Number 24

IEA COMPASS:
BRIEFS IN EDUCATION

USING PROCESS DATA IN LARGE-SCALE ASSESSMENTS

An Example With an eTIMSS Problem Solving and Inquiry Item

SUMMARY

Digital technologies have the potential to revolutionize education by enhancing quality, fairness, and efficiency. However, equitable access to these technologies remains a challenge. ILSAs (international large-scale assessments) have shown that the relationship between digital use and performance varies across countries and over time. To fully harness the potential of digital technologies, pedagogical processes must adapt, and effective teacher training is essential. The study focuses on analyzing the process data associated with a technology-enhanced item in the 2019 eTIMSS PSI (Problem Solving and Inquiry) mathematics assessment. Based on specific didactical hypothesis related to students' conceptions of proofs, the analysis aims to gain insights into students' problem-solving strategies. The findings support the didactical hypothesis and reveal the positive relationship between validation strategies and students' overall PSI scores. Despite limitations, this research can inform practitioners by providing valid and reliable information that promotes effective teaching and learning using digital technologies.

IMPLICATIONS

- Process data from ILSAs can inform educators about students' cognitive processes in technology-enhanced problem-solving. Understanding how students interact and benefit from technologies used in teaching is crucial for informed decision-making.
- The research supports the use of a theory-driven approach to process data, highlighting its potential to enhance assessment validity and deepen understanding of student achievement by revealing students' test-taking strategies and misconceptions.
- The presence of technology-enhanced items and the data they provide underscores that importance for education systems to carefully consider insights during the development of instructional material and assessment strategies.

Authors

Franck Salles: Ministère de l'éducation nationale et de la jeunesse (DEPP)
Aurélie Lacroix: Ministère de l'éducation nationale et de la jeunesse (DEPP)

INTRODUCTION

Digital technologies have the potential to revolutionize education by promoting quality, fairness, and efficiency. Technology empowers teachers to cater to individual student needs, enhance engagement, and provide access to diverse learning resources. However, ensuring equitable access to high-quality digital technologies remains a challenge in many education systems. The COVID-19 pandemic has underscored the importance of supportive policies and conditions to effectively utilize digital tools. Large-scale assessments like TIMSS (Trends in International Mathematics and Science Study), ICILS (International Computer and Information Literacy Study), and PISA (Programme for International Student Assessment) have demonstrated that the relationship between digital use and performance varies across countries and over time.

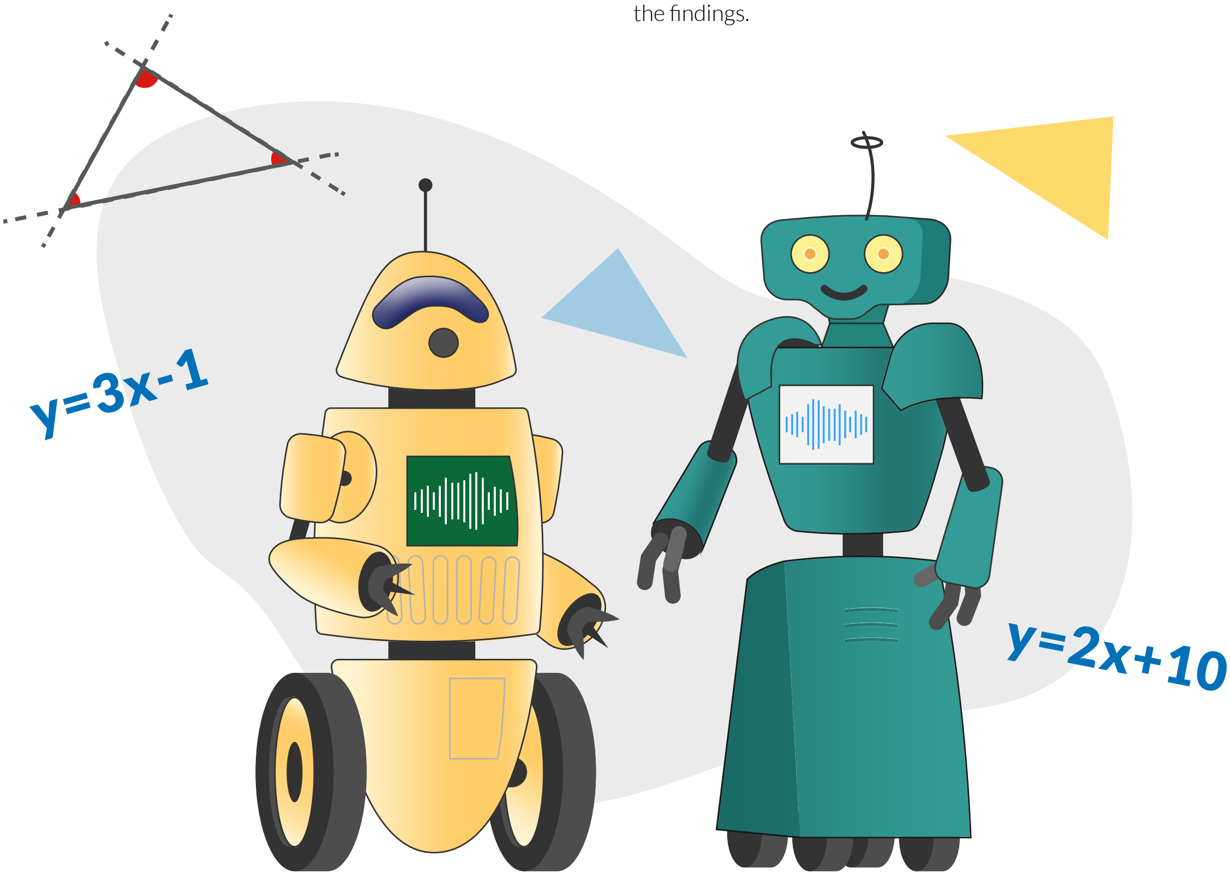
Globally, education institutions need to align valid and reliable information regarding the use of digital technologies with the needs of practitioners and curriculum requirements (OECD, 2023). Due to the computer-based delivery mode, ILSAs play a crucial role in informing policymakers and practitioners about the added benefits of digital technologies in teaching and learning.

Digital technologies offer dynamic learning and assessment experiences through interactive simulations and tools that address dynamic concepts such as mathematical concepts. They facilitate the development of higher-order problem-solving skills by outsourcing procedural tasks to computers, enabling students to focus on strategic thinking. Digital technologies also enrich formats and feedback, fostering increased student engagement and motivation (Drijvers, 2019).

In the field of large-scale assessment, the data collected and used from digital tests potentially open new horizons when it comes to analyzing and informing stakeholders of response processes. Digital assessments enable us to log and time students' interactions with the testing environment (mouse clicks, keystrokes, etc.). The collection and use of process data have advanced quickly in recent years and involve important new areas of activity (Maddox, 2023). Using mostly response time and keystrokes, one area focuses on students' motivation and engagement in relation to test performance (Ercikan et al., 2020). Another area focuses on designing technology-enhanced tasks to capture data in relation to higher-order constructs such as problem-solving or reasoning (Goldhammer et al., 2021; Salles et al., 2020).

The 2019 eTIMSS assessment introduced the PSI Tasks (Mullis et al., 2021), incorporating technology-enhanced items to fully leverage the potential of digital technologies in assessing mathematical skills. By analyzing this process data alongside students' response patterns, researchers can gain valuable insights into their problem-solving strategies, conceptions and misconceptions, and test-taking behaviors.

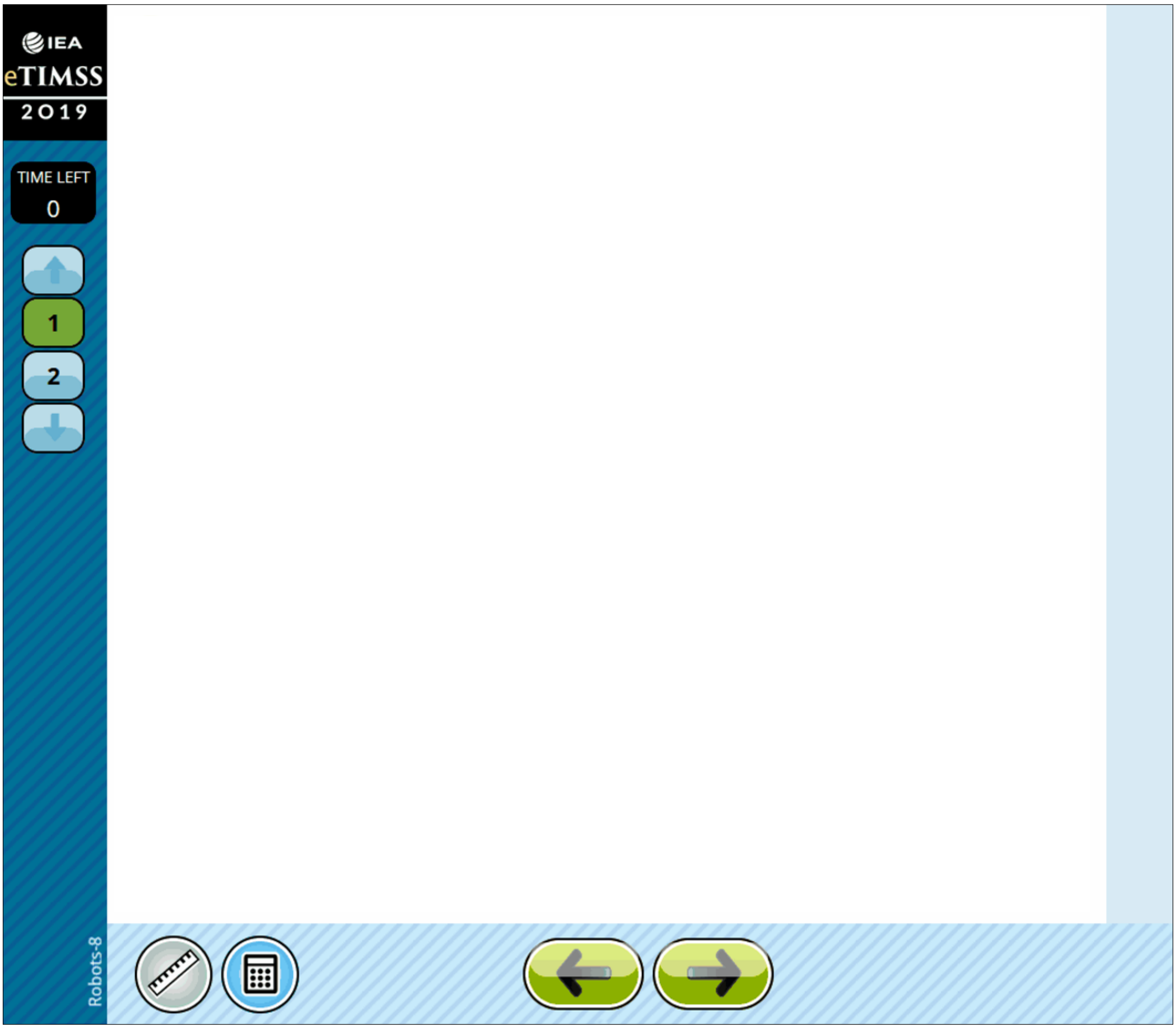
This study focuses on analyzing the process data associated with the Robots item in the 2019 eTIMSS PSI mathematics assessment. The aim is to gain a deeper understanding of students' problem-solving strategies (Salles et al., 2020), in particular, validation strategies. The research employs a didactical approach, guided by theory, to ensure a valid interpretation of the data and provide meaningful insights for educational practitioners when reporting the findings.



NEW DATA AND ANALYSIS

In the 2019 eTIMSS assessment, 22 countries and 5 benchmarking entities participated, with approximately 20,000 eighth-grade students undertaking a PSI test in mathematics and science. A key focus of this analysis is the Robots item, as detailed in the PSI report by IEA and the TIMSS and PIRLS International Study Center at Boston College (Mullis et al., 2021). This item, situated within the algebra content domain and the reasoning cognitive domain, exemplifies an innovative approach to assessment, utilizing technology beyond traditional paper-based methods.

The Robot item in the PSI mathematics 2019 eighth-grade eTIMSS assessment



The Robots item challenges students to discern a linear relationship between two variables, x and y, by inputting x values and observing the automatically generated y outputs. This item is designed to test students' abilities to integrate and synthesize information to discover an underlying mathematical relationship, specifically a linear equation ($y = 2x + 10$), without prior knowledge.[1]

Our study focuses on the strategies students employ in this task, particularly in terms of validation methods. Drawing upon Balacheff's (1988) concepts of pragmatic and intellectual proofs, we analyze student approaches. Pragmatic proofs include naive empiricism, crucial examples, and generic examples, while intellectual proofs rely on formalized notions, definitions, or properties.

► An animated version of the item is available at https://timss2019.org/psi/ch3-robots-items/

A relevant strategy at this grade level is to test a sequence of numbers following a simple arithmetic pattern, such as 1, 2, and 3. This approach allows students to observe output changes and patterns as they input consecutive numbers. However, drawing conclusions based on a few cases may not be robust enough, necessitating a validation method to confirm any conjecture. The limitations of the table tool and available information restrict students' ability to conduct a formal mathematical demonstration. For instance, the linear nature of the relationship is not explicit.

In pragmatic proofs, naive empiricism involves verifying a statement's validity based on a few cases. The crucial example method tests a proposition on a significantly different case, under the assumption that if it holds true in this new case, it is likely universally true. In the Robots item, selecting a number markedly different from the initial sequence and using the calculator to verify the conjecture would constitute a crucial example. The generic example involves demonstrating validity reasons by performing operations on an object as a representative of a broader class. Testing the number 100 in the Robots item serves as a generic example, but the table tool's limitation of not allowing numbers greater than 999 restricts this method's application.

Intellectual proofs, in contrast, do not rely on experimenting but are intellectual constructs based on formalized mathematical concepts. Testing the input value of 0 in the Robot item provides undeniable evidence of the constant term in the linear relationship, exemplifying an intellectual proof.

The analysis utilizes the international log database of the Robots item, examining variables such as the number of tests, time spent, and the use of the provided calculator. These process indicators, derived through feature engineering, include testing an arithmetic sequence, tests with a range greater than 10, testing specific values like 100 and 0, and calculator usage. After data cleaning, 13,000 students from 22 countries who participated in eTIMSS 2019 were included. To prevent any country from being overrepresented, we created a weighting variable so that the sum of student weights is the same for all countries.

RESULTS

Figure 1: Bivariate statistics between the study strategy indicators and the item

Characteristic	Score = 0 (N = 10 127)[1]	Score = 1 (N = 2 797)[1]	Overall (N = 12 924)[1]	p-value[2]
Number of different tests	4 (2.43)	5 (2.43)	5 (2.44)	<0.001
Time on the item (min)	2.35 (1.92)	2.70 (1.99)	2.43 (1.94)	<0.001
Test 0 (YES)	11.4 %	28.7 %	15.1 %	<0.001
Test 100 (YES)	3.2 %	5.5 %	3.7 %	<0.001
Test an arithmetic sequence (YES)	81.8 %	90.3 %	83.7 %	<0.001
Range between 2 X > 10 (YES)	30.9 %	31.6 %	31.1 %	0.5
Used calculator (YES)	18.4 %	17.0 %	18.1 %	0.1
Sex of the student:				<0.001
Girls	51.3 %	45.2 %	50.0 %	
Boys	48.7 %	54.8 %	50.0 %	

Notes:
[1] p% [N]; median (sd)
[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test
Reading note: The mean number of different tests is 4 with a standard error of 2 for students who failed at the item (score coded 79 or 99).
Source: IEA – DEPP MENJ

This reveals statistically significant relationships with the majority of our variables except for the indicator of range between 2X values and the use of the calculator. This may be attributable to some students testing random numbers without adhering to any specific solving strategy.

It is important to note that a small number of students tested 100, resulting in a larger error for this variable. And finally, we see that the majority of students tested an arithmetic sequence as the foundation of their strategy.

Additionally, the study focuses on interactions between some of these variables based on the didactical hypothesis. Specifically, testing an arithmetic sequence and testing 0 are considered indicators of an intellectual proof. Testing an arithmetic sequence and having a range of tested numbers greater than 10 indicates a crucial example. Finally, testing an arithmetic sequence and testing 100 suggests a generic example.

We seek to determine whether there is a relationship between students' mathematical ability as measured by the PSI score (Martin et al., 2020) and the strategies used by students to solve this item. Linear regressions were used for this purpose.

The following table displays results of linear regressions between the PVs (plausible values) in mathematics and the selected variables of interest for each of the validation strategies.

Figure 2: Linear regressions between PVs in mathematics

	Crucial Example Model		Generic Example Model		Intellectual Proof Model	
	Estimator	*Std.Error*	*Estimator*	*Std.Error*	*Estimator*	*Std.Error*
Intercept	471.2***	3.5	461.3***	2.9	460.5***	2.9
Number of different tests	0.1	0.5	0.7	0.5	0.7	0.5
Time on the item (min)	0.3	0.5	0.2	0.5	0.2	0.5
Sex of the student (ref. Girls)	3.2	2.0	3.0	2.0	3.0	2.0
Used calculator	2.4	2.4	2.5	2.4	2.6	2.4
Test an arithmetic sequence : range between 2X>10						
NO:NO	Ref.	Ref.				
NO:YES	-20.8***	3.0				
YES:NO	51.0***	4.4				
YES:YES	54.0***	3.8				
Test an arithmetic sequence : Test 100						
NO:NO			Ref.	Ref.		
NO:YES			30.9***	10.0		
YES:NO			60.5***	2.5		
YES:YES			79.9***	5.8		
Test an arithmetic sequence : Test 0						
NO:NO					Ref.	Ref.
NO:YES					30.7***	9.3
YES:NO					59.8***	2.5
YES:YES					93.7***	3.3
Range between 2X>10 (ref. NO)			- 3.0	2.5	-3.0	2.3
Test 100 (ref. NO)	21.8***	4.8			21.8***	4.8
Test 0 (ref. NO)	33.5***	2.4	33.7***	2.4		

▸ **Notes**:
*** p-value<0.001
Source: IEA – DEPP MENJ

The number of tests conducted, and the time spent on the Robots item had minimal influence on the PSI PVs, as well as the sex of the student and the fact they used a calculator. However, when examining the effects of validation strategies, we observed patterns between the PSI PVs and the strategies employed by students. The combination of strategies used to solve the item exhibited a positive correlation with the students' overall PSI PVs.

The first validation model focuses on the crucial example. We discovered that testing very different numbers without also testing an arithmetic sequence is negatively correlated with the PSI PVs. It is likely that students who tested very different numbers do not have the mathematical ability to follow a problem-solving and validation strategy. On the other hand, students who tested an arithmetic sequence alongside at least one other random number show no significant difference from those who tested only an arithmetic sequence. However, these students have better mathematical ability.

The second model centered around the generic example. Our findings indicate that testing 100 in addition to an arithmetic sequence is more effective than testing solely an arithmetic sequence.

Lastly, the third model explores intellectual proof. Like the model involving the generic example, testing 0 alongside the arithmetic sequence shows a strong and high correlation with mathematical ability.

DISCUSSION

The findings of this study hold significant implications for policymakers in the realm of educational assessment and curriculum development. In an era where digital technologies are increasingly integrated into educational practices, understanding how students interact with and benefit from these technologies is crucial for informed decision-making. Policymakers are tasked with the responsibility of ensuring that educational strategies not only incorporate technological advancements but also effectively leverage them to enhance learning outcomes. This study sheds light on the cognitive processes underlying students' interactions with digital assessment tools, offering valuable insights for the development of policies that support effective and meaningful integration of technology in education.

Our research's validation models support our didactical hypothesis and align with our initial expectations. They reveal that pragmatic proofs, which are simpler and used earlier in learning, operate at a lower cognitive level compared to intellectual proofs. Intellectual proofs, on the other hand, involve applying mathematical properties and knowledge to make generalizations, indicating a more advanced level of understanding. In this hierarchy, the generic example emerges as a more sophisticated form of pragmatic proof, serving as a steppingstone toward intellectual proofs. It not only validates a conjecture based on specific cases but also aids in understanding and expressing the general underlying relationship, as Balacheff's framework suggests.

The eTIMSS PSI data corroborates this hierarchy. Our analysis shows that students who do not employ any validation method in addition to testing the arithmetic sequence tend to score lower. Specifically, the use of the crucial example strategy is associated with a 54-point increase in scores, the generic example with an 80-point increase, and the intellectual proof with a 94-point increase. These findings underscore the positive impact of employing validation strategies on students' performance in the PSI mathematics test, highlighting an area of potential focus for educational policy and curriculum development.

However, it is important to recognize the limitations of our research. Firstly, the "range of tries larger than 10" variable does not fully capture the essence of the crucial experiment strategy, suggesting a need for more nuanced feature engineering. Secondly, the interaction with the table tool limits students' ability to input large numbers, thereby constraining their capacity to test generic examples. Thirdly, drawing broad conclusions from responses to a single question is challenging due to the multitude of factors that can influence student performance. Lastly, the log data, while extensive, does not capture all meaningful actions undertaken by students, such as the use of external tools like paper for jotting down notes.

Despite these limitations, this research offers critical insights for policymakers. It highlights the importance of considering cognitive processes and validation strategies in the design of digital assessment tools and educational curricula. By understanding how students engage with and benefit from these tools, policymakers can make more informed decisions about integrating technology in education. This research underscores the need for policies that not only embrace technological advancements but also ensure that they are pedagogically sound and conducive to deeper learning and understanding. The release of technology-enhanced items and the associated findings emphasize the need for education systems worldwide to consider these insights in the development of instructional materials and assessment strategies.

REFERENCES

Balacheff, N. (1988). *Une étude des processus de preuve en mathématique chez des élèves de collège* [Habilitation à diriger des recherches, Institut National Polytechnique de Grenoble - INPG ; Université Joseph-Fourier - Grenoble I]. https://theses.hal.science/tel-00326426

Drijvers, P. (2019). Digital assessment of mathematics: Opportunities, issues and criteria. *Mesure et Évaluation en Éducation*, *41*(1), 41–66. https://doi.org/10.7202/1055896ar

Ercikan, K., Guo, H., & He, Q. (2020). Use of response process data to inform group comparisons and fairness research. *Educational Assessment*, *25*(3), 179–197. https://doi.org/10.1080/10627197.2020.1804353

Goldhammer, F., Kroehne, U., Hahnel, C., & De Boeck, P. (2021). Controlling speed in component skills of reading improves the explanation of reading comprehension. *Journal of Educational Psychology*, *113*(5), 861–878. https://doi.org/10.1037/edu0000655

Maddox, B. (2023). The uses of process data in large-scale educational assessments. *OECD Education Working Papers*, No. 286, OECD Publishing. https://doi.org/10.1787/5d9009ff-en

Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. https://timssandpirls.bc.edu/timss2019/methods/

Mullis, I. V. S., Martin, M. O., Fishbein, B., Foy, P., & Moncaleano, S. (2021). *Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks*. https://timssandpirls.bc.edu/timss2019/psi/

OECD (2023). *Shaping digital education: Enabling factors for quality, equity and efficiency*. Éditions OECD. https://doi.org/10.1787/bac4dc9f-en

Salles, F., Dos Santos, R., & Keskpaik, S. (2020). When didactics meet data science: Process data analysis in large-scale mathematics assessment in France. *Large-scale Assessments in Education*, *8*(1). https://doi.org/10.1186/s40536-020-00085-y

IEA COMPASS:
BRIEFS IN EDUCATION

About IEA

IEA is an independent, international cooperative of national research institutions and governmental agencies. IEA conducts large-scale comparative studies of educational achievement to; gain an in-depth understanding of the effects of policies and practices within and across education systems; to provide policymakers with insights into student performance; and to research, understand, and improve education worldwide.

Follow us:

@iea_education

IEAResearchInEducation

IEA

David Rutkowski
Series Editor,
Indiana University

Please cite this publication as:
Salles, F., & Lacroix, A. (2024, June). Using Process Data in Large-Scale Assessments: An Example With an eTIMSS Problem Solving and Inquiry Task. *IEA Compass: Briefs in Education* No 24. IEA

ISSN: 2589-70396

Copies of this publication can be obtained from:

IEA Amsterdam
Keizersgracht 311
1016 EE Amsterdam
The Netherlands

By email: secretariat@iea.nl
Website: www.iea.nl

IEA

Researching education, improving learning