AN INTERNATIONAL COMPARISON OF PERFORMANCE ON TIMSS ELEMENTARY MATHEMATICS ITEMS WITH POTENTIAL FOR COVARIATIONAL REASONING

Julien Corven
Illinois State University
jccorve@ilstu.edu

Teo Paoletti
University of Delaware
teop@udel.edu

Allison L. Gantt
University of Delaware
agantt@udel.edu

We previously (Gantt et al., 2023; Paoletti et al., 2021) identified items from the publicly released TIMSS 2011 assessments that had potential for students to employ covariational reasoning as a solution strategy. In this report, we explore the extent to which fourth-grade students' performance on such items in mathematics differed among 26 nations. Using multi-level modeling, we conclude that, in general, fourth-grade students were less successful on mathematics items for which covariational reasoning was a viable strategy than on items for which we could not identify a possible covariational reasoning strategy. However, three countries (Finland, Sweden, and the Netherlands) did not follow this pattern.

Keywords: Assessment, Elementary School Education, Problem Solving

Coordinating two dynamically changing quantities, or reasoning covariationally (Carlson et al., 2002), is a critical skill for students to develop across mathematics and science (e.g., Gantt et al., 2023; Panorkou & Germia, 2021; Paoletti et al., 2022; Sokolowski, 2020). Covariational reasoning can be closely tied to students' construction of mathematical representations (e.g., Confrey & Smith, 1995; Moore et al., 2013; Stevens, 2018; Wilkie, 2020) and to their making sense of particular types of quantitative relationships (e.g., Ellis et al., 2015; Johnson, 2015; Thompson & Thompson, 1996). Despite the growing body of research emphasizing the importance of covariational reasoning across mathematics and science domains (see Gantt et al., 2023 for a synthesis), the majority of the research exploring K-12 students' covariational reasoning has focused on small scale qualitative studies. In this report, we address the need to explore students' covariational reasoning quantitatively using a large data set from the Trends in International Mathematics and Science Study (TIMSS).

Since 1995, the TIMSS assessment has been regularly used to measure students' achievement in mathematics and science in 90 countries (NCES, n.d.). TIMSS data has been used for international comparisons of student achievement (e.g., Archibold, 1999; Chudgar et al., 2013; Mejía-Rodríguez et al., 2021; Wang et al., 2012) and has formed the basis for government reports and decision-making about STEM education in countries around the world (e.g., NCES, 2021; Richardson et al., 2020; Thomson et al., 2020).

The TIMSS assessment is a particularly useful resource to explore students' covariational reasoning as there are indications that such reasoning may be supported differently in various countries. For example, Thompson et al. (2017) explored a large sample of US and Korean teachers' meanings related to constructing graphs. They found large differences between the two countries in terms of teachers' tendencies to exhibit covariational reasoning on tasks that could elicit such reasoning. Thompson and Carlson (2017) also described differences in Japanese and American textbooks' approaches to support students' covariational reasoning (Thompson & Carlson, 2017). Hence, by using the TIMSS assessment, we address the need to explore between-country differences in students' covariational reasoning. The research question guiding this study is: *To what extent do differences exist within and between countries in fourth-grade students' performance on TIMSS mathematics items depending on the item's potential to elicit a covariational reasoning strategy?*

Covariational Reasoning in the TIMSS Assessment

In this report, we build on and extend our prior work (Gantt et al., 2023; Paoletti et al., 2021) where we conducted a content analysis of publicly released TIMSS items from the Grade 4 Mathematics, Grade 8 Mathematics, Grade 4 Science, and Grade 8 Science assessments based on their potential to elicit students' covariational reasoning. We defined an item as having the potential to elicit covariational reasoning (PCR) if we could "1) identify a way a student might conceive two changing quantities and 2) determine some solution strategy that could reasonably entail covariational reasoning" (Gantt et al., 2023, p. 6). For example, one multiple choice Grade 4 Mathematics item prompted: "The scale on a map indicates that 1 centimeter on the map represents 4 kilometers on the land. The distance between two towns on the map is 8 centimeters. How many kilometers apart are the two towns?" (IEA, 2013, p. 12). We classified this item as a PCR item because a student could coordinate one-centimeter changes in the map with four-kilometer changes on land to determine that eight such centimeter changes would result in 32 kilometers on land. This solution entails reasoning covariationally. However, a correct response to this item does not guarantee that a student employed covariational reasoning; for example, a student could have used a memorized procedure or randomly guessed the correct answer.

In Gantt et al. (2023), we categorized approximately one-third of all publicly released items as PCR across all four assessments. Particular to this paper, we coded 27 out of the 73 (37%) publicly released Grade 4 Mathematics items as PCR. We further organized the TIMSS items into content strands across the Grade 4 and Grade 8 Mathematics assessments. Table 1 presents the number of PCR items within each Grade 4 content strand. We also reported the number of PCR items by TIMSS-identified cognitive domain (Knowing, Applying, Reasoning; Mullis et al., 2009). In Grade 4 Mathematics, we identified 6 PCR Knowing items (out of 28, 21%), 10 PCR Applying items (out of 29, 35%), and 11 PCR Reasoning items (out of 15, 73%).

Table 1: Grade 4 Mathematics PCR Items as Percentage of Total Items by Content Strand

3. Content strand	4. Total items	5. PCR items	6. % (PCR/ Total)
7. Number (Whole numbers; fractions & decimals)	8. 29	9. 14	10. 48%
11. Algebra (Patterns & relationships; number sentences)	12. 11	13. 6	14. 55%
15. Geometry (2- & 3-dimensional shapes; points, lines & angles)	16. 24	17. 1	18. 4%
19. Statistics (Reading & interpreting; organizing & representing)	20. 9	21. 6	22. 67%

Methods

Our data corpus consisted of the TIMSS 2011 publicly available fourth-grade mathematics student performance data (TIMSS 2011 International Database, 2022). We restricted our investigation to data from 26 member nations of the Organization for Economic Cooperation and Development (OECD). Each code for each student's response to each item was converted to a binary score (correct or incorrect). Because we had only coded the potential for covariational reasoning (PCR) for publicly released TIMSS assessment items (Gantt et al., 2023), we removed performance data for all other (not released) items from the dataset. Students who completed forms of the TIMSS assessment that included no publicly released items were excluded. We considered the students who remained in the data set to be randomly selected from among all students who took the assessment, since the assignment of forms to students was randomized within each classroom. Due to the TIMSS sampling methodology and the distribution of publicly released fourth-grade mathematics items within forms of the assessment, students included in the data set contained performance data for at least half of the items presented to them.

We created multi-level (items nested within students) logistic regression models separately for each country. The unconditional (intercept-only) model gave the log-odds of an average student getting an average item correct. The conditional model incorporated the binary predictor of PCR for each item. In this second model, the intercept represented the log-odds of an average student getting an average non-PCR item correct, and the sum of the intercept and the coefficient for PCR represented the log-odds of an average student getting an average PCR-coded item correct. We converted all log-odds to probabilities in our results for ease of interpretation. Although a correct answer to a PCR item does not imply that a student employed covariational reasoning, we interpret a smaller probability within the PCR category as possibly indicating that an average student from that country was less likely to reason covariationally for such a problem.

Finally, for each country, we conducted an ANOVA test to compare the conditional model to the unconditional model to see for which countries the conditional model was a statistically significant improvement over the unconditional model. A statistically significant ANOVA test indicated that the conditional model significantly reduced the variance of the residuals from the data over the unconditional model, meaning that adding the PCR predictor significantly improved the model's predictive capabilities.

Results

Table 2 contains conditional model results (as probabilities) for 26 OECD member nations.

Table 2: Probability Results of Conditional Models for 26 OECD Member Nations

23. oun try	24. o n- P C R	25. C R	26. if f.	27. ountr y	28. o n- P C R	29. C R	30. iff .
31. ustr alia	32. 5 7 5	33. 5 2 5	34. .0 5 0 *	35. orea	36. 7 9 4	37. 7 7 3	38. .0 21 *

39. ustria	40. 574	41. 493	42. .081*	43. etherlands	44. 613	45. 642	46. 029*
47. elgium	48. 660	49. 606	50. .054*	51. ew Zealand	52. 493	53. 469	54. .024*
55. hile	56. 488	57. 421	58. .067*	59. orthern Ireland	60. 734	61. 679	62. .055*
63. zech Republic	64. 575	65. 540	66. .035*	67. orway	68. 529	69. 466	70. .063*
71. enmark	72. 617	73. 593	74. .024*	75. oland	76. 522	77. 471	78. .051*
79. ngland	80. 667	81. 614	82. .053*	83. ortugal	84. 664	85. 579	86. .085*
87. inland	88. 628	89. 624	90. .004	91. lovak Republic	92. 563	93. 485	94. .078*
95. ermany	96. 614	97. 572	98. .042*	99. lovenia	100. 581	101 480	102. .101*
103. ungary	104. 626	105 502	106 .124*	107. pain	108. 523	109 446	110. .077*
111. reland	112. 611	113 564	114 .047*	115. weden	116. 528	117 520	118. .008
119. taly	120. 580	121 474	122 .106*	123. urkey	124. 494	125 401	126. .093*

127. apan	128 736	129 705	130 .031*	131. nited States	132. 672	133 592	134. .080*

135. *Note:* * represents a statistically significant difference at $p < .001$. Lack of a * indicates the difference was not statistically significant at $p < .05$.

As Table 2 shows, for all the countries in our sample except Finland, the Netherlands, and Sweden, the probability of an average student getting a non-PCR item correct exceeds the probability of an average student getting a PCR item correct. For nearly all countries in the sample, the conditional model is an improvement over the unconditional model and the coefficient for PCR in the model is statistically significant ($p < .001$) and negative. The first result shows that analyzing performance data with the PCR status of an item better predicts student performance than analyzing the data without this distinction, whereas the second result demonstrates that students performed significantly better on non-PCR than PCR items.

As noted above, three countries do not follow this pattern. For Finland and Sweden, the PCR coefficients in the conditional models are not statistically significant, meaning there was no significant difference in average fourth-grade student performance on non-PCR and PCR items on the mathematics assessment. Unsurprisingly, the conditional model is also not a significant improvement over the unconditional model for Finland ($p = .377$) and Sweden ($p = .106$). Finally, for the Netherlands, the conditional model is a significant improvement over the unconditional model ($p < .001$), but the sign of the PCR coefficient is positive instead of negative. The positive coefficient indicates that the average fourth-grade student in the Netherlands was more likely to answer a PCR mathematics item correctly than a non-PCR item.

Discussion

Our results suggest that, in most of the countries for which we analyzed performance data, the average student had a lower probability of getting an average PCR item correct compared to a non-PCR item. Although prior research indicates that emphases on covariational reasoning in teaching and curricula vary from country to country (e.g., Japan and Korea compared to the US; Thompson & Carlson, 2017; Thompson et al., 2017), we found similar differences in how students in these three countries performed on PCR and non-PCR items. In other words, although the average student in Japan and Korea outperformed the average student in the US on both PCR and non-PCR items from the fourth-grade mathematics TIMSS assessment, there are significant gaps in performance between these item groups in all three countries. We notice, however, that the gaps are smaller in Japan and Korea than in the US, and we intend to conduct follow up analyses to determine whether the difference in gaps is statistically significant.

We note that Finland, Sweden, and the Netherlands, the three countries for which we found either no difference in performance on PCR and non-PCR items or better performance on PCR items than non-PCR items, are geographically close to each other. We hypothesize that differences in written curricula, teacher knowledge, or instructional practices might explain our findings. However, an important limitation of this investigation is that the quantitative analyses we conducted cannot explain differences, only document them. Further qualitative research will be needed to contextualize our findings and possibly explain why these differences occurred.

We emphasize that the analyses we present here are preliminary, and there are other possible explanations for our findings that are not yet addressed. For example, it may be that differences in performance in different content areas or in different cognitive domains are confounded with differences in performance on PCR items. We intend to conduct follow-up analyses using just the data from items in the Number and Algebra strands and incorporating cognitive domain as a separate predictor to address this question. We also will extend these analyses to Grade 4 Science and Grade 8 Mathematics and Science items to see if the differences we found persist. Finally, we focus on within-country differences in this report and only began to explore relative performance differences between countries by noting that some countries do not follow the expected pattern. Future analyses will extend and better contextualize our findings. In doing so, we hope to determine where students are succeeding in learning to reason covariationally and identify best practices from qualitative analyses of curricula and teaching.

References

Archibold, R. C. (1999). 10 states plan new math curriculum and exams. *New York Times*. Retrieved February 8, 2023 from https://archive.nytimes.com/www.nytimes.com/library/national/050699math-exams-edu.html

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education*, *33*(5), 352–378. https://doi.org/10.2307/4149958

Chudgar, A., Luschei, T. F., & Zhou, Y. (2013). Science and mathematics achievement and the importance of classroom composition: Multicountry analysis using TIMSS 2007. *American Journal of Education*, *119*(2), 295–316. https://doi.org/10.1086/668764

Confrey, J., & Smith, E. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education*, *26*(1), 66–86. https://doi.org/10.2307/749228

Ellis, A. B., Özgür, Z., Kulow, T., Williams, C. C., & Amidon, J. (2015). Quantifying exponential growth: Three conceptual shifts in coordinating multiplicative and additive growth. *The Journal of Mathematical Behavior*, *39*, 135–155. https://doi.org/10.1016/j.jmathb.2015.06.004

Gantt, A. L., Paoletti, T., & Corven, J. (2023). Exploring the prevalence of covariational reasoning across mathematics and science using TIMSS 2011 assessment items. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-023-10353-2

International Association for the Evaluation of Educational Achievement [IEA]. (2013). *TIMSS 2011 grade 4 released mathematics items*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA, and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands. https://nces.ed.gov/timss/pdf/TIMSS2011_G4_Math.pdf

Johnson, H. (2015). Secondary students' quantification of ratio and rate: A framework for reasoning about change in covarying quantities. *Mathematical Thinking and Learning*, *17*(1), 64–90. https://doi.org/10.1080/10986065.2015.981946

Mejía-Rodríguez, A. M., Luyten, H., & Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: An exploration of student and parent data of TIMSS 2015. *International Journal of Science and Mathematics Education, 19*(6), 1229–1250. https://doi.org/10.1007/s10763-020-10100-x

Moore, K, C., Paoletti, T., & Musgrave, S. (2013). Covariational reasoning and invariance among coordinate systems. *Journal of Mathematical Behavior*, *32*(3), 461–473. https://doi.org/10.1016/j.jmathb.2013.05.002

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. TIMSS & PIRLS International Study Center Lynch School of Education, Boston College. https://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Frameworks.pdf

National Center for Education Statistics [NCES]. (n.d.). *Participating countries: Trends in International Mathematics and Science Study (TIMSS)*. Retrieved February 20, 2023, from https://nces.ed.gov/timss/participation.asp

National Center for Education Statistics [NCES]. (2021). *TIMSS 2019 U.S. highlights web report*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/timss/results19/index.asp

Panorkou, N., & Germia, E. F. (2021). Integrating math and science content through covariational reasoning: The case of gravity. *Mathematical Thinking and Learning*, *23*(4), 318–343. https://doi.org/10.1080/10986065.2020.1814977

Paoletti, T., Gantt, A., & Corven, J. (2021). Items with potential for covariational reasoning across mathematics and science TIMSS assessments. In D. Olanoff, K. Johnson, & S. Spitzer (Eds.), *Proceedings of the forty-third*

annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 88–93). Philadelphia, PA.
Paoletti, T., Lee, H. Y., Rahman, Z., Vishnubhotla, M., & Basu, D. (2022). Comparing graphical representations in mathematics, science, and engineering textbooks and practitioner journals. *International Journal of Mathematical Education in Science and Technology*, *53*(7), 1815–1834. https://doi.org/10.1080/0020739X.2020.1847336
Richardson, M., Isaacs, T., Barnes, I., Swensson, C., Wilkinson, D., & Golding, J. (2020). *Trends in International Mathematics and Science Study (TIMSS) 2019: National report for England*. UCL Institute of Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/941351/TIMSS_2019_National_Report.pdf
Sokolowski, A. (2020). Developing covariational reasoning among students using contexts of formulas: (Are current formula notations in physics aiding graph sketching?). *The Physics Educator*, *2*(4), 2050016. https://doi.org/10.1142/S266133952050016X
Stevens, I. E. (2018). The parallelogram problem: Supporting covariational reasoning in the construction of formulas. In T. E. Hodges, G. J. Roy, & A. M. Tyminski (Eds.), *Proceedings of the 40th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 624–627). University of South Carolina & Clemson University.
Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, Part II: Mathematical knowledge for teaching. *Journal for Research in Mathematics Education*, *27*(1), 2–24. https://doi.org/10.2307/749194
Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 421–456). National Council of Teachers of Mathematics.
Thompson, P. W., Hatfield, N. J., Yoon, H., Joshua, S., & Byerley, C. (2017). Covariational reasoning among US and South Korean secondary mathematics teachers. *Journal of Mathematical Behavior*, *48*, 95–111. https://doi.org/10.1016/j.jmathb.2017.08.001
Thomson, S., Wernert, N., Rodrigues, S., & O'Grady, E. (2020). *TIMSS 2019 Australia highlights*. Australian Council for Educational Research. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=timss_2019
TIMSS 2011 International Database. (2022). Retrieved February 18, 2023 from https://timssandpirls.bc.edu/timss2011/international-database.html
Wang, Z., Osterlind, S. J., & Bergin, D. A. (2012). Building mathematics achievement models in four countries using TIMSS 2003. *International Journal of Science and Mathematics Education, 10*(5), 1215–1242. https://doi.org/10.1007/s10763-011-9328-6
Wilkie, K. J. (2020). Investigating students' attention to covariation features of their constructed graphs in a figural pattern generalisation context. *International Journal of Science and Mathematics Education, 18*(2), 315–336. https://doi.org/10.1007/s10763-019-09955-6