

Chapter 3

Methodology Used to Analyze Student Misconceptions, Errors, and Misunderstandings in TIMSS



Abstract The Trends in International Mathematics and Science Study (TIMSS) and TIMSS Advanced assessments are a good source of data for the study of student misconceptions, errors, and misunderstandings in physics and mathematics. After examining the available range of TIMSS and TIMSS Advanced data, five countries that participated in the TIMSS 2015 and TIMSS Advanced 2015 assessments, and all, or most, of the prior TIMSS assessments, were selected for study (Italy, Norway, Slovenia, the Russian Federation, and the United States) to maximize the cross-country comparisons that could be made across grade levels and assessment years. A complete review of the TIMSS and TIMSS Advanced assessment frameworks and content mapping (to determine related topics and items across grades and assessment cycles) identified the set of items that measure misconceptions, errors, and misunderstandings related to the topics of gravity and linear equations. Item-level statistics (the percentage of students who provided the correct answer, and the percentage demonstrating the misconception, error, or misunderstanding) were used to make comparisons across countries at each grade level overall and by gender. In addition to analyzing gender differences, examining trends in patterns of misconceptions, errors, and misunderstandings over time provided important information across countries.

Keywords Diagnostic data • Errors • International large-scale assessment • Item statistics • Misconceptions • Student achievement • *t*-test • Trend analysis • Trends in International Mathematics and Science Study (TIMSS) • Italy • Norway • Russian Federation • Slovenia • United States

3.1 TIMSS and TIMSS Advanced Data

TIMSS and TIMSS Advanced assessments have been measuring trends in international mathematics and science achievement since 1995, based on nationally representative samples of students in each participating country at grade four, grade eight, and the final year of secondary school (for students taking advanced

coursework in physics and mathematics). TIMSS has been administered every four years for six assessment cycles¹ (namely in 1995, 1999, 2003, 2007, 2011, and 2015), while TIMSS Advanced has been administered at three points in time (1995, 2008, and 2015). Following the release of the international reports from each assessment, the IEA releases international databases for secondary analyses. In addition, after each assessment, a portion of the assessment items (and scoring guides) are released, while at least half are retained as secure items for future assessment cycles. In both assessments, items may be released after one, two, or three assessment cycles.

This report used assessment items and student performance data from the TIMSS and TIMSS Advanced assessments conducted across all assessment cycles from 1995 to 2015. The set of countries administering TIMSS and TIMSS Advanced varies for each assessment cycle. We report on five countries that participated in the TIMSS Advanced 2015 assessment and in all, or most, of the TIMSS grade eight and grade four mathematics and science assessments since 1995: Italy, Norway, the Russian Federation, Slovenia, and the United States (see Tables 3.1, 3.2, and 3.3). These countries were selected from the nine countries participating in TIMSS Advanced 2015 (Table 3.1) to maximize the data available to answer the research questions (see Sect. 1.4). All selected countries participated at all three grade levels

Table 3.1 Participation of countries in TIMSS Advanced assessments, by cycle

Country	TIMSS Advanced					
	Advanced mathematics			Physics		
	1995	2008	2015	1995	2008	2015
France	•	–	•	•	–	•
Italy	•	•	•	–	•	•
Lebanon	–	•	•	–	•	•
Norway	–	•	•	•	•	•
Portugal	–	–	•	–	–	•
Russian Federation	•	•	•	•	•	•
Slovenia	•	•	•	•	•	•
Sweden	•	•	•	•	•	•
United States	•	–	•	•	–	•

• Indicates participation in that assessment cycle

– Indicates no participation

Notes This table includes the nine countries that participated in the TIMSS Advanced 2015 assessment. Five countries were selected for inclusion in this study (Italy, Norway, the Russian Federation, Slovenia, and the United States). See Appendix for TIMSS Advanced data considerations for 1995, 2008, and 2015

Source TIMSS Advanced 1995, 2008, and 2015 assessments. Copyright International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from Mullis et al. (2016a, Appendix MA.1 and Appendix PA.1)

¹In 1999, the TIMSS assessment was only administered at grade eight.

Table 3.2 Participation of countries in TIMSS grade eight assessments, by cycle

Country	TIMSS grade 8					
	1995	1999	2003	2007	2011	2015
France	●	–	–	–	–	–
Italy	○	●	●	●	●	●
Lebanon	–	–	●	●	●	●
Norway	●	–	●	●	●	●
Portugal	●	–	–	–	–	–
Russian Federation	●	●	●	●	●	●
Slovenia	●	○	●	●	●	●
Sweden	●	–	●	●	●	●
United States	●	●	●	●	●	●

- Indicates participation in that assessment cycle
- Indicates participation but data not comparable for measuring trends to 2015
- Indicates no participation

Notes This table includes the nine countries that participated in the TIMSS Advanced 2015 assessment. Five countries were selected for inclusion in this study (Italy, Norway, the Russian Federation, Slovenia, and the United States). See Appendix for TIMSS grade eight data considerations for 1995, 1999, and 2015

Source TIMSS 1995, 1999, 2003, 2007, 2011, and 2015 assessments. Copyright International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from Mullis et al. (2016b, Appendix A.1)

in the 2015 assessments and were missing data for no more than one assessment cycle at any grade level. The five selected countries thus permit the greatest number of comparisons across countries, grade levels, and assessment cycles.

TIMSS assesses mathematics and science achievement at two grade levels and so has two target populations: all students enrolled in grade four and all students enrolled in grade eight (or the equivalent grades in each country). The TIMSS Advanced physics and mathematics populations are defined as students in their final year of secondary school who are currently taking (or who had previously taken) the TIMSS Advanced-eligible courses in physics or advanced mathematics² (Martin et al. 2014). (More information is provided about the TIMSS and TIMSS Advanced populations in Appendix.)

The TIMSS Advanced population is a select group reflecting one-quarter or less of final-year students in most countries in 2015. The coverage index (percentage of the corresponding age cohort covered by the TIMSS Advanced physics and advanced mathematics student populations) was lower in physics than in advanced mathematics in all five countries included in the study (Tables 3.4 and 3.5). For physics, the coverage index in 2015 ranged from about 5% coverage in the Russian Federation and the United States, to 18% in Italy (Table 3.4). There were some

²TIMSS Advanced-eligible courses are defined as those that cover most of the topics outlined in the TIMSS Advanced physics and mathematics assessment frameworks.

Table 3.3 Participation of countries in TIMSS grade four assessments, by cycle

Country	TIMSS grade 4				
	1995	2003	2007	2011	2015
France	–	–	–	–	•
Italy	◦	•	•	•	•
Lebanon	–	–	–	–	–
Norway	•	•	•	•	•
Portugal	•	–	–	•	•
Russian Federation	–	•	•	•	•
Slovenia	•	•	•	•	•
Sweden	–	–	•	•	•
United States	•	•	•	•	•

• Indicates participation in that assessment cycle

◦ Indicates participation but data not comparable for measuring trends to 2015

– Indicates no participation

Notes This table includes the nine countries that participated in the TIMSS Advanced 2015 assessment. Five countries were selected for inclusion in this study (Italy, Norway, the Russian Federation, Slovenia, and the United States). See Appendix for TIMSS grade four data considerations for 1995 and 2015. TIMSS was not administered at grade four in 1999

Source TIMSS 1995, 2003, 2007, 2011, and 2015 assessments. Copyright International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from Mullis et al. (2016b, Appendix A.1)

differences in the physics coverage index across the assessment years, with increases seen in Italy, the Russian Federation, and the United States, and decreases seen in Norway and Slovenia between 2015 and 2008 or 1995. In particular, the percentage of students studying physics at an advanced level in Italy increased from 4% in 2008 to 18% in 2015, which indicates a progressively more inclusive sample of students. In contrast, the percentage of students decreased in Slovenia from 39% in 1995 to 7–8% in 2008 and 2015, reflecting a more restricted sample of students. For advanced mathematics, the coverage index in 2015 ranged from 10 to 11% in the Russian Federation and the United States, to 34% in Slovenia (Table 3.5; see Appendix for any additional data considerations between 1995 and 2015).

In addition to the overall coverage index, the percentages of female and male students in the TIMSS Advanced populations (final-year students taking advanced coursework in physics or mathematics) varied across countries and may differ from the percentages in the full population of students in their final year of secondary school. Boys were more likely to undertake advanced physics coursework than girls in all five countries (Table 3.4); only about 30% of advanced physics students in Norway and Slovenia, and about 40% of students in Italy, the Russian Federation, and the United States were female. The percentage of females in physics did not change substantially across assessment years in any country. In contrast to physics, the percentage of female students taking advanced mathematics (Table 3.5) was lower than males in Italy and Norway (about 40%), higher than males in Slovenia

Table 3.4 Coverage index and percentages of female and male students in TIMSS Advanced physics samples, by country and year

Country	Coverage index (%) ¹						Percentage of students ²											
	2015			2008			1995			2015			2008			1995		
	Female			Male			Female			Male			Female			Male		
Italy	18.2	3.8	–	46 (1.1)	40 (2.4)	–	54 (1.1)	60 (2.4)	–	74 (1.8)	74 (1.8)	–	–	–	–	–	–	
Norway	6.5	6.8	8.4	29 (1.2)	29 (1.7)	26 (1.8)	71 (1.2)	71 (1.7)	26 (1.8)	74 (1.8)	74 (1.8)	26 (1.8)	71 (1.7)	71 (1.7)	74 (1.8)	74 (1.8)	74 (1.8)	
Russian Federation	4.9	2.6	1.5	42 (1.2)	45 (1.3)	46 (2.0)	58 (1.2)	55 (1.3)	46 (2.0)	54 (2.0)	54 (2.0)	46 (2.0)	55 (1.3)	55 (1.3)	54 (2.0)	54 (2.0)	54 (2.0)	
Slovenia	7.6	7.5	38.6	30 (1.7)	27 (1.2)	28 (3.7)	70 (1.7)	73 (1.2)	28 (3.7)	72 (3.7)	72 (3.7)	28 (3.7)	73 (1.2)	73 (1.2)	72 (3.7)	72 (3.7)	72 (3.7)	
United States	4.8	–	2.7	39 (1.6)	–	43 (4.7)	61 (1.6)	–	43 (4.7)	57 (4.7)	57 (4.7)	43 (4.7)	–	–	–	–	–	

Notes

– Indicates no participation

¹ The physics coverage index is the percentage of the corresponding age cohort covered by the TIMSS Advanced target population (students in their final year of secondary school who were taking or had previously taken TIMSS Advanced-eligible physics courses)

² Standard errors are provided in brackets

Source TIMSS Advanced 1995, 2008, and 2015 assessments. Copyright International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from Mullis et al. (2016a, Exhibits P1.4 and P1.7)

Table 3.5 Coverage index and percentages of female and male students in TIMSS Advanced 2015 mathematics samples, by country

Country	Coverage index (%) ¹	Percentage of students ²	
		Female	Male
Italy	24.5	37 (1.3)	63 (1.3)
Norway	10.6	38 (1.4)	62 (1.4)
Russian Federation	10.1	50 (1.3)	50 (1.3)
Slovenia	34.4	60 (1.1)	40 (1.1)
United States	11.4	49 (0.9)	51 (0.9)

Notes Data for 2015 only, since there were no TIMSS Advanced mathematics items related linear equations from TIMSS 1995 or TIMSS 2008 included in the study

¹The mathematics coverage index is the percentage of the corresponding age cohort covered by the TIMSS Advanced target population (students in their final year of secondary school who were taking or had previously taken TIMSS Advanced-eligible mathematics courses)

²Standard errors are provided in brackets. Because of rounding some results may appear inconsistent

Source TIMSS Advanced 2015 assessment. Copyright International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from Mullis et al. (2016a, Exhibits M1.4 and M1.7)

(about 60%), and about equal to males in the Russian Federation and the United States.

All results in this report are based on item-level statistics available using the TIMSS and TIMSS Advanced international databases from each assessment cycle, including the weighted percent correct for each country and the percentage of students in each item response category (see Sect. 3.2.3). Item-level statistics were computed for each country, as well as on average across the five countries included in the study (overall and by gender). Example items used in this report include “restricted-use” items³ from the TIMSS 2015 assessments, as well as released items from prior assessment years. Although all example items are released or restricted-use items, appropriate non-released (secure) items from TIMSS 2015 were included in the analyses of patterns in misconceptions, but are not shown in the report.

³The 2015 “restricted-use” items are those designated by the TIMSS & PIRLS International Study Center for use as examples in the international report as well as by participating countries in their national reports or for research purposes, such as this IEA thematic report. Example items from 2015 included in this report are used with permission from the IEA. Secure items from 2015 are discussed but are not shown in the report.

3.2 Methodology

Our methodology consisted of three major components: (1) assessment framework review and content mapping to identify the set of items measuring the selected topics in our study (gravity and linear equations); (2) evaluation of diagnostic item-level performance data to identify the specific performance objectives measured by these items and to provide evidence of specific types of misconceptions, errors, and misunderstandings; and (3) analyses of the percentage of students demonstrating these misconceptions, errors, and misunderstanding to report patterns across countries by grade level, gender, and assessment year.

3.2.1 *Assessment Framework Review and Content Mapping*

To determine how mathematics and science concepts progress from the lower grades in TIMSS to TIMSS Advanced, topics covered in the 2015 TIMSS Advanced assessment frameworks were mapped to related topics at grades four and eight in the TIMSS 2015 frameworks. In the TIMSS and TIMSS Advanced frameworks, the greatest degree of content overlap across grades four, eight, and 12 is in the physics topic area of mechanics (forces and motion) and the mathematics content area of algebra, resulting in adequate numbers of assessment items across grades to report on patterns of misconceptions. Within topics, a set of framework objectives were identified at each grade level that were then used to select the items used in the study.

As described in Chap. 1, this study focuses on two specific topics: gravity in physics and linear equations in algebra. We determined the set of TIMSS 2015 and TIMSS Advanced 2015 framework objectives that measured these topics (or precursor topics) across grade levels for gravity (Table 1.1) and linear equations (Table 1.2). Since the TIMSS and TIMSS Advanced frameworks have been revised over the past 20 years, content mapping also included mapping the TIMSS framework objectives in 1995, 1999, 2003, 2007, and 2011, and the TIMSS Advanced framework objectives in 1995 and 2008, to the corresponding TIMSS 2015 framework objectives.

3.2.2 *Evaluation of Item-Level Performance Data*

Once the specific TIMSS and TIMSS Advanced framework objectives related to gravity and linear equations were identified, sets of items for each topic (16 for physics and 28 for mathematics) from the grade four, grade eight, and TIMSS Advanced assessments were assembled and reviewed. First, the TIMSS Advanced

2015 items were evaluated to determine the performance objectives measured by each item and the specific types of misconceptions, errors, and misunderstandings demonstrated by students across the five TIMSS Advanced countries chosen for the study (Italy, Norway, the Russian Federation, Slovenia, and the United States).⁴ Then, TIMSS items from across the assessment cycles at grades four and eight that measured related or precursor concepts were evaluated for evidence of specific misconceptions, errors, and misunderstandings at the lower grade levels.⁵

Evidence of misconceptions, errors, and misunderstandings was determined by examining patterns in the item-level performance data. For multiple-choice (MC) items, this involved distractor analysis, or examining the incorrect options to determine common errors and misconceptions that may be demonstrated by students who choose those options. For constructed-response (CR) items (where students provide a written response), response patterns were determined based on the nature of student responses as defined in the scoring guides that accompany the items. In TIMSS and TIMSS Advanced, scoring guides provide item-specific criteria to differentiate between correct, partial, and incorrect student responses and use two-digit diagnostic codes to track specific misconceptions or errors (i.e., to differentiate between different types of partial and incorrect responses). This initial item evaluation used item statistics (i.e., the weighted percentage distributions of students in each country choosing each MC response option or each CR item response category in the scoring guide) obtained from the international data almanacs available on the TIMSS & PIRLS International Study Center website (<https://timssandpirls.bc.edu/>).

Further content analysis of the set of items covering the topics of gravity and linear equations at each grade level identified a set of performance objectives (four in physics and nine in mathematics) that were measured by these items across the grade levels. These performance objectives are based on the set of TIMSS and TIMSS Advanced items selected for the study and are more specific than the broader TIMSS and TIMSS Advanced framework objectives outlined in Chap. 1. Some performance objectives were assessed at only one grade level, while others were measured by items at two grade levels (i.e., TIMSS Advanced/grade eight or grade eight/grade four) or at all three grade levels (for physics only). For the items measuring each performance objective, we identified the misconceptions, errors, or misunderstandings that may be demonstrated by different types of incorrect student responses. There were from one to six items measuring each type of misconception,

⁴Additional TIMSS Advanced items from 1995 and 2008 were also evaluated for physics. Mathematics only included items from TIMSS Advanced 2015.

⁵The TIMSS testing schedule permits the same cohort of students to be assessed over time (e.g., grade four students in 2007 are grade eight students in 2011, and grade 12 students in 2015). This report does not directly measure changes in specific misconceptions, errors, and misunderstandings over time for the same cohort of students due to limitations in the available item-level data. This raises some potential considerations and implications for future research in this area (see Sect. 5.4).

error, and misunderstanding. (See Sect. 1.2 for detailed definitions of the terms, and Chap. 4 for an overview of performance objectives, misconceptions, errors, and misunderstandings, and the set of items used in the study.)

3.2.3 Reporting Patterns in Percent Correct and Percent with Misconceptions, Errors, and Misunderstandings by Grade, Country, Gender, and Assessment Year

All of the analyses used to report on the percent correct and percentage of students with misconceptions, errors, and misunderstandings were conducted using the IEA's International Database (IDB) Analyzer (Version 4.0) *Percentages* function (IEA 2018). The IDB Analyzer uses a jackknife repeated replication (JRR) procedure to compute estimates and standard errors for a variety of statistics, such as average scores and percent correct (see Appendix for further technical notes). We do not provide standard errors in the tables and figures in this book (supplementary materials providing standard errors for all estimates are available for download at www.iea.nl/publications/RfEVol9).

Four types of analyses were used to produce the item-level statistics shown in the report.

Percent Correct

This is the percentage of students receiving credit on each item. For MC and short CR items (each worth one score point), this reflects the percentage of students who provided a correct answer. For extended CR items, this reflects the weighted percentage of students receiving full credit (two points) or partial credit (one point). For example, on an item where 10% of students received full credit and 10% received partial credit, the weighted percent correct is 15%, which reflects the percentage of students receiving full credit (10%) plus half the percentage receiving partial credit (5%). Percent correct was computed for all items in each country (overall and by gender). When reporting percent correct on the set of items in physics and mathematics, data from the most recent assessment was used for each item.

Percentage of Students with Misconceptions, Errors, and Misunderstandings

Two different types of item-level analyses were used to determine these percentages:

- (1) Specific types of misconceptions and misunderstandings reflected items where a single response option (in the case of MC items) or a single scoring guide category (in the case of CR items), or multiple response options or multiple scoring guide categories were identified to track and report on a particular type of misconception or misunderstanding. The percentage of students with the specific misconception (in physics) or misunderstanding (in mathematics) was calculated as the sum of the percentages of students in each of the relevant options or score categories. Specific misconceptions and misunderstandings apply to 11 items in physics (10 MC and one CR) and three items in

mathematics (all MC). For two of the MC items in physics, one response option measured one type of misconception and others measured a second type; two separate analyses were conducted to obtain the percentages for both types of misconceptions.

- (2) General types of misunderstandings reflected items where there were no specific misconceptions, errors, or misunderstandings tracked. All that could be determined was whether or not a student was able to demonstrate the understanding or ability required for the performance objective measured by the item. For these items, the percentage of students with a more general type of misunderstanding reflected all students who did not answer the item correctly. This included students who attempted the item but provided an incorrect response (including invalid responses or off-task comments), as well as those students who did not answer the item (omitted responses).⁶ General types of misunderstandings apply to six items in physics (one MC and five CR) and 26 items in mathematics (12 MC and 14 CR). The majority of these items were constructed response and many required students to explain their answer or show their work. In the TIMSS scoring guides, the general incorrect code 79 covers any type of incorrect response, including “crossed out, erased, stray marks, illegible, or off task” responses. When including blanks (omitted responses), we assumed that students who reached the item, but did not respond, did not have the understanding necessary to answer the question (i.e., similar in nature to responses that contain stray marks or off-task comments). This is consistent with the assumption underlying TIMSS scale scores, where omitted responses are treated as incorrect in scaling. The alternative would be to remove the blanks (omitted responses) from the sample, which would underestimate the percentage of students who did not demonstrate conceptual understanding.

The codes used for specific or general types of misconceptions, errors, and misunderstandings, and the corresponding value labels in the TIMSS data files are provided for all physics and mathematics items in Appendix (Tables A.1 and A.2).⁷ The percentage of students with misconceptions, errors, and misunderstandings was computed for all items in each country (overall and by gender). For trend items administered in multiple assessments, the percentage of students was reported for each assessment year.

Average Percent Correct and Average Percent with Misconceptions, Errors, and Misunderstandings

These averages reflect the percent correct (or percent with misconceptions, errors, or misunderstandings) in each country averaged across the countries that have data for the item. For most items, this reflects the average across all five countries.

⁶The “percent omitted” does not include the percent “not reached;” a response that is “not reached” is treated as a missing response and is not included in the denominator for the percent correct or percent with misconceptions.

⁷The separate analyses for the physics items that measured two different types of misconceptions were identified by two different versions (V1 and V2; see Table A.1).

However, there were some assessment years where data were not available for all countries, and the averages were based only on three or four countries.

3.2.4 Statistical Comparisons

Differences in the percent correct and the percentage of students with misconceptions, errors, and misunderstandings were computed (1) between each country and the average across the five countries, (2) between female and male students within each country, and (3) across assessment years for the trend items. The appropriate t -tests were used for all comparisons involving these item-level statistics, and indicators of statistical significance are provided in all data tables and figures that provide comparisons. A difference was considered “significant” when the probability (p) associated with the t -test was less than 0.05 (i.e., the probability is at least 95% that the reported difference is “real” and not due to chance). We used the following t -tests for each type of comparison.

- (1) For comparisons between the percentages in each country and the average across the five countries, there is overlap between the samples (i.e., each country is part of the average). In such cases, a part-whole t -test was used to account for this overlap:

$$t = \frac{(est_j - est_i)}{\sqrt{se_i^2 + (1 - 2p)se_j^2}}$$

Where est_i is the estimated average percentage for the five countries; est_j is the estimated percentage for one country; se_i and se_j are the respective corresponding standard errors; and p is the proportion of the five countries represented by each country (0.2).

- (2) For within-country gender differences, there are two types of t -tests that can be used depending on the student samples: independent (when there are independent random samples of female and male students drawn from the population) and non-independent (when this is not the case). For independent random samples, the independent t -test is appropriate:

$$t = \frac{(est_{female} - est_{male})}{\sqrt{(se_{female})^2 + (se_{male})^2}}$$

where est_{female} and est_{male} are the estimates for the percentage of females and males, respectively, and se_{female} and se_{male} are the corresponding standard errors of these percentages. The independent t -test can be calculated using the output

from the IDB Analyzer, where the JRR procedure is used to determine the separate percentages and standard errors for females and males.

However, in the TIMSS and TIMSS Advanced assessments, the samples of female and male students are not independent, since they are in the same schools and classrooms selected to take the assessments. Therefore, the correct t -test for non-independent samples requires the standard error of the difference between the percentage of females and the percentage of males:

$$t = \frac{(est_{female} - est_{male})}{se(est_{female} - est_{male})}$$

The standard error of the difference, $se(est_{female} - est_{male})$, takes into account the covariance (cov) between females and males for dependent samples:

$$se_{(est_{female} - est_{male})} = \sqrt{se_{(est_{female})}^2 + se_{(est_{male})}^2 - 2cov(est_{female}, est_{male})}$$

To obtain the appropriate standard errors, the JRR procedure must be conducted on the female–male percentage difference. The version of the IDB Analyzer that we used (Version 4.0), employs the JRR procedure to obtain standard errors for the percent of females and the percent of males. It does not, however, allow for jackknifing the gender differences for these item-level statistics (percent correct or percent misconceptions, errors, and misunderstandings). Therefore, the independent standard errors (and computed t -tests) obtained using the IDB Analyzer are approximations that do not take into account the covariance between females and males. These approximations are acceptably accurate if the covariances are small in comparison to the standard errors of the percentage of females and percentage of males. For the item-level statistics, this is expected to be the case due to the design of TIMSS, where only a small number of students take each item. Generally, about four students in each school or class will take each item (Martin et al. 2014).

To determine the magnitude of the covariances for gender differences, we conducted analyses for selected items using the EdSurvey R Package⁸ (NCES [National Center for Education Statistics] 2018) *Gap* function. The *Gap* function applies the JRR technique to the difference between the percentage of females and males. The output includes the standard error of the difference and the covariance. In the tested cases, we found that the covariance for the

⁸EdSurvey is an R statistical package developed by American Institutes for Research (AIR) and commissioned by the National Center for Education Statistics (NCES). EdSurvey is tailored to the processing and analysis of NCES large-scale education data with appropriate procedures. EdSurvey Version 2.0.3 is designed for the analysis of national and international education data from the NCES, including TIMSS and TIMSS Advanced. For more information, see: <https://nces.ed.gov/nationsreportcard/researchcenter/software.aspx>

item-level statistics was very small. We then ran analyses on the same items using the IDB Analyzer and compared the standard errors and t -tests obtained for gender differences using the two different methods. The standard errors using the correct non-independent method (*EdSurvey* R Package) and those using the independent method (IDB Analyzer) were approximately the same (to the nearest at the 0.0001%), and the significance of reported differences was not affected.⁹ Thus, for convenience, we used the output from the IDB Analyzer for the gender differences and applied the approximate independent t -tests for all items. Additional information on both software packages, as well as example outputs, are provided in Appendix.

- (3) The differences between years for trend items are based on independent samples. Thus, the standard independent t -test was used:

$$t = \frac{(est_{year1} - est_{year2})}{\sqrt{(se_{year1})^2 + (se_{year2})^2}}$$

where est_{year1} and est_{year2} are the estimates for the percentage of students in the two assessment years being compared, and se_{year1} and se_{year2} are the corresponding standard errors.

3.3 Addressing the Research Questions

As described in Sect. 3.2.2, we reviewed the set of TIMSS and TIMSS Advanced items that measured student understanding of the key concepts (gravity in physics and linear equations in mathematics), administered in each assessment year from 1995 to 2015. As a consequence, we established performance objectives that could be assessed by the items across grade levels and the types of student misconceptions, errors, and misunderstandings demonstrated on these items. This enabled us to report student performance on the TIMSS and TIMSS Advanced items related to gravity and linear equations across countries by grade level, gender, and assessment year to answer the three research questions.

⁹See Appendix (Tables A.3 and A.4) for comparison of the output from the IDB Analyzer and *EdSurvey* R package.

3.3.1 Research Question 1

What are common types of student misconceptions, errors, and misunderstandings in grade four, grade eight, and the final year of secondary school (TIMSS Advanced students), and how do they compare across countries?

To answer the first research question, we examined items from TIMSS and TIMSS Advanced administered at each grade level that demonstrated specific types of student misconceptions, misunderstandings, and errors. We determined the percentage of students for each response type by country and on average across the five countries included in the study. Response patterns provided evidence of the nature and extent of students' misconceptions, misunderstandings, and errors. We presented released example items at each grade level to illustrate the different types of misconceptions, errors, and misunderstandings. Each example item exhibit showed the item; the scoring guide (for CR items) or correct answer (for MC items); and other item information, including the TIMSS item ID,¹⁰ year(s) administered, and the performance objective assessed by the item.¹¹

3.3.2 Research Question 2

How do student misconceptions, errors, and misunderstandings differ by gender?

To answer the second research question, we determined the percentage of male students and percentage of female students demonstrating each type of misconception, misunderstanding, and error. We prepared tree graphs showing the gender differences across countries at each grade level.

3.3.3 Research Question 3

How persistent are patterns in misconceptions, errors, and misunderstandings over time?

To answer the third research question, we plotted figures for each country showing the percentage of students demonstrating the specific types of misconceptions, misunderstandings, and errors over multiple assessment years based on the set of trend items.

¹⁰The item IDs are those used in the TIMSS and TIMSS Advanced databases and released item sets, allowing readers to access all the released items used in this report.

¹¹The performance objectives are those developed for this report. These generally are more detailed than the broader TIMSS and TIMSS Advanced framework objectives and reflect the specific set of items included.

References

- IEA. (2018). *IDB Analyzer (Version 4.0)*. Hamburg, Germany: IEA. Retrieved from <http://www.iea.nl/data>.
- Martin, M. O., Mullis I. V. S., & Foy, P. (2014). TIMSS Advanced 2015 assessment design. In I. V. S. Mullis, & M. O. Martin (Eds.), *TIMSS Advanced 2015 assessment frameworks* (pp. 85–98). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from: <https://timssandpirls.bc.edu/timss2015-advanced/frameworks.html>.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016a). *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., Foy, P., & Hooper, M. (2016b). *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/>.
- NCES. (2018). *EdSurvey (Version 2.0.3)*. US Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <https://cran.r-project.org/package=EdSurvey>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

