in Table 1.

Subdomain	Group 1	Group 2	Group 3
Reasoning	0.00	0.00	0.00
Formulating	-3.03	3.95	-5.15
Employing	-0.11	7.05	-4.08
Interpreting	3.78	4.69	-6.07

Table 1: Indicators for the three group profiles for the subdomain processes.

Group 1 is characterised by a stronger mean score in interpreting and a weaker mean score in formulating, with reasoning and employing in between these two. The average homogeneity score for Group 1 was the smallest of the three groups (10.5), suggesting that these countries can also be characterised by relatively consistent scores across the four process subdomains, including the mathematical reasoning domain.

Group 2 is characterised by a weaker mean score in mathematical reasoning and a stronger employing mean score, with formulating and interpreting in between these two.

Group 3 is characterised by a relatively stronger mean score in mathematical reasoning in contrast to the other three process subdomains. Group 3 also had the largest average homogeneity score of 13.7 suggesting a wider spread of subdomain scores than in the other groups. However, the similar indicator values for the formulating, employing and interpreting subdomains suggest that the performance in these areas were similar to each other.

The most likely latent profile membership of the highest-performing countries is plotted in Figure 1. Countries have been ordered by their mean score for mathematics, and the figure shows the mean subdomain scores for each country with a mathematics mean performance significantly higher than the OECD average. The error bars represent the 95% confidence interval for the respective subdomain mean. (The graph showing the most likely profile membership of all participating countries will be shared in the presentation).

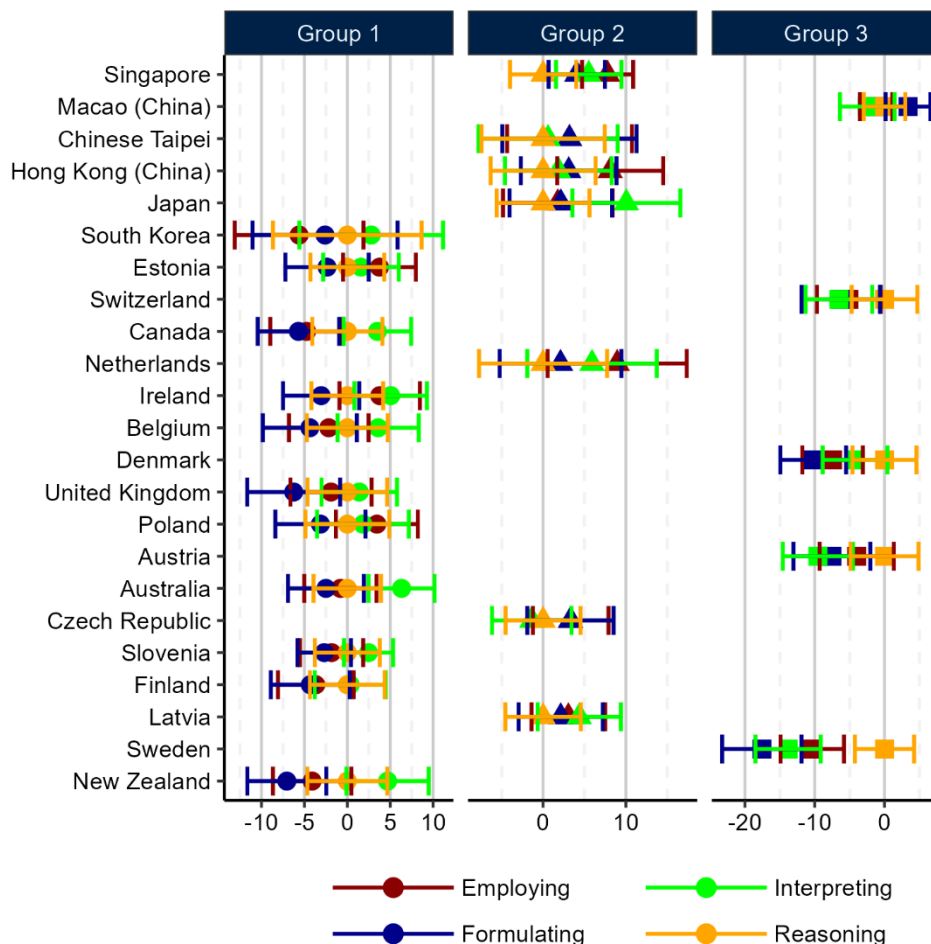


Figure 1: Subdomain mean scores for the highest-performing countries in PISA 2022.

The average mathematics score for countries most likely to be in Group 1 was 451, compared to an average score of 436 for Groups 2 and 3, suggesting that countries in Group 1 have stronger mathematics performance on average than countries in Groups 2 or 3. This was also the group with the highest average performance in mathematical reasoning with an average of 449, compared to an average of 430 for countries most likely to be in Group 2 and an average of 439 for countries most likely to be in Group 3.

The majority of the 6 top performing countries, which all performed significantly higher in mathematics than the other participating countries, were most likely to be in Group 2. However, the majority of the other countries most likely to be in Group 2 score significantly lower than the OECD average in mathematics (18 out of 27

countries). In contrast, less than half of the countries most likely to be in Group 1 had an average mathematics score significantly lower than the OECD average (13 out of 28 countries). English-speaking countries were also more likely to be in Group 1.

Over two-thirds of these higher-performing countries had homogeneity scores that were lower than the average homogeneity score across all participating countries. Of the top 6 performing countries, only Japan had a homogeneity score (13.1) that was larger than the average across all participating countries. Other higher-performing countries with an above-average homogeneity score included Sweden (21.0), New Zealand (15.8), Canada (13.9), Denmark (13.4), Austria (13.1) and The Netherlands (12.8).

## **DISCUSSION**

The analysis identified three profiles of performance in the PISA 2022 mathematics process subdomains. As well as illustrating different profiles of performance involving mathematical reasoning and problem-solving processes, they also illustrate some important cultural and linguistic differences. Johansson and Strietholt (2019) found that country-level strengths or weaknesses in mathematics content areas persist over time, suggesting that these profiles result from cultural contexts, including national curricula and national education policies and practices.

The majority of East Asian countries were most likely to be in Group 2, with a relatively stronger performance in employing and a relatively weaker performance in reasoning. The majority of European and English-speaking countries were most likely to be in Group 1, characterised by a stronger mean score in interpreting and a weaker mean score in formulating. These profiles suggest that culture and language have a substantial impact on students' mathematics performance in ILSAs. Furthermore, for countries most likely to be in Group 1, the mean interpreting score was often the highest score. This may reflect the emphasis on model validation and interpretation present in European mathematics education research (Niss, 1994; Geisler, 2021).

The highest 6 performing countries were all East Asian, which in the past has led to what Sellar and Linard (2013) referred to as the phenomenon of “looking East” to identify policies and practices explaining this high performance. The majority of these countries were most likely to be in Group 2, which was characterised by a weaker relative average performance in mathematical reasoning but a stronger relative average performance in employing; that is, employing mathematical concepts, facts and procedures. This suggests an emphasis on fluency with mathematical content, rather than on reasoning. However, these characteristics are not sufficient for high performance in mathematics more generally, as the majority of countries with this profile of performance scored significantly below the OECD average in mathematics.

The PISA 2022 mathematics framework puts mathematical reasoning at the centre of the problem-solving process. However, only the profile of countries in Group 1 had average performance scores in reasoning at the middle of the problem-solving process measures. The profile of Group 3 suggests a greater emphasis on mathematical

reasoning than on the other problem-solving processes, while the profile of Group 2 suggests a greater emphasis on problem-solving processes than on mathematical reasoning. This analysis problematises the relationship between mathematical reasoning and the problem-solving processes and further research is needed to examine this relationship further.

The analysis in this paper assumes that PISA provides a valid measure of mathematics outcomes for different countries. While the participating countries influence the nature of the mathematics assessed, the content is based on what the OECD views as what students need to learn for today's (and tomorrow's) world (OECD, 2023a). This may align in different ways to country curricula. Furthermore, this analysis only examines the national picture, and it is also important to look within groups. The PISA 2022 results also found differences in mathematics performance between girls and boys, as well as depending upon a student's socioeconomic background. Finally, the differences in mean performance in each of the process subdomains were small in the majority of countries. While this paper has focused on comparisons with the performance in mathematical reasoning, the results also suggest that there is considerable variation in performance in the interpreting subdomain. Other models focusing on the relative performance in interpreting may lead to other performance profiles.

## References

- Deng, Z. & Gopinathan, S. (2016). PISA and high-performing education systems: explaining Singapore's education success. *Comparative Education*, 53(4), 499–472. <https://doi.org/10.1080/03050068.2016.1219535>
- Fan, L. & Zhu, Y. (2007). From convergence to divergence: the development of mathematical problem solving in research, curriculum, and classroom practice in Singapore. *ZDM*, 39, 491–501. <https://doi.org/10.1007/s11858-007-0044-1>
- Geisler, S. (2021). Data-based modelling with experiments – students' experiences with model-validation. In M. Inpraita, N. Changsri, & N. Boonsena (Eds.). *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education*, Vol 2, pp. 289–296. Khon Kaen, Thailand: PME.
- Johansson, S. & Strietholt, R. (2019). Globalised student achievement? A longitudinal and cross-country analysis of convergence in mathematics performance. *Comparative Education*, 55(4), 536–556. <https://doi.org/10.1080/03050068.2019.1657711>
- Kadijevich, D. M., Stephens, M., Solares-Rojas, A. & Guberman, R. (2023). Impacts of TIMSS and PISA on mathematics curriculum reforms. In Y. Shimizu & R. Vithal (Eds.), *Mathematics Curriculum Reforms around the World*, (pp. 359–374). Springer. [https://doi.org/10.1007/978-3-031-13548-4\\_22](https://doi.org/10.1007/978-3-031-13548-4_22)
- Lim, W. K. (2010). Asian Education Must Change to Promote Innovative Thinking. *Nature* 465 (7295): 157–157. <https://doi.org/10.1038/465157a>
- Lingard, B. (2017). *The OECD's PISA: Purposes, impact and usage*. Paper presented at the International symposium "Educational governance in the era of international assessment". Japan Women's University. <https://www.jeas.jp/jeas/wp->

[content/uploads/2017/10/Lingard-B.-The-OECD%E2%80%99s-PISA-Purposes-Impact-and-Usages.pdf](https://content/uploads/2017/10/Lingard-B.-The-OECD%E2%80%99s-PISA-Purposes-Impact-and-Usages.pdf)

Niss, M. (1994). Mathematics in Society. In R. Biehler, R. W. Scholz, R. Sträßer, & B. Winkelmann (Eds.). *Didactics of Mathematics as a Scientific Discipline* (pp. 367-378). Kluwer Academic Publishers.

OECD (2023a). *PISA 2022 Assessment and Analytical Framework*, PISA, OECD Publishing.

OECD (2023b). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*, PISA, OECD Publishing, <https://doi.org/10.1787/53f23881-en>

Olivares, D., Lupiáñez, J. L., & Segovia, I. (2021). Roles and characteristics of problem solving in the mathematics curriculum: a review. *International Journal of Mathematical Education in Science and Technology*, 52(7), 1079–1096.

Rojano, T., & Solares-Rojas, A. (2018). The mathematics curriculum design from an international perspective: Methodological elements from a comparative analysis. In Y. Shimizu & R. Vithal (Eds.), *School mathematics curriculum reforms: Challenges, changes and opportunities. Proceedings of the twenty-fourth ICMI Study conference* (pp. 475–482). International Commission on Mathematical Instruction.

Scrucca, L., Fraley, C., Murphy, T. B. & Raftery, A. E. (2023). *Model-based clustering, classification and density estimation using mclust in R*. Chapman and Hall. <https://doi.org/10.1201/9781003277965>

Serçe, F. & Acar, F. (2021). A comparative study of secondary mathematics curricula of Turkey, Estonia, Canada, and Singapore. *Journal of Pedagogical Research*, 5(1), pp. 216–242. <https://doi.org/10.33902/JPR.2021167798>

Stacey, K., Almuna, F., et al. (2015). PISA's influence on thought and action in mathematics education. In K. Stacey & R. Turner (Eds.), *Assessing mathematical literacy. The PISA experience* (pp. 275–306). Springer.

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>

Valero, P. (2023). Mathematical subjectivation: death sentence or chances for a terrestrial life. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds.). *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 1, pp 53–68). University of Haifa, Israel: PME.