Chapter 5
Conclusions About Using TIMSS and TIMSS Advanced Data to Explore Student Misconceptions, Errors, and Misunderstandings in Physics and Mathematics

Abstract Assessment items from twenty years of TIMSS and TIMSS Advanced assessments enabled the identification of specific types of student misconceptions, errors, and misunderstandings related to two core concepts (gravity in physics and linear equations in mathematics). Results across grade levels, genders, and assessment years for five countries (Italy, Norway, the Russian Federation, Slovenia, and the United States) were compared. In physics, misconceptions and misunderstandings related to gravity were common across all five countries; for most misconceptions at each grade level, at least 25% of students demonstrated the misconception, and, in some countries, >50% of students demonstrated certain misconceptions. Errors and misunderstandings related to linear equations were extremely common across all five countries; on average >50% of students demonstrated errors at each grade level. Gender differences were found at all three grade levels, but to a greater extent in physics than in mathematics. Classroom teachers who are aware of the misconceptions or types of errors students may make will be able to plan for and provide additional support to their students when they are teaching these concepts. TIMSS resources can provide in-depth information about students' level of understanding, and their misconceptions and errors, across a range of core mathematics and science concepts. Access to released assessment items, scoring rationales, and actual student responses may allow researchers to undertake even richer secondary data analysis.

Keywords Errors · Gravity · International large-scale assessment · Linear equations · Mathematics · Misconceptions · Physics · Science · Student achievement · Trend analysis · Trends in International Mathematics and Science Study (TIMSS)

This report illustrates how item-level diagnostic data from TIMSS and TIMSS Advanced can be used to provide in-depth information about students' level of understanding, and specific types of misconceptions, errors, and misunderstandings, related to core physics and mathematics concepts across grade levels (specifically

T. Neidorf et al., *Student Misconceptions and Errors in Physics and Mathematics*, IEA Research for Education 9, https://doi.org/10.1007/978-3-030-30188-0_5

gravity and linear equations in this study). We (1) summarize the results across both physics and mathematics; (2) discuss limitations and further applications of our methodology; (3) consider implications related to instruction in physics and mathematics; and (4) describe some implications for future TIMSS assessment design and reporting.

5.1 Summary of Results Across Physics and Mathematics

The frequency of specific types of student misconceptions, errors, and misunderstandings related to gravity and linear equations at each grade level varied across the five countries included in the study: Italy, Norway, the Russian Federation, Slovenia, and the United States. We compare misconceptions, errors, and misunderstandings for both physics and mathematics by: (1) patterns in misconceptions, errors, and misunderstandings across countries and grade levels; (2) gender differences in misconceptions, errors, and misunderstandings; and (3) trends in misconceptions, errors, and misunderstandings over time (see Tables 4.1 and 4.21 for the specific codes used to refer to misconceptions, errors, and misunderstandings related to gravity and linear equations).[1]

5.1.1 Patterns in Misconceptions, Errors, and Misunderstandings Across Countries and Grades

In analyzing the patterns in student misconceptions, errors, and misunderstandings related to gravity and linear equations (Tables 5.1 and 5.2), we determined the average percentage of students with the misconception, error, or misunderstanding across the corresponding set of items.

In physics (Table 5.1), misconceptions and misunderstandings related to gravity were generally quite common across all five countries. For most misconceptions at each grade level, on average across items, at least 25% of students demonstrated the misconception, and, in some countries, at least 50% of all students demonstrated certain misconceptions. In TIMSS Advanced, misconceptions held by $\geq 50\%$ of students included P1B ("objects thrown upward have no acceleration at their maximum height") in Italy, P2 ("the time on the way up and the time on the way down are not equal") in both Italy and the Russian Federation, and P1C ("gravitational acceleration is always in the direction of motion/velocity") in the United States. At grade four, misconceptions held by $\geq 50\%$ of students included P3B ("gravity alone cannot cause an object initially at rest to start moving") in Italy, and misconception P4C ("gravity can make objects move in other directions that are not

[1]The codes are also defined in the Notes on the tables in Chap. 5.

'down' toward the surface of Earth") in Norway. In contrast, at grade eight there were no misconceptions demonstrated by $\geq 50\%$ of students in any country. There were three misconceptions (one at each grade level) where in all or nearly all countries <25% of students demonstrated the misconception: P1A ("gravitational force (acceleration) acting on objects near Earth's surface is not constant but changes with the height of the object above the surface") in TIMSS Advanced, P4A ("gravitational force causes objects to fall 'down' (in an 'absolute downward' direction in space) rather than toward the center of Earth") at grade eight, and P4B ("gravity pushes upward on objects sitting on a solid surface and on objects that are moving upward") at grade four.

In mathematics (Table 5.2), errors and misunderstandings related to linear equations were extremely common across all five countries; on average, >50% of students showed most types of errors at each grade level. Errors and misunderstandings with lower percentages of students across countries were M3B ("demonstrates confusion between slope and intercept") and M6 ("not able to translate relationships shown in table form into a mathematical equation") at grade eight, and M8 ("not able to identify a correct set of numbers that follow a given relationship/rule") at grade four.

5.1.2 Gender Differences in Misconceptions, Errors, and Misunderstandings

Gender differences in misconceptions, errors, and misunderstandings related to gravity (Table 5.3) and linear equations (Table 5.4) were found at all three grade levels, but to a greater extent in physics than in mathematics. In these summary exhibits, the percentages shown reflect the maximum female–male difference across the items measuring each misconception, error, or misunderstanding.

On average across the five countries, there were gender differences found for all but three gravity misconceptions or misunderstandings: P1C ("gravitational acceleration is always in the direction of motion/velocity") and P3A ("gravity acts only on falling objects, but not on objects at rest or moving upward") in TIMSS Advanced, and P4A ("gravitational force causes objects to fall "down" (in an "absolute downward" direction in space) rather than toward the center of Earth") at grade eight. In comparison, average gender differences were found for about half of the errors or misunderstandings related to linear equations. In physics, there were higher percentages of female students with the misconceptions related to gravity in all countries, with the exception of Italy for misconception P1B ("objects thrown upward have no acceleration at their maximum height where the instantaneous velocity is zero") in TIMSS Advanced, where the percentage of males was higher.

In mathematics, there were five types of errors or misunderstandings related to linear equations with significantly higher percentages of males in at least one country. This applied to four misunderstandings at grade eight: M3B ("demonstrates confusion between slope and intercept of an equation") in Italy and the

Table 5.1 Summary of physics misconceptions and misunderstandings related to gravity across items at each grade level, by country, 1995–2015

Misconceptions and misunderstandings	Items	Percentage of students with misconception or misunderstanding (averaged across items) (%)																	
		TIMSS Advanced						Grade 8						Grade 4					
		ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries
Performance objective 1: Determine the acceleration of thrown objects (after they are released)																			
P1A	TIMSS Advanced: Item 1A_V1	16	4	6	7	7	8												
P1B	TIMSS Advanced: Item 1A_V2	71	39	48	44	41	48												
P1C	TIMSS Advanced: Item 2	–	27	34	44	53	39												
Performance objective 2: Determine the time duration between different points on the path of a thrown object																			
P2	TIMSS Advanced: Item 1B	80	37	51	28	36	46												
Performance objective 3: Determine the effect of gravitational force on moving objects or on objects at rest																			
P3A	TIMSS Advanced: Item 3 Grade 8: Items 4, 5, 6, 7, 8	33	22	42	3	–	25	45	43	37	30	31	37						
P3B	Grade 8: Item 9 Grade 4: Items 10, 11							42	32	26	27	22	30	55	41	33	49	33	42
Performance objective 4: Identify the direction of the force due to gravity																			
P4A	Grade 8: Item 12							19	8	13	11	17	14						
P4B	Grade 4: Items 13, 14_V1													25	25	6	11	18	17
P4C	Grade 4: Items 14_V2, 15, 16													38	50	33	46	31	40

Key

≥ 50% of students

25.0–49.9 % of students

< 25% of students

Notes Physics misconceptions and misunderstandings: P1A = gravitational force (acceleration) acting on objects near Earth's surface is not constant but changes with the height of the object above the surface, P1B = objects thrown upward have no acceleration at their maximum height where the instantaneous velocity is zero (the instant it stops moving upward and reverses direction), P1C = gravitational acceleration is always in the direction of motion/velocity (rather than a constant acceleration directed toward the center of Earth), P2 = the time on the way up and the time on the way down are not equal (the downward acceleration due to gravity is not treated as constant), P3A = gravity acts only on falling objects, but not on objects at rest (on the ground or sitting on another surface) or on objects that are moving upward, P3B = gravity alone cannot cause an object initially at rest to start moving; it requires another force/push, P4A = gravitational force causes objects to fall "down" (in an "absolute downward" direction in space) rather than toward the center of Earth, P4B = gravity pushes upward on objects sitting on a solid surface and on objects that are moving upward, P4C = gravity can move objects in other directions that are not "down" toward the surface of Earth. ITA = Italy, NOR = Norway, RUS = Russian Federation, SVN = Slovenia, and USA = United States

– Data not available (see Appendix for country-specific notes)

Table 5.2 Summary of mathematics errors and misunderstandings related to linear equations across items at each grade level, by country, 1995–2015

Errors and misunderstandings	Items	Percentage of students with error or misunderstanding (averaged across items) (%)																	
		TIMSS Advanced						Grade 8						Grade 4					
		ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries
Performance objective 1: Interpret the solution to a system of linear equations to answer a question or solve a problem in real life contexts																			
M1	TIMSS Advanced: Item 1A	78	38	60	64	43	57												
Performance objective 2: Solve systems of linear equations in two variables																			
M2A	TIMSS Advanced: Item 1B G8: Item 2	75	57	64	61	60	63	78	76	75	68	61	72						
M2B	Grade 8: Items 3, 4, 5							87	88	60	83	69	77						
Performance objective 3: Interpret the meanings of slope and y-intercept in linear equations or graphs																			
M3A	Grade 8: Item 6							98	89	90	86	69	86						
M3B	Grade 8: Items 7, 8, 9							22	30	12	22	18	19						
Performance objective 4: Relate algebraic equations to their graphical representations (and vice-versa)																			
M4A	Grade 8: Item 10							79	90	52	89	56	73						
M4B	Grade 8: Items 11, 12							65	78	48	63	55	62						
Performance objective 5: Write equations to represent situations																			
M5	Grade 8: Items 13, 14, 15, 16							58	72	36	56	44	48						
Performance objective 6: Given pairs of numbers in tables or ordered pairs, generate an algebraic equation of the relationship between the two variables																			
M6	Grade 8: Items 17, 18, 19							44	43	37	30	31	37						
Performance objective 7: Given pairs of numbers in tables or ordered pairs, generate a verbal description of the relationship																			
M7A	Grade 8: Item 20							48	60	31	47	32	44						
M7B	Grade 4: Items 21, 22, 23													60	77	50	71	54	62
Performance objective 8: Given a verbal description of a relationship between a set of numbers, generate pairs of whole numbers that follow that relationship (rule)																			
M8	Grade 4: Items 24, 25, 26													40	49	26	42	36	39
Performance objective 9: Apply algebraic thinking to solve simple real-life problems involving unknowns																			
M9	Grade 4: Items 27, 28													74	60	58	73	68	66

Key
- ≥ 50% of students
- 25.0–49.9% of students
- < 25%

Notes Mathematics errors and misunderstandings: M1 = not able to use slope and intercept to provide an argument in support of the solution to a real-life problem situation, M2A = not able to apply the procedure correctly to solve a real-life problem situation, M2B = not able to apply the procedure correctly to solve non-contextualized problems, M3A = not able to relate slope with steepness of a line, M3B = demonstrates confusion between slope and intercept of an equation, M4A = not able to correctly identify the graph of an equation, M4B = not able to translate graphical representations into a mathematical equation or verbal description of a linear relationship, M5 = not able to translate verbal descriptions into a correct mathematical equation, M6 = not able to translate relationship shown in table form into a mathematical equation, M7A = not able to generate a correct verbal description given a specific relationship in the form of ordered pairs, M7B = not able to generate a correct verbal description given a specific relationship shown in table form, M8 = not able to identify a correct set of numbers that follow a given relationship/rule, M9 = not able to apply algebraic thinking to solve simple real-life problems involving unknowns. ITA = Italy, NOR = Norway, RUS = Russian Federation, SVN = Slovenia, and USA = United States

Table 5.3 Summary of gender differences in physics misconceptions and misunderstandings related to gravity across items at each grade level, by country, 1995–2015

Misconceptions and misunderstandings	Items	Percentage of students with misconception or misunderstanding (maximum female–male difference across items) (%)																	
		TIMSS Advanced						Grade 8						Grade 4					
		ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries
Performance objective 1: Determine the acceleration of thrown objects (after they are released)																			
P1A	TIMSS Advanced: Item 1A_V1	9	.	6	.	10	5												
P1B	TIMSS Advanced: Item 1A_V2	-9	20	.	17	16	9												
P1C	TIMSS Advanced: Item 2	—	.	.	.	.	.												
Performance objective 2: Determine the time duration between different points on the path of a thrown object																			
P2	TIMSS Advanced: Item 1B	13	15	16	.	14	13												
Performance objective 3: Determine the effect of gravitational force on moving objects or on objects at rest																			
P3A	TIMSS Advanced: Item 3 Grade 8: Items 4, 5, 6, 7, 8	.	.	.	.	—	.	21	.	13	.	8	7						
P3B	Grade 8: Item 9 Grade 4: Items 10, 11							.	.	.	9	.	4	11	11	.	.	11	7
Performance objective 4: Identify the direction of the force due to gravity																			
P4A	Grade 8: Item 12							.	.	.	.	.	.						
P4B	Grade 4: Items 13, 14_V1													.	.	.	.	8	3
P4C	Grade 4: Items 14_V2, 15, 16													.	18	9	17	8	11

Key

- Higher percentage of female students with the misconception or misunderstanding
- Higher percentage of male students with the misconception or misunderstanding
- . No statistically significant gender difference

Notes Physics misconceptions and misunderstandings: P1A = gravitational force (acceleration) acting on objects near Earth's surface is not constant but changes with the height of the object above the surface, P1B = objects thrown upward have no acceleration at their maximum height where the instantaneous velocity is zero (the instant it stops moving upward and reverses direction), P1C = gravitational acceleration is always in the direction of motion/velocity (rather than a constant acceleration directed toward the center of Earth), P2 = the time on the way up and the time on the way down are not equal (the downward acceleration due to gravity is not treated as constant), P3A = gravity acts only on falling objects, but not on objects at rest (on the ground or sitting on another surface) or on objects that are moving upward, P3B = gravity alone cannot cause an object initially at rest to start moving; it requires another force/push, P4A = gravitational force causes objects to fall "down" (in an "absolute downward" direction in space) rather than toward the center of Earth, P4B = gravity pushes upward on objects sitting on a solid surface and on objects that are moving upward, P4C = gravity can move objects in other directions that are not "down" toward the surface of Earth. ITA = Italy, NOR = Norway, RUS = Russian Federation, SVN = Slovenia, and USA = United States

– Data not available (see Appendix for country-specific notes)

Table 5.4 Summary of gender differences in mathematics errors and misunderstandings related to linear equations across items at each grade level, by country, 1995–2015

Errors and misunderstandings	Items	Percentage of students with error or misunderstanding (averaged across items) (%)																	
		TIMSS Advanced						Grade 8						Grade 4					
		ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries	ITA	NOR	RUS	SVN	USA	Average of countries
Performance objective 1: Interpret the solution to a system of linear equations to answer a question or solve a problem in real life contexts																			
M1	TIMSS Advanced: Item 1A	12	.	15	18	.	8												
Performance objective 2: Solve systems of linear equations in two variables																			
M2A	TIMSS Advanced: Item 1B G8: Item 2	9	.	12	19	.	7	.	.	.	.	.	.						
M2B	Grade 8: Items 3, 4, 5							.	.	.	5	.	4						
Performance objective 3: Interpret the meanings of slope and y-intercept in linear equations or graphs																			
M3A	Grade 8: Item 6							.	.	6	.	.	.						
M3B	Grade 8: Items 7, 8, 9							-5	.	.	.	-5	-4						
Performance objective 4: Relate algebraic equations to their graphical representations (and vice-versa)																			
M4A	Grade 8: Item 10							.	.	.	6	.	.						
M4B	Grade 8: Items 11, 12							.	10	11	.	.	4						
Performance objective 5: Write equations to represent situations																			
M5	Grade 8: Items 13, 14, 15, 16							.	.	–11	.	-6	.						
Performance objective 6: Given pairs of numbers in tables or ordered pairs, generate an algebraic equation of the relationship between the two variables																			
M6	Grade 8: Items 17, 18, 19							11	.	–10	.	.	.						
Performance objective 7: Given pairs of numbers in tables or ordered pairs, generate a verbal description of the relationship																			
M7A	Grade 8: Item 20							.	.	.	–10	.	.						
M7B	Grade 4: Items 21, 22, 23													.	11	.	.	.	.
Performance objective 8: Given a verbal description of a relationship between a set of numbers, generate pairs of whole numbers that follow that relationship (rule)																			
M8	Grade 4: Items 24, 25, 26													.	15	.	.	–7	.
Performance objective 9: Apply algebraic thinking to solve simple real-life problems involving unknowns																			
M9	Grade 4: Items 27, 28													.	.	.	9	.	5

Key

- Higher percentage of female students with the error or misunderstanding
- Higher percentage of male students with the error or misunderstanding
- . No statistically significant gender difference

Notes Mathematics errors and misunderstandings: M1 = not able to use slope and intercept to provide an argument in support of the solution to a real-life problem situation, M2A = not able to apply the procedure correctly to solve a real-life problem situation, M2B = not able to apply the procedure correctly to solve non-contextualized problems, M3A = not able to relate slope with steepness of a line, M3B = demonstrates confusion between slope and intercept of an equation, M4A = not able to correctly identify the graph of an equation, M4B = not able to translate graphical representations into a mathematical equation or verbal description of a linear relationship, M5 = not able to translate verbal descriptions into a correct mathematical equation, M6 = not able to translate relationship shown in table form into a mathematical equation, M7A = not able to generate a correct verbal description given a specific relationship in the form of ordered pairs, M7B = not able to generate a correct verbal description given a specific relationship shown in table form, M8 = not able to identify a correct set of numbers that follow a given relationship/rule, M9 = not able to apply algebraic thinking to solve simple real-life problems involving unknowns. ITA = Italy, NOR = Norway, RUS = Russian Federation, SVN = Slovenia, and USA = United States

United States, M5 (“not able to translate verbal descriptions into a correct mathematical equation”) in the Russian Federation and the United States, M6 (“not able to translate relationship shown in table form into a mathematical equation”) in the Russian Federation, and M7A (“not able to generate a correct verbal description given a specific relationship in the form of ordered pairs”) in Slovenia. There was also one misunderstanding at grade four (M8, “not able to identify a correct set of numbers that follow a given relationship/rule”) in the United States.

These results, based on item-level data, track what was found in scale score gender differences in the international reports from TIMSS and TIMSS Advanced in 2015 (Martin et al. 2016; Mullis et al. 2016a, b). In the five countries included in our study, males generally outperformed females in the relevant science content domains that covered the gravity topic in physics (mechanics and thermodynamics in TIMSS Advanced, physics at grade eight, and physical science at grade four). In contrast, there were fewer and smaller gender differences in the mathematics content domains that covered linear equations (algebra in TIMSS Advanced and grade eight, and number at grade four), and not all of these favored males. At grade eight, females scored higher than males for algebra in Italy, Slovenia, and the United States. Both in the item-level percentage of students with misconceptions, errors, or misunderstandings in this report, and in the subscale scores in the international reports, gender differences in both physics and mathematics were generally higher in TIMSS Advanced than at the lower grade levels. However, there were differences between physics and mathematics in the patterns of gender differences across grades in each country.

5.1.3 Trends in Patterns of Misconceptions, Errors, and Misunderstandings Over Time

The trend patterns across both physics (Figs. 4.23 and 4.24) and mathematics (Figs. 4.49 and 4.50) indicate some interesting differences over the assessment years in the frequency of misconceptions, errors, and misunderstandings demonstrated on items related to gravity and linear equations across countries and grade levels.[2]

Italy

There were very few measurable differences in the percentage of students with misconceptions, errors, and misunderstandings over time. Significant differences were found in mathematics for one item at grade four, where the frequency of misunderstanding M7B decreased between 2003 and 2015, and for one item at grade eight, where the frequency of misconception M3A increased slightly between 2011 and 2015. In physics, misconception P3A decreased in frequency between 2011 and 2015.

[2]The definitions for each misconception, error, or misunderstanding code referred to in this section can be found in Chap. 4 (Tables 4.1 and 4.21) and also in the notes in Tables 5.1 and 5.2.

Norway

There were no changes in the frequency of gravity misconceptions or misunderstandings at either grade four or grade eight. In mathematics, there was one item at grade four that showed a decrease in the frequency of misunderstanding M8 from 2007 to 2015. In contrast, there were two mathematics items at grade eight where the frequency increased for the types of errors and misunderstandings: in M4B (between 2007 and 2015) and M7A (between 1995 and 2003), and one item measuring misunderstanding M3A, where the frequency decreased between 2011 and 2015.

Russian Federation

Across grades and subjects, the greatest number of items showing trend differences was in the Russian Federation (10 items total). Most of the trend differences were in grade four, where the percentage of students with misconceptions related to gravity (three of three items, measuring misconceptions P3B and P4B) and misunderstandings related to linear equations (five of seven items, measuring misunderstandings M7B, M8 and M9) decreased over time. In physics, the frequency of misconception P3B also decreased at grade eight. The only case of an increase occurred in grade eight mathematics (misunderstanding M3A).

Slovenia

The number of items with trend differences were greater in grade eight than grade four for physics (two versus three items) and greater in grade four than grade eight (four versus three items) for mathematics. At grade eight, there were some significant decreases over time in the frequency of misconceptions, errors, or misunderstandings, and errors related to gravity (P3A and P3B) and linear equations (M3A and M5). In mathematics, however, there was an increase in misunderstanding M7A at grade eight between 1995 and 2003. At grade four, the frequency of misconceptions and misunderstandings decreased on one item in physics (measuring misconception P3B) and on four items in mathematics (measuring misunderstandings M7B, M8, and M9).

United States

The number of items with trend differences in the United States was the same as in the Russian Federation (10 items total in both countries). In contrast to the Russian Federation, however, the majority of trend differences in the United States were in grade eight mathematics, where the frequency decreased across assessment years on five items measuring errors or misunderstandings for M2B, M3A, M4A, M6, and M7A, and increased on one item (misunderstanding M4B). In physics, the frequency of misconceptions related to gravity were found to decrease on one item at grade eight (misconception P3A) and one item at grade four (misconception P3B).

5.2 Limitations and Further Applications of the Methodology

For our study, we used item-level data from the TIMSS international database (https://www.iea.nl/data) and, therefore, we were limited by the specific types of diagnostic data provided. In large-scale assessments like TIMSS, there is always a balance between the resources required for scoring, maintaining high reliability of scoring, and collecting diagnostic data that will provide information for tracking specific types of misconceptions, errors, and misunderstandings. Generally, for mathematics items there is a correct response and an incorrect response, with only a few items in the set that we used for our study being scored using a two-point scoring guide, where one point was given for a partial response. Similarly, there were only a few CR items worth one score point that used diagnostic scoring guides to track specific types of incorrect responses. In the case of physics items, there were slightly more CR items that used diagnostic scoring guides to track particular types of incorrect responses. For future studies similar to ours, more items with scoring codes that track the different types of errors that students make would be useful, particularly in mathematics.

The information produced by MC items is also limited by the guessing factor involved for such items. In general, for the same misconception, error, or misunderstanding, the percentage of students demonstrating the misconception or error may be higher for CR items than for MC items. The information provided by MC items could be enhanced if the distractors tracked important types of conceptual misunderstanding rather than the computational errors that students can make while solving the problem.

For the CR items, unless there was a specific diagnostic code to track particular misconceptions, errors, or misunderstandings, the reporting of more general misunderstandings and errors included all incorrect responses (including blanks). In doing this, we assumed that students who left the item blank did not know how to apply the concept or mathematical procedure in order to solve the problem, similar to other incorrect responses where students do not make an attempt at the item (e.g., random marks or off-task comments). However, it is difficult to know why students did not answer the item. Therefore, the percentage of students with misunderstandings or errors on these types of items may be inflated.

The TIMSS and TIMSS Advanced assessments are designed to provide reliable overall scores for science (or physics in TIMSS Advanced) and mathematics, and for each content domain. However, the sample sizes for the item-level statistics used in this report (percent correct and percentage of students demonstrating different types of misconceptions, errors, and misunderstandings) are relatively low.[3] As a result, many of the observed differences across countries, genders, and assessment years were not statistically significant. Also, as result of the booklet

[3]The number of students responding to each item is shown in the supplementary tables provided at www.iea.nl/publications/RfEVol9.

rotation scheme used in the TIMSS assessment design, only about one in every seven students get the same item; for TIMSS Advanced, about one in every three students get the same item.[4] This means a very small number of students in each class take the same item, which particularly affects the ability to report gender differences within countries.

To generalize beyond students' performance on individual items, a larger set of items that measure each type of misconception, error, or misunderstanding would be needed in each assessment cycle. In that case, "misconception indices," based on the average percentage of students with misconceptions across items, could be computed and tested for reliability in order to compare the frequency of these misconceptions on a broader range of items across countries and grade levels.

In addition, it would be interesting to follow a cohort of students to track the percentage of students with particular misconceptions, errors, and misunderstandings over time (e.g., students who were grade four in 2007, grade eight in 2011, and then, TIMSS Advanced in 2015). This would provide international data for understanding how students conceptualize a topic of interest as they progress through the grades and how similar or different the patterns in misconceptions, errors, and misunderstandings are across countries. Again, more items related to the topic of interest would be needed in each assessment cycle for a reliable measure.

While this report focused on specific types of misconceptions, errors, and misunderstandings related to the topics of gravity in physics and linear equations in mathematics, the general methodology that we describe can be applied to a range of science and mathematics topics covered in TIMSS and TIMSS Advanced to trace misconceptions, errors, and misunderstandings across two or three grade levels and better understand students' performance on those topics in science and mathematics. Another area that countries could continue to explore is the pattern of misconceptions, errors, and misunderstandings at one grade only, as was done in the United States for TIMSS Advanced (Provasnik et al. 2019). This could produce rich information about the misconceptions, errors, and misunderstandings that students at a specific grade have across different content domains.

We examined differences in misconceptions, errors, and misunderstandings by gender, but there are many other demographic variables available in TIMSS and TIMSS Advanced that could be analyzed. Countries could also look at differences by region, school type, or course type, as was done in the TIMSS Advanced report for the United States (Provasnik et al. 2019).

A better understanding of the misconceptions, errors, and misunderstandings over the assessment years could be achieved by investigating what is happening at the country level in the education system. A change in the curriculum, a change in the approach to teaching, or a change in the emphasis on the various types of learning strategies that could have resulted in a change in the pattern of

[4]For additional information on the assessment design for TIMSS 2015 and TIMSS Advanced 2015, please see https://timssandpirls.bc.edu/timss2015/downloads/T15_FW_Chap4.pdf and https://timssandpirls.bc.edu/timss2015-advanced/downloads/TA15_FW_Chap4.pdf, respectively.

misconceptions, errors, and misunderstandings made by students in different assessment years, merits further investigation. This kind of information, along with the methodology that we used for this report, could support teachers' and educators' efforts to improve instruction in the classroom. While it is beyond the scope of this report to explore curricular changes in the five different countries included in our study, further research could focus on this aspect. The TIMSS and TIMSS Advanced encyclopedias, teacher questionnaires, and country-level curriculum questionnaires, and results from the test curriculum matching analyses provide context for results from this type of study in terms of possible changes in policy, curriculum, or instruction across assessment cycles or grades (Martin et al. 2016; Mullis et al. 2016a, b, c, d). It should be noted, however, that any future research connecting curriculum changes to patterns and trends in the specific types of misconceptions, errors, and misunderstandings discussed in this report would likely require a more detailed analysis of curriculum documents from each country.

5.3 Implications Related to Instruction

In this report, we have discussed different types of misconceptions, errors, and misunderstandings related to gravity and linear equations that were demonstrated by TIMSS Advanced students in their final year of secondary school, and showed how these were connected to related misconceptions and a lack of foundational understanding about these concepts at grades four and eight. By identifying specific misconceptions, errors, and misunderstandings related to these core concepts, the findings from this type of study support the teaching, learning, and reinforcement of core concepts throughout school. Classroom teachers who are aware of the misconceptions or types of errors that students may make will be able to plan for and provide additional support to their students when they are teaching these concepts. Using released TIMSS and TIMSS Advanced items as additional resources may enable science and mathematics educators to identify misconceptions, develop pre-assessments, and provide focused instruction for their students.

In physics, our study showed that many TIMSS Advanced students still have difficulty understanding the effects of constant acceleration due to gravity on motion. The types of misconceptions related to gravity (and to forces and motion in general) described in previous smaller-scale studies across different grade levels were found to persist in the nationally representative TIMSS and TIMSS Advanced samples, including TIMSS Advanced students who had taken more advanced coursework in physics. In particular, it is of concern that many students in TIMSS Advanced across countries did not grasp the concept that the force (acceleration) due to gravity is a constant for thrown objects, instead indicating there was no acceleration at the maximum height and that acceleration was always in the direction of motion/velocity, rather than a constant acceleration directed toward the center of Earth. The misconception held by TIMSS Advanced students that acceleration due to gravity is not constant may arise from related misconceptions about the force of gravity at earlier grades.

The TIMSS data revealed that a lack of basic understanding of gravitational force at the lower grades can lead to misconceptions at higher grade levels, including the misconceptions that gravity acts only on falling objects, that gravity alone cannot cause an object initially at rest to start moving without another force/push, and that the force due to gravity is directed upward for an object at rest sitting on a surface or for objects that are moving upward.

Based on the types of gravity misconceptions found across grade levels, it is important for teachers at all grades to expose their students to a broad range of problem-solving contexts that will develop and evaluate their ability to apply their understanding of the concepts related to force and motion. In addition, pre-assessments and hands-on activities have been found to be important in identifying and addressing student misconceptions and developing their knowledge of forces (Darling 2012).

In mathematics, our report showed various conceptual stages where students have problems or make errors on the items involving linear equations that have been discussed in previous studies (Simon and Blume 1994; Stump 2001; Kalchman and Koedinger 2005; Caglayan and Olive 2010). These are the areas where focused instruction is needed for students to make the leap toward being well-versed in that concept. For example, one of the findings was that a higher percentage of students at grade eight were able to translate a graphical representation into a verbal description as opposed to an algebraic equation. This could mean that students are able to understand the relationship represented by the graph of a line, but they are not well-versed in the symbolic representation of a line, what each symbol means, and how they are related. Instruction needs to focus on these aspects, with an emphasis on understanding that goes beyond using equations to find the value of one variable when the other is given.

Similarly, students at each grade level find solving real-life problems more difficult than solving non-contextualized mathematics problems (item 1 in TIMSS Advanced, item 15 at grade eight, and items 24 and 25 at grade four). Students have difficulty solving real-life problems that require reading the context, understanding it, and then translating the problem into mathematics language to find what they need to do to solve the problem. Instruction across the grade levels needs to include more and different types of application problems that go beyond pure computation.

5.4 Implications for Future TIMSS Assessment Design and Reporting

While TIMSS is designed primarily to monitor system-level achievement trends in a global context, another important outcome of the study is the diagnosis of common learning difficulties in mathematics and science, as evidenced by misconceptions and errors (Mullis and Martin 2013a). Thus, TIMSS items and associated scoring guides are developed to allow identification of widespread student misunderstandings that, in turn, could lead to curricular or instructional improvements (Mullis and Martin 2013b). For example, TIMSS MC items use plausible distracters that are based on likely student errors or misconceptions.

CR items are scored using the TIMSS two-digit diagnostic scoring system, which allowed us to classify responses based on the method used in solving a problem, and track common errors or misconceptions. However, because scoring of CR items is a significant cost factor for the TIMSS countries, diagnostic scoring codes for specific response types are developed parsimoniously, such that only the codes with apparent value for educational improvement are included in the scoring guides (Mullis and Martin 2013b). As a result, the TIMSS item-level diagnostic data are limited to pre-defined distractors and diagnostic codes included to capture only the predominant correct and incorrect approaches/strategies used by students across all participating countries.

Despite this design restriction, our report demonstrated that access to specific TIMSS resources, namely released assessment items, CR item scoring guides, and item-level diagnostic data, can provide in-depth information about students' level of understanding and their misconceptions and errors across a range of core mathematics and science concepts. In addition to these critical TIMSS resources, future cycles of TIMSS may consider offering two additional resources: access to more complete scoring rationales for both CR and MC items, and actual student responses. Such resources would allow even richer secondary data analysis of mathematics and science concepts, and misconceptions, errors, and misunderstandings.

TIMSS items and scoring guides are developed with great care and thoughtfulness, with specific reasons for including each MC distractor item and each response code for the scoring guides of the CR items. Researchers would benefit greatly from having access to the rationales for the inclusion of specific distractors and specific response codes in TIMSS items.

Access to scoring rationales can be coupled with the potential benefits of eTIMSS, an electronic version of TIMSS. The 2019 administration of TIMSS begins the transition to administering the assessments in the eTIMSS digital format, allowing enhanced assessment of complex areas of the TIMSS framework that are difficult to measure with the paper-and-pencil format. In addition, eTIMSS will be able to capture students' actual responses to items in an easily accessible digital format. Traditionally, TIMSS provides access to achievement data files containing the actual responses to the MC items and the codes assigned to the CR items through the TIMSS scoring guides. Starting with the 2019 cycle, eTIMSS has the potential to provide access to a new international data file for students' responses that are captured via keyboard/number pad input. This new TIMSS resource has high value for researchers, since it potentially provides even deeper insights into what students know and are able to do, including common misconceptions, errors, and misunderstandings.

As discussed in Sect. 5.2, a more focused effort on providing diagnostic outcomes from TIMSS would require the inclusion of a larger number of items at each grade level that measure certain core concepts and misconceptions of interest. Also, sets of items related to a particular concept would need to be kept secure and administered in multiple assessments in order to track trends in students' understanding and how their misconceptions about concepts develop or vary over time.

The TIMSS and TIMSS Advanced assessments cover the framework objectives in each content domain with enough items to permit subscale reporting. However, each individual topic is measured by a small number of items distributed across the assessment booklets. Since each booklet includes only a portion of the total item pool, only a small subset of students in each country are likely to take items related to a particular topic. Therefore, while scores are provided at the content domain level, it is not possible to obtain reliable student-level data on a set of items that measure a particular topic within a content domain. To provide the best diagnostic information, students would have to take multiple items related to a specific topic in a single assessment (not possible with the current assessment design) in order to generalize beyond performance on individual items. One possible way to accomplish this would be to select one topic to explore in more depth and develop a block of 10–15 items that measure particular types of misconceptions, errors, and misunderstandings related to this topic. These special item blocks would be administered to a subset of students in the national samples, providing enough student-level data to support diagnostic reporting of the selected topic.

As also discussed in Sect. 5.2, it would be interesting to follow the same cohort of students across grade levels to track how their conceptual understanding of a concept develops with schooling over the years. TIMSS has a "quasi-longitudinal" design that permits this type of study, with the grade four and grade eight assessments being conducted every four years (see https://www.iea.nl/timss). However, in order to track the patterns of misconceptions, errors, and misunderstandings across grade levels, a change would be needed in the assessment design to include a block of cross-grade items (or a related block of items at each grade level) that measure a particular topic in consecutive assessment cycles. TIMSS Advanced has been administered less often than TIMSS,[5] so measuring the same cohort of students from grade four to the final year of secondary school would require putting TIMSS and TIMSS Advanced on the same assessment schedule. Even if a cohort is not tracked across all three grade levels though, monitoring the frequency of

[5]The 2015 assessment year was unusual in that all three assessments were administered, and there are data available for the same cohort of students (e.g., 2007 grade 4, 2011 grade 8, and 2015 TIMSS Advanced).

misconceptions, errors, and misunderstandings related to one topic of interest between grade four and grade eight could be a useful addition for future TIMSS cycles.

References

Caglayan, G., & Olive, J. (2010). Eighth grade students' representations of linear equations based on a cups and tiles model. *Educational Studies in Mathematics, 74*(2), 143–162.

Darling, G. (2012). How does force affect motion? *Science and Children, 50*(2), 50–53.

Kalchman, M., & Koedinger, K. R. (2005). Teaching and learning functions. In M. S. Donovan, & J. D. Bransford (Eds.), *How students learn: History, mathematics, and science in the classroom* (pp. 351–393). Washington, DC: National Academies Press.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: http://timssandpirls.bc.edu/timss2015/international-results/.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013a). *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html.

Mullis, I. V. S., & Martin, M. O. (2013b). *TIMSS 2015 item writing guidelines.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://timssandpirls.bc.edu/publications/timss/T15_item_writing_guidelines.pdf.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: http://timssandpirls.bc.edu/timss2015/international-results/.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016b). *TIMSS Advanced 2015 international results in advanced mathematics and physics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: http://timssandpirls.bc.edu/timss2015/international-results/advanced/.

Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (Eds.) (2016c). *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: http://timssandpirls.bc.edu/timss2015/encyclopedia/.

Mullis, I. V. S., Martin, M. O., & Loveless, T. (2016d). *20 years of TIMSS: International trends in mathematics and science achievement, curriculum, and instruction* (pp. 23–37). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: http://timss2015.org/timss2015/wp-content/uploads/2016/T15-20-years-of-TIMSS.pdf.

Provasnik, S., Malley, L., Neidorf, T., Arora, A., Stephens, M., Balestreri, K., Perkins, R., & Tang, J. H. (2019, in press). *U.S. performance on the 2015 TIMSS Advanced mathematics and physics assessments: A closer look* (NCES 2017-020). Washington, DC: US Department of Education, National Center for Education Statistics.

Simon, M. A., & Blume, G. W. (1994). Mathematical modeling as a component of understanding ratio-as-measure: A study of prospective elementary teachers. *Journal of Mathematical Behavior, 13,* 183–197.

Stump, S. L. (2001). High school precalculus students' understanding of slope as measure. *School Science and Mathematics, 101*(2), 81–89.