5 What are the strengths and weaknesses of students in Türkiye?

This chapter analyses students' performance in Türkiye in different sub-domains and individual test items in Mathematics, Science and Reading in the International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS) and the OECD Programme for International Student Assessment (PISA) assessments in Grades 4, 8 and at 15 years of age. It identifies content and cognitive areas of strength and weakness in Türkiye's performance in comparison to the OECD average.

Analyses of student performance in Chapter 2 revealed substantial improvements over time in mathematics and science in International Association for the Evaluation of Educational Achievement (IEA) Trends in International Mathematics and Science Study (TIMSS) Grades 4[1] and 8, and in reading, mathematics and science in the OECD Programme for International Student Assessment (PISA). This improvement was driven by a significant decline in the share of low performers in all grades, subjects and in both assessments. Moreover, this improvement brought Türkiye's performance closer to the OECD average in all subjects of assessment. However, those findings only show the average level of performance and the trends over time. This chapter takes a deeper look, examining student performance in sub-domains of the main subjects of assessment and individual test items.

Box 5.1. What the data tell us

Science

- Physics is an area of strength throughout schooling (except in Grade 8 in 2019).
- Chemistry emerged as a strength in Grade 8 in 2019.
- Life science (Grade 4)/Earth science (Grade 8) are areas of weakness.

Mathematics

- The content sub-domains of number in Grade 4 and algebra in Grade 8 are areas of strength.
- Mathematical reasoning – which refers to solving pure mathematical or real-life problems using logical, systematic thinking – is a weakness in Grade 4 but becomes a strength by Grade 8.
- The sub-domain of applying became a strength in Grade 4 in 2019. In Grade 8, the sub-domain of reasoning is a strength and knowing became a strength in 2019.

Reading

- At age 15, narration and argumentative texts and the cognitive process of evaluating and reflecting are areas of strength, while description texts and the cognitive process of locating information are areas of weakness. Those results might reflect exposure to certain types of texts and activities at school and the coverage of the curriculum.

This chapter relies on an in-depth analysis of performance in different domains and sub-domains of mathematics, science and reading and on success rates on individual test items for those domains and sub-domains. The analysis uses data from TIMSS 2015 and TIMSS 2019 (Grades 4 and 8) (Box 5.2) and from the PISA 2018 international surveys. Note that "domain" and "subject" of assessment are used interchangeably. The domains (and sub-domains) of assessment across TIMSS 2015 Grades 4 and 8 are set out in Table 1.5 and those for TIMSS 2019 in Table 5.2.

Box 5.2. TIMSS 2019 additional analysis

Initially, when the item analysis of this chapter was developed in 2020, only the data from TIMSS 2015 were available. At a later stage, the analysis was expanded using the data from TIMSS 2019 at the request of the Ministry of National Education in Türkiye.

Conducting item analysis of Türkiye's TIMSS 2019 data is of particular interest because, in the 2019 TIMSS assessment, Türkiye chose to assess students in Grade 5 rather than Grade 4 as in previous rounds of TIMSS. TIMSS provides countries with the choice of assessing different grades, if

students do not fall under the minimum average age at the time of testing, to provide the most appropriate match between their curricula and the content of the TIMSS assessments. The change in the assessed year brings Türkiye closer to the age assessed in most countries since students in Türkiye start school slightly earlier than in other countries on average (IEA, 2020[1]). Item analysis of the 2019 data provides a perspective on this change by showing how a student performed in specific areas of the curriculum. A new curriculum was also implemented in Türkiye in 2017/18, so the 2019 analysis may provide a perspective on how the reform is shaping changes in learning.

It is important to note that in 2019 TIMSS started transitioning to computer-based assessment. In 2019, countries were offered the choice between the paper assessment or the new eTIMSS assessment TIMSS 2019 took steps to ensure as much comparability as possible between paper- and computer-based assessments (IEA, 2020[1]). However, it is not possible to establish exactly how the change in assessment mode affected the country's performance in 2019. The TIMSS 2019 Item Equivalence Study found that the mode effect may be stronger in some countries than others (Fishbein et al., 2018[2]).

Source: IEA (2017[3]), *TIMSS 2019 Assessment Frameworks*, https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/mathematics-framework/; IEA (2020[1]), *Methods and Procedures: TIMSS 2019 Technical Report*, https://timssandpirls.bc.edu/timss2019/methods/ (accessed on 6 December 2021); Fishbein, B. et al. (2018[2]), "The TIMSS 2019 Item Equivalence Study: Examining mode effects for computer-based assessment and implications for measuring trends", https://doi.org/10.1186/s40536-018-0064-z.

Table 5.1. Sub-domains assessed by TIMSS, 2015

	Maths		Science	
	Grade 4	Grade 8	Grade 4	Grade 8
Content domains	Number (50%)	Number (30%)	Life science (45%)	Biology (35%)
	Geometric shapes and measures (35%)	Algebra (30%)	Physical science (35%)	Chemistry (20%)
	Data display (15%)	Geometry (20%)	Earth science (20%)	Physics (25%)
		Data and chance (20%)		Earth science (20%)
Cognitive domains	Knowing (40%) Applying (40%) Reasoning (20%)	Knowing (35%) Applying (40%) Reasoning (25%)	Knowing (40%) Applying (40%) Reasoning (20%)	Knowing (35%) Applying (35%) Reasoning (30%)

Source: Mullis, I. et al. (2016[4]), *Student Achievement*, http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-achievement/ (accessed on 5 March 2018); Martin, M. et al. (2016[5]), *TIMSS 2015 International Results in Science*, TIMSS & PIRLS International Study Center, Boston.

Table 5.2. Sub-domains assessed by TIMSS, 2019

	Maths		Science	
	Grade 4	Grade 8	Grade 4	Grade 8
Content domains	Number (50%)	Number (30%)	Life science (45%)	Biology (35%)
	Measurement and geometry (30%)	Algebra (30%)	Physical science (35%)	Chemistry (20%)
	Data (20%)	Geometry (20%)	Earth science (20%)	Physics (25%)
		Data and probability (20%)		Earth science (20%)

	Maths		Science	
	Grade 4	Grade 8	Grade 4	Grade 8
Cognitive domains	Knowing (40%) Applying (40%) Reasoning (20%)	Knowing (40%) Applying (40%) Reasoning (20%)	Knowing (35%) Applying (40%) Reasoning (25%)	Knowing (35%) Applying (35%) Reasoning (30%)

Note: Some of the domain names changed between 2015 and 2019 but the content has remained largely the same:
- In Grade 4 mathematics: Geometric shapes and measures have changed in Measurement and Geometry and Data display has changed in Data.
- In Grade 8 mathematics: Data and chance have changed in Data and Probability.

Also, the shares of each content domain covered in the assessment slightly changed between 2015 and 2019, to better reflect the curricula, standards and frameworks of the participating countries. For further information about the topics covered within the sub-domains, please see IEA (2013[6]) and (2017[3]).
Source: Mullis, I. et al. (2016[4]), *Student Achievement*, http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-achievement/ (accessed on 5 March 2018); Martin, M. et al. (2016[5]), *TIMSS 2015 International Results in Science*, TIMSS & PIRLS International Study Center, Boston.

In PISA, the analysis in this chapter focuses primarily on reading, since this was the major domain in 2018 (see Chapter 1). The PISA 2018 reading framework is based on the processes of locating information, understanding and evaluating and reflecting across single or multiple-source texts (Table 5.3). For further information about each of these processes, please see *PISA 2018 Results (Volume 1): What Students Know and Can Do* (OECD, 2019[7]).

Table 5.3. Reading processes assessed by PISA across text source, 2018

Reading processes	Single source text (65%)	Multiple-source text (35%)
Locating information (25%)	Scanning and locating (15%)	Searching for and selecting relevant text (10%)
Understanding (45%)	Representing literal meaning (15%) Integrating and generating inferences (15%)	Integrating and generating inferences (15%)
Evaluating and reflecting (30%)	Assessing quality and credibility, and Reflecting on content and form (20%)	Corroborating and handling conflict (10%)

Note: Percentages in parenthesis show the share of items by process and text source.
Source: OECD (2019[7]), *PISA 2018 Results (Volume I): What Students Know and Can Do*, https://dx.doi.org/10.1787/5f07c754-en.

An explanation of the analyses of performance in sub-domains and on individual test items in this chapter

This chapter relies on two types of analyses: analysis of student performance on sub-domains of assessment and analyses of student performance on individual test items, which are then mapped to different sub-domains and item characteristics.

Sub-domain analysis

Both the TIMSS and PISA assessments assess student performance in a number of sub-domains of the three main subjects: reading, mathematics and science (PISA) and mathematics and science (TIMSS) (see Chapter 1). While student performance across sub-domains tends to be strongly correlated, variation from a country's average performance in one subject indicates a relative strength or weakness in the

country's performance (e.g. the analysis in TIMSS and PISA might identify a relative strength in geometry because, in comparison with the national average in mathematics, performance in this sub-domain is higher than expected).

Analysis by item characteristics

Analysing the characteristics of the individual test items reveals if performance is associated with certain item characteristics, such as an item's content and cognitive sub-domain, its format (multiple-choice or open-ended) and its difficulty. The success rate of each item for students in Türkiye is compared to the average success rate for all participating OECD countries. Conspicuous items, indicating strengths or weaknesses, are those on which students in Türkiye performed particularly well or less well in comparison with the OECD average (Box 5.3).

The analyses are carried out for mathematics and science in Grades 4 and 8 in TIMSS 2015 and 2019 and for mathematics, science and reading in PISA 2018. Strengths or weaknesses identified through those two methods might shed light on different emphasises in teaching or curricula.

Box 5.3. Selection of conspicuous test items

The analysis seeks to identify the items which stand out as indicating particular strengths or weaknesses of students in Türkiye. Those items could be selected in a number of ways. The most straightforward method is to compare the success rates of students in Türkiye to the OECD average. However, such an approach would be distorted when comparing items with relatively low or high success rates overall. For instance, a difference of 10% between success rates for Türkiye and the OECD average is less notable when the two success rates are 95% and 85% than when they are 55% and 45% – in other words, the relationship between getting items correct and the items' difficulty is not linear (OECD, 2013[8]).

In order to transform success rates into a linear metric, the logit transformation of those percentages should be computed. This transformation has the effect of "stretching out" very low and very high success rates in comparison with success rates close to 50%. A logit value of 0 means that the item has a success rate of 50%; positive logits mean higher success rates and negative logits mean lower success rates.

For each test item, the logit transformation of the success rate for students in Türkiye is compared to the average of participating OECD countries. Conspicuous items are those with at least one standard deviation (both sides) between the success rate for Türkiye and that for the OECD.

The average of participating OECD countries is used for both PISA and TIMSS analyses to make the composition of the comparator group of countries similar across both surveys.

Once conspicuous items have been identified, they are divided into two groups: items on which students did particularly well (items representing strength) and those on which they did not do well (items representing weakness). Then both groups of items are described in terms of the content and cognitive sub-domains they represent. If a content (e.g. geometry) or cognitive (e.g. reasoning) sub-domain is disproportionately represented among items representing strength, then it is possible to conclude that this sub-domain represents an area of strength and vice versa.

Source: OECD (2013[8]), *Lessons from PISA 2012 for the United States*, https://dx.doi.org/10.1787/9789264207585-en.

Performance in different aspects of mathematics and science competency in Grade 4

Sub-domain analysis: In Grade 4 2015, physical science is an area of strength while geometry and life science are weaknesses

Comparing student performance in the different sub-domains in TIMSS 2015 Grade 4 to the overall average for Türkiye:

- In mathematics, geometry and reasoning are identified as areas of relative weakness.
- No sub-domain of mathematics was identified as an area of strength (Table 5.4).
- In science, physical science is an area of strength and life science is a relative weakness (see Table 5.5).

Table 5.4. Performance in mathematics sub-domains in Grade 4 (TIMSS 2015)

Country	Mean performance in mathematics	Performance in mathematics content domains			Performance in mathematics cognitive domains		
		Data display	Geometry	Number	Knowing	Applying	Reasoning
Germany	522	535	531	515	524	515	535
Poland	535	538	534	534	517	541	546
Russia	564	573	557	567	556	566	570
Türkiye	483	476	475	489	491	482	466
OECD average	524	526	526	523	523	525	527
TIMSS international average	509	504	507	510	509	508	507

Note: Where the terms "strong" and "weak" are used, they denote that the average score for the domain in question is either significantly above or below Türkiye's overall mathematics or science average score.
Relative weakness in Türkiye.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

Table 5.5. Performance in science sub-domains in Grade 4 (TIMSS 2015)

Country	Mean performance in science	Performance in science content domains			Performance in science cognitive domains		
		Earth science	Life science	Physical science	Knowing	Applying	Reasoning
Germany	528	519	528	532	527	529	532
Poland	547	540	557	540	544	554	542
Russia	567	562	569	567	569	568	561
Türkiye	483	480	472	496	478	486	483
OECD average	524	521	527	521	523	525	525
TIMSS international average	505	499	506	503	503	504	502

Note: Where the terms "strong" and "weak" are used, they denote that the average score for the domain in question is either significantly above or below Türkiye's overall mathematics or science average score.
Relative weakness in Türkiye.
Relative strength in Türkiye.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

Sub-domain analysis: In Grade 4 2019, performance in science is consistent across sub-domains

In 2019, students in Türkiye performed 40 points in mathematics and 43 in science higher compared with the country's performance in 2015. While students performed higher overall on average, the analysis here presents areas of relative weakness or strength in comparison to Türkiye's overall performance. A country with high average scores may still have areas of relative strength or weakness because students in that country do not perform uniformly across the assessed content.

Comparing student performance in the different sub-domains in TIMSS 2019 Grade 4 to the overall average for Türkiye:

- In mathematics, data and reasoning are identified as areas of relative weakness (Table 5.6). Compared to TIMSS 2015 Grade 4 , the data sub-domain has become an area of relative weakness while performance in geometry is no longer a weakness and now exceeds the mean performance in mathematics. As in 2015, reasoning in mathematics remains an area of relative weakness.
- In science, in contrast with 2015, students performed consistently across all assessed areas with no domain emerging as a strength or a weakness (Table 5.7).

Table 5.6. Performance in mathematics sub-domains in Grade 4 (TIMSS 2019)

Country	Mean performance in mathematics	Performance in mathematics content domains			Performance in mathematics cognitive domains		
		Data	Measurement and geometry	Number	Knowing	Applying	Reasoning
Germany	521	515	531	517	523	514	531
Poland	520	524	529	513	509	521	527
Russia	567	560	571	567	555	571	573
Türkiye	523	510	527	525	514	531	509
OECD average	529	531	533	530	527	528	530
TIMSS international average	501	499	504	508	503	505	503

Note: Where the terms "strong" and "weak" are used, they denote that the average score for the domain in question is either significantly above or below Türkiye's overall mathematics or science average score.
Relative weakness in Türkiye.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

Table 5.7. Performance in science sub-domains in Grade 4 (TIMSS 2019)

Country	Mean performance in science	Performance in science content domains			Performance in science cognitive domains		
		Earth science	Life science	Physical science	Knowing	Applying	Reasoning
Germany	518	509	521	518	520	516	518
Poland	531	529	534	526	524	538	525
Russia	567	554	570	572	562	572	569
Türkiye	526	524	519	538	531	528	521
OECD average	526	525	527	522	526	524	527
TIMSS international average	491	501	506	505	509	507	509

Note: Where the terms "strong" and "weak" are used, they denote that the average score for the domain in question is either significantly above or below Türkiye's overall mathematics or science average score.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

Item analysis 2015: Performance in individual test items in Grade 4

In TIMSS 2015, Grade 4 students answered 178 test items in mathematics and 200 in science. The conspicuous items (Box 5.3) in mathematics included 29 test items in which students did relatively well in comparison with the OECD average and 26 in which they performed relatively lower. In science, students did relatively well in 29 test items and relatively lower in 28.

Below, the attributes of those items are discussed with the aim of identifying patterns that characterise the performance of students in Türkiye. It is worth noting that sub-domain analysis identifies whether students are doing better in a particular sub-domain relative to the national average in that domain. In contrast, item analysis compares success rates on individual test items to average success rates of a group of countries – in this case, OECD countries. Hence, the former is based on national comparison within a country of reference while the latter is based on an international comparison between the country of reference and other countries.

In mathematics, number and knowing are areas of strength

In mathematics, the 29 items (out of 178) in which students did particularly well comprise (Figure 5.1):

- Nineteen items out of 96 covering the content sub-domain of number (representing 20% of all items covering the number sub-domain).
- Nineteen items out of 74 (representing 26% of items) covering the knowing sub-domain.

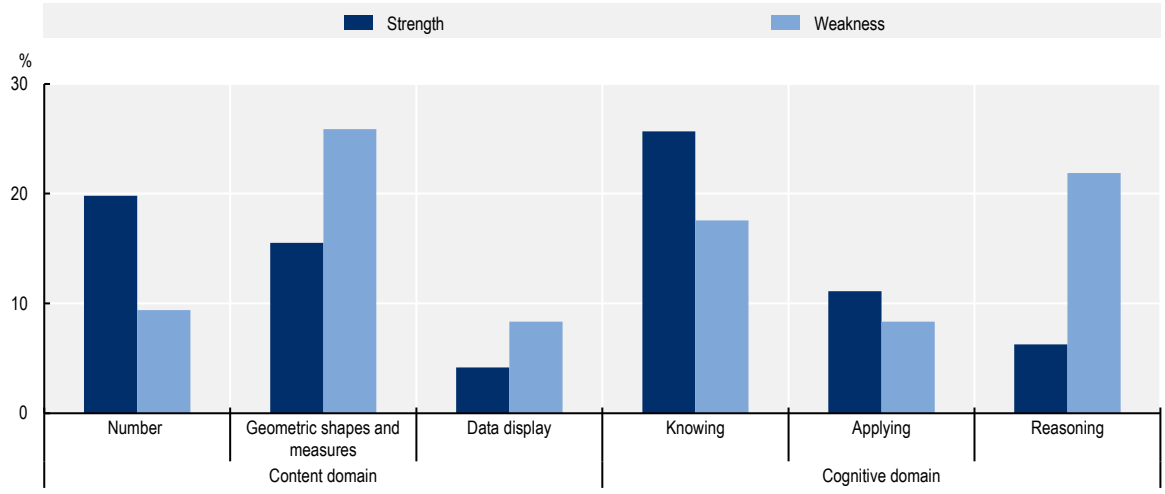
In mathematics areas of weakness include geometry and reasoning

In mathematics, the 26 items (out of 178) in which students did particularly less well comprise (Figure 5.1):

- Nine items out of 96 covering the content sub-domain of geometry (representing 16% of geometry items).
- Seven items out of 32 (representing 22% of items) covering the reasoning sub-domain.

Figure 5.1. Conspicuous items in mathematics in Grade 4 (TIMSS 2015)

Items in which students performed particularly well and less well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2016[9]), *TIMSS 2015*, https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

StatLink https://stat.link/lr10i3

The test items for the content and cognitive domains "number" and "knowing" in which students did well were relatively easy (i.e. had a level of difficulty lower than the average for their domain). In contrast, the test items in which students did well in geometry had a level of difficulty close to the average for all items in geometry, and those for the cognitive domain of applying were more difficult than the average. The two cognitive functions of applying and reasoning where students were less strong require a higher level of cognitive skills than the knowing sub-domain. Knowing entails familiarity with mathematical concepts while reasoning and applying require logical and systematic thinking and the application of mathematics in a variety of contexts (Vincent-Lancrin et al., 2019[11]).

Students in Türkiye also performed better in multiple-choice questions than open-ended questions. Open-ended questions have been used in international assessments, such as PISA, to assess higher-order skills such as applying knowledge and reasoning. For those questions, students often have to explain their answers and provide solid arguments to support those answers. At the time of the 2019 OECD report *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Türkiye*, teaching and testing in Türkiye made frequent use of multiple-choice tests so students may have had less exposure to open-ended questions than their peers in other countries (Kitchen et al., 2019[12]).

In general, the findings may reflect possible variations in curriculum coverage in terms of content domains but also in terms of coverage of knowledge and skills. For instance, students in Türkiye have a relative strength in knowledge about mathematics but weakness when it comes to skills such as mathematical reasoning.

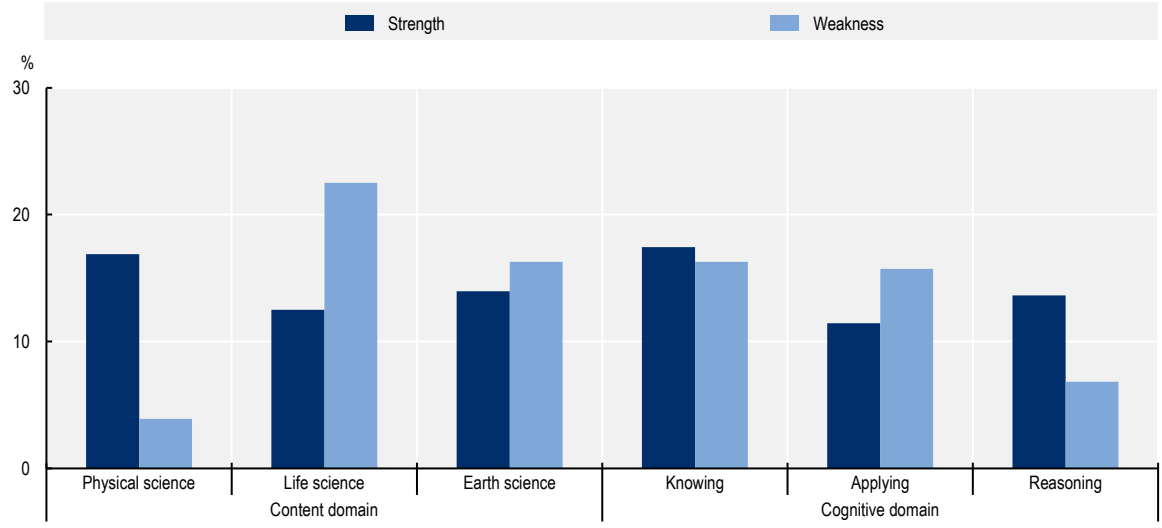
In science, physical science and reasoning are areas of strength

In science, students in Türkiye did well in (Figure 5.2):

- Physical science (13 out of 77 items, representing 17% of all items covering this domain).
- Reasoning (6 out of 44 items, representing 14% of all items in this category).

Figure 5.2. Conspicuous items in science in Grade 4 (TIMSS 2015)

Items in which students performed particularly well and less well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

StatLink https://stat.link/ry97u4

In terms of difficulty, students did well in items that are more difficult than the average of all science items. This is in line with the finding that students in Türkiye perform particularly well in science, which is a national strength, and suggests that teaching and learning support students in mastering even difficult items and content.

Life science and applying sub-domains are areas of weakness

- The items representing weakness are dominated by those covering life science (18 out of 80, representing 23%) (Figure 5.2).
- About 16% of items (11 out of 70) covering the applying sub-domain were found to represent an area of weakness.

As in mathematics, students did particularly well in multiple-choice science items and less well in open-ended questions.

Item analysis 2019: Performance in individual test items in Grade 4

In the computer-based assessment of TIMSS 2019, Grade 4 students answered 262 test items in mathematics and 240 in science. The conspicuous items in mathematics included 34 test items in which students in Türkiye did relatively well in comparison with the OECD average and 37 in which they performed relatively lower. In science, students did relatively well in comparison with the OECD average in 35 test items and relatively less well in 38.

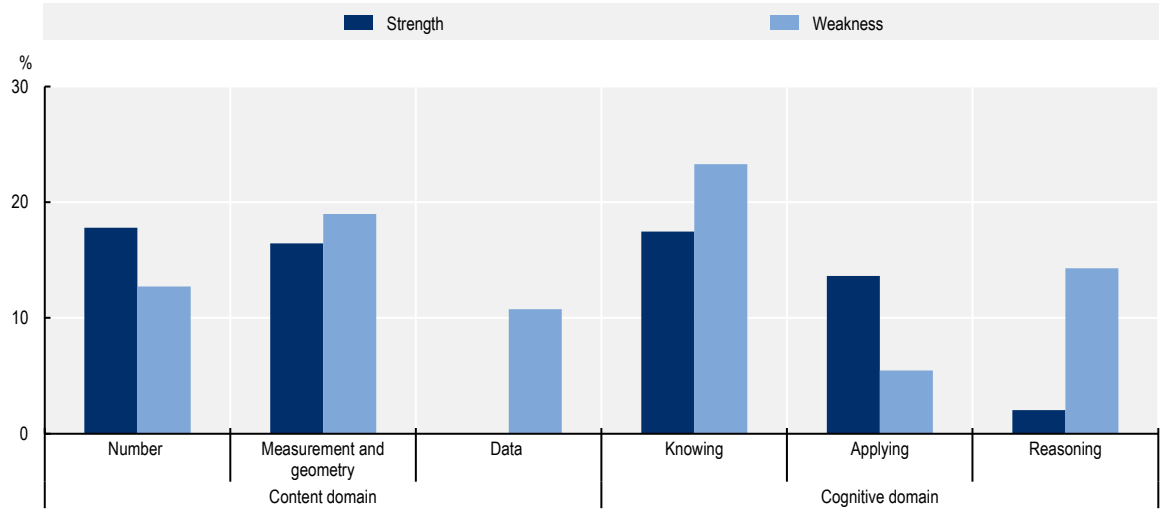
In mathematics, number and applying are areas of strength

In mathematics, the 34 items (out of 262) in which students did particularly well comprise (Figure 5.3):

- Twenty-one items out of 118 covering the content sub-domain of number (representing 18% of all items covering the number sub-domain).
- Fifteen items out of 110 (representing 14% of items) covering the applying sub-domain.

Figure 5.3. Conspicuous items in mathematics in Grade 4 (TIMSS 2019)

Items in which students performed particularly well and less well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

StatLink https://stat.link/3tn6bd

In mathematics, the 37 items (out of 262) in which students did particularly less well comprise (Figure 5.3):

- Seven items out of 65 covering the content sub-domain of data (representing 11% of data items).
- Twenty-four out of 103 (representing 23% of items) covering the knowing sub-domain.
- Seven items out of 49 (representing 14% of items) covering the reasoning sub-domain.

From 2015 to 2019 item analysis: Strengths and weaknesses in Grade 4 mathematics

In mathematics, the number sub-domain was a strength in both TIMSS 2015 and 2019, which might suggest that curriculum coverage, as well as teaching, is particularly strong in this sub-domain across both Grade 4 – at the end of primary education – and Grade 5 – at the start of lower secondary education. The applying sub-domain – which requires students to be able to apply mathematical skills across a variety of contexts – also became a strength in 2019. The change in student performance in the application of mathematical skills between 2015 and 2019 might reflect recent curricula reform, which places greater emphasis on applying knowledge and skills to solve problems (Box 5.4). In terms of difficulty, the items where students performed well in mathematics were more difficult than average, suggesting that students are well prepared to tackle some complex content.

In the cognitive sub-domains, both the 2015 and 2019 analyses identified reasoning as an area of weakness. Reasoning in mathematics requires high-level cognitive skills as individuals must go beyond solving routine problems to integrating knowledge and skills to solve multistep problems in unfamiliar, complex situations (IEA, 2017[3]). The items where students performed less well in data, knowing and

reasoning had a higher level of difficulty than the questions on average while students tended to manage easier questions in these sub-domains well.

Box 5.4. Curriculum reform in Türkiye

Comprehensive curriculum updates were carried out for all education levels in Türkiye between 2016 and 2018 in primary, lower and upper secondary education. The objective of recent curricula reforms was to better reflect children's interests, talents and character in the curricula while reducing the number of course hours, rearranging time for breaks and leisure activities, and streamlining the content. For secondary education, the objective was to reduce the number of compulsory course hours and provide greater diversity of courses while aligning the curriculum with the needs of higher education. Another major objective was to reorient the curriculum and refocus it on applied skills rather than knowledge. The emphasis on applied skills is reflected in the introduction of: workshops to translate knowledge into living skills; in-depth field courses, projects and applied studies; and learning in non-school environments such as natural, historical and cultural places, centres for science and arts, and museums. For example, the frequency of conducting a science experiment at the Grade 4 level, after the introduction of the new curriculum, was above the international average (World Bank, 2021[13]).

In the recent reforms, particular emphasis was given to the curriculum in the Sciences and Social Sciences High Schools with the aim of enabling students to build in-depth knowledge in the disciplines of natural and social sciences. In Vocational and Technical Anatolian High Schools, the emphasis of the reforms is on helping students to develop the skills needed by industry. This includes refocusing the curriculum on practical skills needed in the job market and introducing skills-based workshops to shift the focus from knowledge to 21st century skills.

The recent curricula reforms started by reviewing curricula in other countries and by working with experts on the different subjects including those not assessed by international surveys like PISA and TIMSS (e.g. civics, social sciences and humanities). Topics of international and current interest such as climate change and sustainability were introduced to the curriculum with the aim of making it more aligned with global developments. Examples include an elective course on environment for Grades 6, 7 and 8. These changes were implemented across the different phases of education with a simplified version of the curriculum for open high schools. After the development of the new curriculum, teachers were invited to participate in training programmes organised by the Directorate General for Teacher Training. Trained teachers were then expected to train other teachers in their provinces and to provide feedback about curriculum implementation to the Monitoring and Evaluation Department.

Sources: Government of Turkey (2019[14]), *Strategic Plan 2019-2023*; World Bank (2021[13]), "Türkiye's improvements in the quality of learning: TIMSS 2019 results", https://www.worldbank.org/en/country/Türkiye/brief/Türkiyes-improvements-in-the-quality-of-learning (accessed on 21 May 2021).

In science, physical science is an area of strength

In science, students in Türkiye did well in (Figure 5.4):

- Physical science (15 out of 82 items, representing 18% of all items covering this domain).

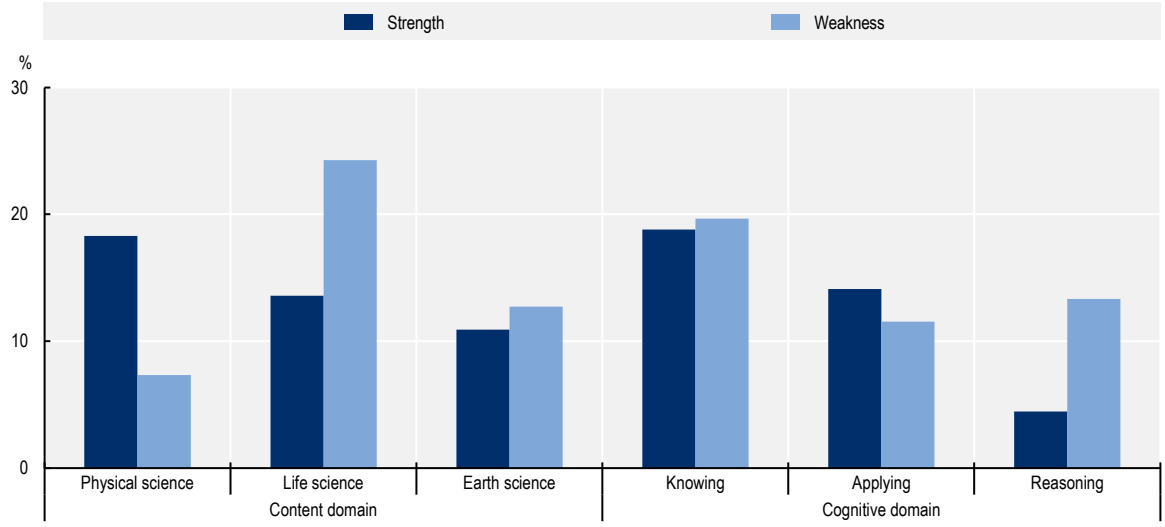
Life science and reasoning sub-domains are areas of weakness

- The items representing a weakness are dominated by those covering life science (25 out of 103, representing 24%) (Figure 5.4).

- About 13% of items covering reasoning (6 out of 45) were found to represent an area of weakness.

Figure 5.4. Conspicuous items in science in Grade 4 (TIMSS 2019)

Items in which students performed particularly well and less well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

StatLink https://stat.link/gfdlqr

From 2015 to 2019 item analysis: Strengths and weaknesses in Grade 4 science

Across 2015 and 2019 in science, students in Türkiye performed consistently well in physical science, while the life science sub-domain emerges as a weakness. As in mathematics, the items where students performed well were more difficult than average, again suggesting that students have the knowledge and skills to do well when faced with difficult questions. In contrast, the life science items where students did less well had a lower difficulty than average. As students do not do well even in comparatively easier life science items, their performance might reflect the balance of content in Türkiye's curriculum with less time devoted to life science. In contrast, the strength of student performance in physical science might suggest that it is a domain that receives greater emphasis. Students' high performance in physical science is also facilitated by their strength in number skills and knowledge since the two content domains are highly related (Rossdy et al., 2019[15]).

Also, in line with the results in mathematics, there was an improvement in the applying sub-domain in science which was a weakness in 2015. Applying in science requires high-level skills with individuals using their knowledge to compare, contrast and classify groups of objects or materials, relating knowledge of a science concept to a specific context, generating explanations and solving practical problems (IEA, 2017[3]).

Similarly, in line with the performance across different mathematics sub-domains, reasoning in science also emerges as a weakness in 2019. Reasoning in science refers to using evidence and science understanding to analyse, synthesise and generalise, often in unfamiliar situations and complex contexts (IEA, 2017[3]). Students also did not perform well even in easier items, suggesting that they might not be exposed to many activities or content that develops reasoning skills. The identification of reasoning as a

weakness in both science and mathematics in 2019 also suggests that the pattern of performance might reflect emphasis in the curriculum and pedagogical approaches.

Performance in different aspects of mathematics and science competency in lower secondary education

Sub-domain analysis: In Grade 8 2015, physics continues to be a strength while earth science is a weakness

In line with student performance in Grade 4, physics was a relative strength for students in Türkiye while earth science (which shares some similar content areas with the life science sub-domain in Grade 4, which was a weakness) was a relative weakness (Table 5.9). In mathematics, however, in contrast to student performance in Grade 4, reasoning in mathematics was a relative strength in Grade 8 (Table 5.8). All other differences were not found to be statistically significant.

Table 5.8. Mathematics sub-domains in Grade 8 2015

Country	Mean performance in Mathematics	Performance in mathematics content domains			Performance in mathematics cognitive domains			
		Algebra	Data and chance	Number	Geometry	Knowing	Applying	Reasoning
Russia	538	558	507	533	536	543	541	520
Türkiye	458	459	467	447	463	447	460	472
OECD average	507	496	515	511	503	503	508	510
TIMSS international average	481	483	472	482	477	481	480	479

Note: Where the terms “strong” and “weak” are used, they denote that the average score for the domain in question is either significantly above or below Türkiye’s overall mathematics or science average score.
Relative strength in Türkiye.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

Table 5.9. Science sub-domains in Grade 8 2015

Country	Mean performance in science	Performance in science content domains				Performance in science cognitive domains		
		Chemistry	Earth science	Biology	Physics	Knowing	Applying	Reasoning
Russia	544	558	532	539	548	558	538	538
Türkiye	493	493	477	491	506	489	492	495
OECD average	519	513	523	521	516	517	519	521
TIMSS international average	485	485	481	483	485	484	485	484

Note: Where the terms “strong” and “weak” are used, they denote that the average score for the domain in question is either significantly above or below Türkiye’s overall mathematics or science average score.
Relative weakness in Türkiye.
Relative strength in Türkiye.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

Sub-domain analysis: In Grade 8 2019, reasoning is a strength in mathematics

In line with student performance in Grade 8 in 2015, reasoning in mathematics continues to be an area of relative strength (Table 5.10). In both TIMSS 2015 and 2019, reasoning was a weakness in Grade 4 but became a strength in Grade 8, perhaps suggesting that the curriculum in Türkiye focuses on developing this cognitive domain in lower secondary education.

Table 5.10. Mathematics sub-domains in Grade 8 2019

Country	Mean performance in mathematics	Performance in mathematics content domains				Performance in mathematics cognitive domains		
		Algebra	Data and chance	Number	Geometry	Knowing	Applying	Reasoning
Russia	543	560	517	541	540	550	543	536
Türkiye	496	493	502	493	490	494	491	504
OECD average	513	506	518	513	513	510	514	514
TIMSS international average	489	497	489	495	497	497	496	499

Note: Where the terms “strong” and “weak” are used, they denote that the average score for the domain in question is either significantly above or below Türkiye’s overall mathematics or science average score.
Relative strength in Türkiye.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

In line with the pattern that emerged in science in Grade 4 in 2019, performance across all content domains and cognitive domains in Grade 8 was consistent, with no domain being statistically different from the mean (Table 5.11).

Table 5.11. Science sub-domains in Grade 8 2019

Country	Mean performance in science	Performance in science content domains				Performance in science cognitive domains		
		Chemistry	Earth science	Biology	Physics	Knowing	Applying	Reasoning
Russia	543	551	533	543	540	543	543	543
Türkiye	515	516	509	513	518	506	515	524
OECD average	519	510	525	520	516	516	518	522
TIMSS international average	490	490	487	493	491	492	492	491

Note: Where the terms “strong” and “weak” are used, they denote that the average score for the domain in question is either significantly above or below Türkiye’s overall mathematics or science average score.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

Item analysis 2015: Performance in individual test items in Grade 8

In TIMSS 2015, Grade 8 students answered 215 test items in mathematics and 260 in science. The conspicuous items in mathematics included 30 test items in which students did particularly well in comparison with the OECD average and 32 in which they did poorly. In science, students did particularly well in 31 test items and poorly in 32.

In mathematics, areas of strength include geometry, data and chance, algebra and reasoning

Students performed relatively equally, doing well (and less well) across a number of sub-domains. However, items in which students did particularly well included (Figure 5.5):

- Nine out of 43 items covering geometry (representing 21% of all items in this sub-domain).
- Seven out of 41 items covering data and chance (representing 17% of all items in this sub-domain).
- Nine out of 63 items covering algebra (representing 14% of all items in this sub-domain).
- Eight out of 46 items covering reasoning (representing 17% of all items in this sub-domain).

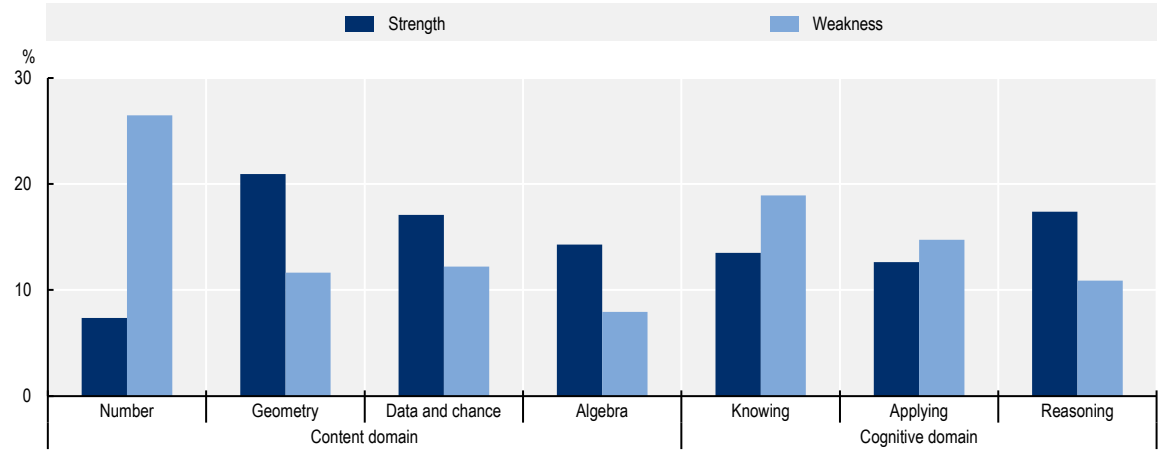
In mathematics, weaknesses include the sub-domains of number and knowing

- Eighteen out of 68 items covering the number sub-domain (representing 26% of all items in this sub-domain) (Figure 5.5).
- Fourteen out of 74 items covering the knowing sub-domain (representing 19% of all items in this sub-domain).

As with mathematics and science in Grade 4, students did better in multiple-choice questions and less well in open-ended questions. Moreover, average test item difficulty did not vary much between items representing strengths and those representing weaknesses. The two exceptions were knowing and multiple-choice questions, where the items representing a strength were more difficult than those representing a weakness. Variations in difficulty between items representing strengths and weaknesses could reflect exposure to certain topics in the curriculum, certain teaching practices emphasising particular cognitive skills, or certain assessment tools used more frequently in classrooms in Türkiye, such as multiple-choice questions.

Figure 5.5. Conspicuous items in mathematics in Grade 8 (TIMSS 2015)

Items in which students performed well and not well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2016[9]), *TIMSS 2015*, https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

In comparison with Grade 4, there was a reversal in the areas of strength and weakness in Grade 8. Geometry and reasoning, which were areas of weakness in Grade 4, became areas of strength in Grade 8, and the reverse is true for number and knowing. Data from PISA and TIMSS show that students in Türkiye

tend to perform less well in mathematics compared with the other domains of reading and science. While the reasons for this cannot be identified through the data alone, this might reflect a less consistent approach across the curriculum and teaching in mathematics in the country. The performance in mathematics contrasts notably with science, where there are clear consistencies in the patterns of strength and weakness across grades and domains in Türkiye.

In science, physics is an area of strength

- Twelve conspicuous physics items out of 62 (representing 19% of all physics items) fell into the group of items representing strengths (Figure 5.6)

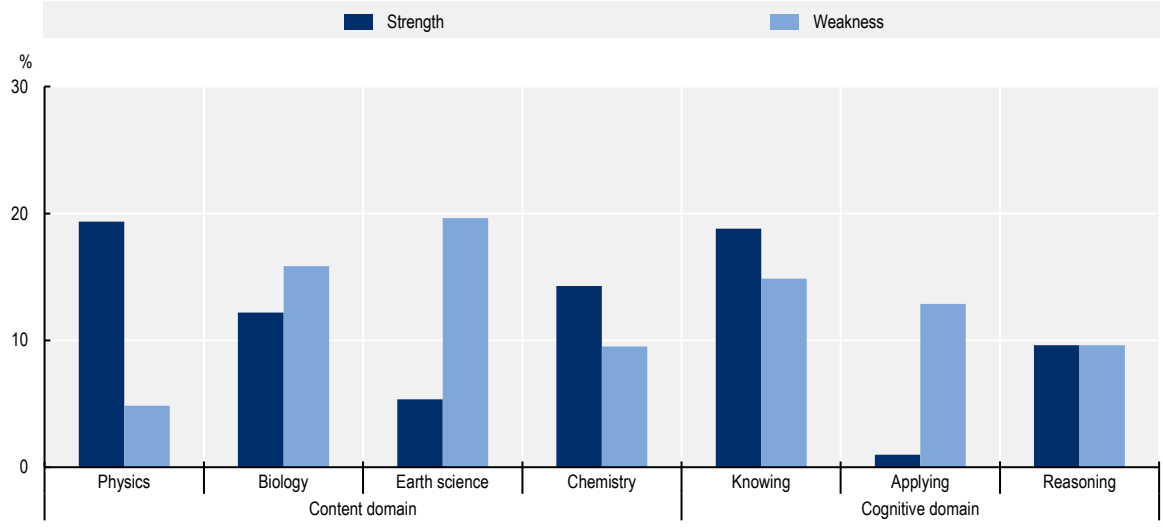
In science, earth science and applying are areas of weakness

- Eleven out of 56 items covering earth science (representing 20% of all items in this sub-domain) fell into the group of items representing weaknesses (Figure 5.6).
- Thirteen out of 101 items covering the cognitive domain of applying (representing 13% of those items) also fell into this group.

The test items covering physics in which students did particularly well were more difficult than those representing weaknesses. This seems counterintuitive and could be for a number of reasons. One might be because students make less effort in items that they perceive to be easier, another might be differences in emphasis in the physics curriculum or approaches to teaching and learning in physics in Türkiye. For the rest of the content and cognitive domains the average difficulty did not vary hugely between items representing strengths and those representing weaknesses. In comparison with Grade 4, physics remains an area of strength while applying scientific knowledge remains an area of weakness.

Figure 5.6. Conspicuous items in science in Grade 8 (TIMSS 2015)

Items in which students performed well and not well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2016[9]), *TIMSS 2015,* https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021).

Item analysis 2019: Performance in individual test items in Grade 8

In mathematics, areas of strength include algebra, knowing and reasoning

Items in which students did particularly well included (Figure 5.7):

- Ten out of 74 items covering algebra (representing 14% of all items in this sub-domain).
- Twenty out of 93 items covering knowing (representing 22% of all items in this sub-domain).
- Six out of 53 items covering reasoning (representing 11% of all items in this sub-domain).

In mathematics, weaknesses include geometry, data and probability and applying

Items in which students did less well included (Figure 5.7):

- Thirteen out of 60 items covering geometry (representing 22% of all items in this sub-domain).
- Eight out of 59 items covering data and probability (representing 14% of all items in this sub-domain).
- Twenty-two out of 134 items covering applying (representing 16% of all items in this sub-domain).

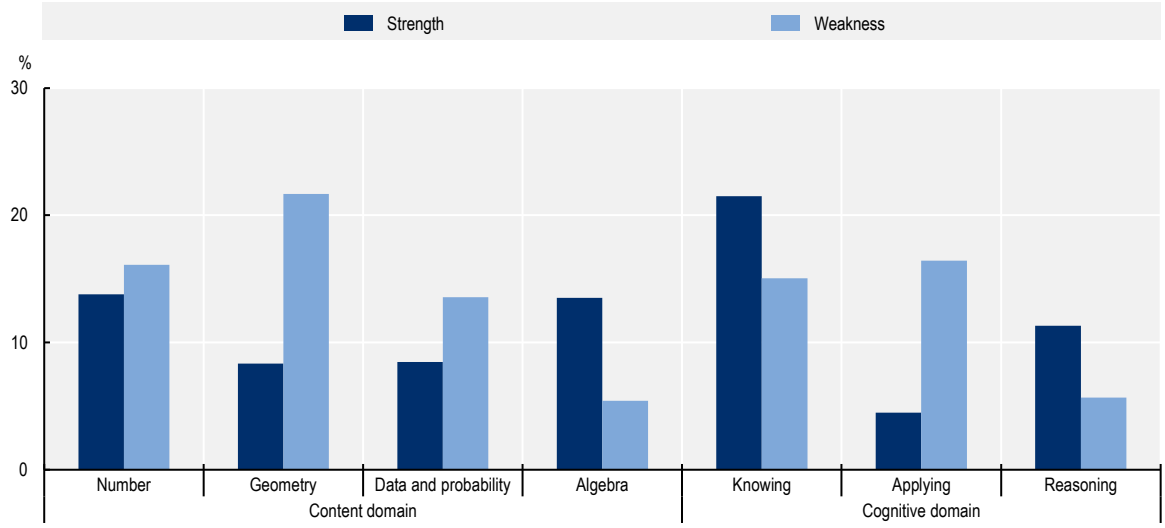
From 2015 to 2019 item analysis: Strengths and weaknesses in Grade 8 across years in mathematics

From both the analysis of 2015 and 2019, algebra emerges as a strong area of content in Türkiye in Grade 8. The consistency of algebra as a strength might reflect the focus of the curriculum in Türkiye. However, aside from this, there is no clear pattern in the strengths and weaknesses in mathematics over 2015 and 2019 or in comparison with Grade 4. This might reflect the impact of curriculum changes or more broadly a less consistent approach to mathematics teaching across schooling.

In Grade 8 mathematics, students performed better in multiple-choice questions. This is consistent with the results from 2015 and might be related to the frequent use of multiple-choice questions in classroom tests at school in Türkiye (Kitchen et al., 2019[12]). Regarding difficulty, in mathematics, students did well in items that were more difficult on average compared to other items, suggesting that they master these areas, such as number in Grade 4 and algebra in Grade 8. At the same time, the items in which students performed less well were more difficult on average, indicating that students are still able to respond well to questions covering the basis of geometry, data and probability and applying.

Figure 5.7. Conspicuous items in mathematics in Grade 8 (TIMSS 2019)

Items in which students performed well and not well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

StatLink https://stat.link/q7u0b3

In science, chemistry is an area of strength

Items in which students did particularly well included (Figure 5.8):

- Sixteen out of 63 items covering chemistry (representing 25% of all items in this sub-domain).

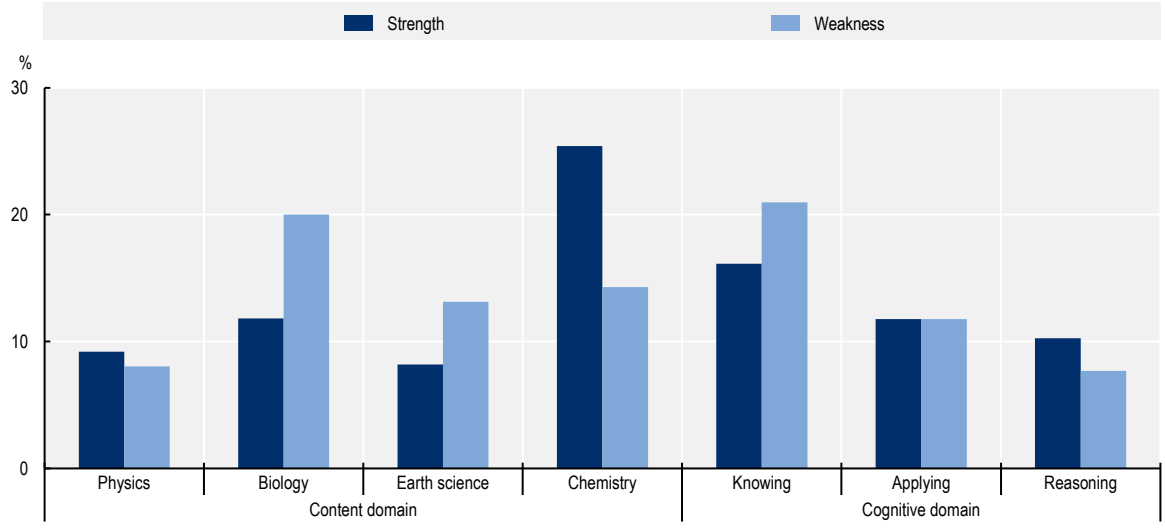
In science, biology, earth science and knowing are areas of weakness

Items in which students did less well included (Figure 5.8):

- Twenty-two out of 110 items covering biology (representing 20% of all items in this sub-domain).
- Eight out of 61 items covering earth science (representing 13% of all items in this sub-domain).
- Twenty-six out of 124 items covering knowing (representing 21% of all items in this sub-domain).

Figure 5.8. Conspicuous items in science in Grade 8 (TIMSS 2019)

Items in which students performed well and not well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: IEA (2020[10]), *TIMSS 2019*, https://timssandpirls.bc.edu/timss2019/international-database/, (accessed on 20 December 2021).

StatLink https://stat.link/ofu5jd

From 2015 to 2019 item analysis: Strengths and weaknesses in Grade 8 across years in science

In Grade 8 in 2019, students performed particularly well in chemistry. Students in Türkiye are strong in both algebra and chemistry, which reflects the shared skills that underpin these subjects – chemistry uses algebra to express relationships between quantities and substances (Cunningham and Whelan, 2014[16]). In line with Grade 4 in 2019, students improved in the domain of applying and this is might be related to the new curriculum implemented in 2018 (Box 5.4). In science, earth science was an area of weakness in both 2015 and 2019, which might point to this sub-domain having less coverage in Türkiye's curriculum.

In science, students performed well in items that were relatively easy compared to the other items in the same domains showing that even in the weaker ones – biology, earth science and the cognitive domains of knowing items – students are able to respond to easier questions covering the basics. However, students struggled with more difficult items, which might suggest that the curriculum or pedagogy does not cover or prepare students for topics that are more difficult. Only in chemistry, students performed well in difficult items. This reflects the high performance of students in Türkiye in chemistry compared to other OECD countries.

Performance in different aspects of reading competency in upper secondary education

At the age of 15, almost all students in Türkiye were in upper secondary education. Students who sat the PISA 2018 test responded to a series of test items in mathematics, reading and science selected from a pool of 243 reading items, 70 mathematics items and 115 science items. Since reading was the major domain of assessment in PISA 2018, more test items were used to assess this domain than for

mathematics and science, so the analysis in this chapter focuses on reading. Moreover, sub-scales were constructed only for reading and not for mathematics and science.

Sub-domain analysis: At age 15, understanding and evaluating and reflecting are areas of strength

In each of the domains – reading, mathematics and science – PISA items differ in terms of content and cognitive process (Table 5.3). Items are also differentiated by proficiency (with levels ascending with difficulty) and question format. Chapter 1 provides a full description of the assessment framework. The reading test items were characterised by a number of attributes such as: cognitive process (understand, locate information, evaluate and reflect), text source (single or multiple texts), item format (i.e. simple multiple-choice, complex multiple-choice, open-ended), text type (e.g. argumentative, narrative, exposition) and text format (i.e. continuous, non-continuous and mixed). The design of the reading test is described in detail in the PISA 2018 assessment and analytical framework (OECD, 2019[17]).

Comparing student performance in the different reading sub-domains to the overall average for Türkiye, understanding, evaluating and reflecting, in addition to proficiency with single texts, were found to be areas in which students did better than the national average for reading as a whole (Table 5.12). It is possible that those results are driven by curriculum coverage or by student exposure to teaching practices that emphasise certain cognitive skills.

Table 5.12. Reading sub-domains in PISA

Country	Mean performance in reading	Mean performance in reading process sub-scales			Mean performance in the single- and multiple-source sub-scales	
		Locate information	Understand	Evaluate and reflect	Single text	Multiple text
Germany	498	498	494	497	494	497
Poland	512	514	514	514	512	514
Russia	479	479	480	479	477	482
Türkiye	466	463	474	475	473	471
OECD average	487	487	486	489	485	490

Note:
Relative strength in Türkiye.
Source: OECD (2021[18]), "PISA: Programme for International Student Assessment", https://dx.doi.org/10.1787/data-00365-en (accessed on 21 May 2021).

Item analysis: Performance in individual test items in upper secondary education

In reading, areas of strength included evaluating and reflecting

- In terms of cognitive processes, only one process stood out: evaluate and reflect (Figure 5.9). Out of 63 test items, 15 items (24%) were identified as representing a strength against only four items representing a weakness (6%). In evaluating and reflecting, students should reflect on the content and form of the text and critically assess the quality and validity of the information therein.

The PISA items can also be differentiated in terms of the types of sources that students were required to engage with. Particular strengths among students in Türkiye included:

- Out of the 45 test items covering narration texts, 11 items (24% of all items covering this text type) were identified as conspicuous items in which Turkish students did exceptionally well. The information in narration texts refers to the properties of objects in time. Narration texts typically require students to answer questions relating to when, in what sequence and why characters in stories behave as they do.
- Six out of 38 items covering argumentative text (representing 16% of those items) were found to represent an area of strength. Argumentation texts present the relationship among concepts or propositions and often answer "why" questions.
- Turkish students did particularly well in test items covering continuous texts, with 26 out of 158 items identified as representing an area of strength (representing 16% of all items covering this type). Continuous texts are formed by sentences organised into paragraphs. Examples of continuous texts include newspaper reports, essays, novels, short stories, reviews and letters. In contrast, no test items covering non-continuous texts fell into the category representing strength. Examples of non-continuous texts include lists, tables, graphs, diagrams, advertisements, schedules, catalogues, indices and forms.

No major differences in average difficulty were observed between items in the category representing a strength and those representing a weakness with the exception of the cognitive process of evaluating and reflecting. For this sub-domain, the items representing an area of strength were easier than those representing a weakness. This could reflect variations in students' effort when answering test questions, as they may put more effort into answering difficult questions rather than easy ones. It could also reflect exposure to certain teaching practices or particular topics in the curriculum that are associated with the cognitive process of evaluating and reflecting.

In reading, area of weakness across the cognitive domains was locating information

- Twelve out of 49 items (24%) covering the cognitive process of locating information fell into the category of conspicuous items representing a weakness (Figure 5.9). Locating information consists of accessing and retrieving information within a piece of text and searching for and selecting relevant text.

Other weaknesses reflected the types of sources students were required to engage with including description, multiple-source and non-continuous texts.

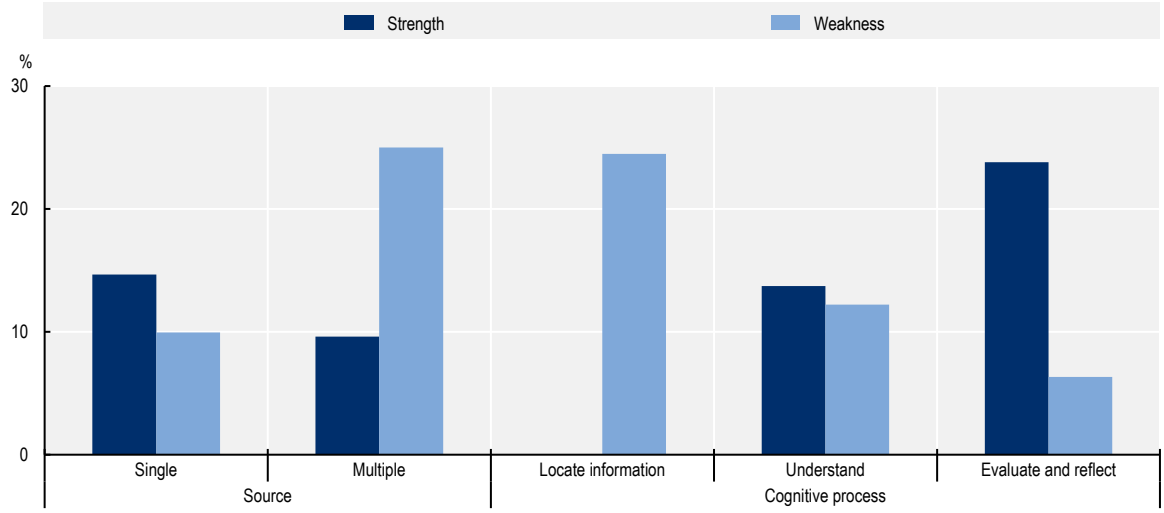
- Six out of 30 items (20%) covering description texts were identified as conspicuously weak items. Description texts are texts where the information refers to the properties of objects in space. Such texts typically provide an answer to "what" questions.
- Students also did worse in test items covering multiple-source texts with 38% of all items covering this type (8 out of 21 items) being identified as representing an area of weakness.
- Students did particularly worse in test items covering non-continuous texts such as lists, tables, graphs, diagrams, advertisements, schedules, catalogues and forms. Six out of 30 (20%) test items fell into this category.

These results could reflect certain exposure at school to particular types of text or instructional strategies. Students are probably more exposed to continuous texts such as newspapers, reports, essays and novels than non-continuous texts such as tables, graphs, diagrams, indices and forms.

Finally, students found it more difficult to engage with complex, multiple-choice questions, with eight out of 32 test items (25%) being identified as representing an area of weakness. No major differences in average difficulty were observed between items in the category representing a strength and those representing a weakness.

Figure 5.9. Conspicuous items in reading in PISA 2018

Items in which students performed well and not well



Note: The items represented in the figure are conspicuous, precisely top (strength) and bottom (weakness) items. Average (neither strength nor weakness) as well as non-conspicuous items were excluded.
Source: OECD (2021[18]), "PISA: Programme for International Student Assessment", https://dx.doi.org/10.1787/data-00365-en.

StatLink https://stat.link/2vjith

In science, students did well in test items covering evaluating and designing scientific enquiry and in open-ended test questions while in mathematics no clear patterns emerged

Since mathematics and science were not the major domains of assessment in PISA 2018, fewer items were included in the assessment and fewer were identified as conspicuous. Therefore:

- In science, students did well in the content domain of physical systems (6 items representing 16% of all items in this group) and did worse in items covering living systems (8 items representing 17% of all items in this group).
- Students also did particularly well in test items covering the cognitive process of "evaluate and design scientific enquiry" with 5 out of 30 items (17%) identified as representing an area of strength.
- Eleven out of 49 (24%) test items covering the cognitive process of explaining phenomena scientifically were identified as conspicuous items representing a weakness.

Students did well in open-ended science questions and less well in complex multiple-choice ones. In physics and in the cognitive domain of explaining phenomena scientifically, the items in which students did particularly well were slightly more difficult than those in which they did poorly. In mathematics, only 14 conspicuous items (8 representing strength and 6 representing weakness) were identified out of 70 test items without a clear pattern emerging in terms of item characteristics.

Conclusions

By analysing the results from TIMSS Grade 4 and 8 in 2015 and 2019, as well as PISA 2018, physics emerges as an area of strength across all levels of schooling (except in Grade 8 in 2019). This may be related to the curriculum's focus as well as strong pedagogical practices that support students to do particularly well in this domain. Students also performed very well in chemistry in Grade 8 in 2019. Earth science and life science are areas of weakness for students in Türkiye in all grades. These two domains are closely related and the students' performance may reflect relatively less exposure to this kind of content. From TIMSS 2015 to TIMSS 2019, applying scientific knowledge improved, possibly reflecting the curriculum reform that took place in 2017/18.

In mathematics, in contrast with science, there is less consistency across areas of strength and weakness over the different levels of school and different assessments. However, sub-domains number in Grade 4 and algebra in Grade 8 were identified as areas of strength. Students' strength in these domains may facilitate their competency in the complex science domains of physics and chemistry since these mathematical skills underpin the skills required in these two science domains (Rossdy et al., 2019[15]). Overall, strengths and weaknesses that emerge from this chapter's analysis of the international assessments might be associated with curricula coverage or teaching and learning practices in Türkiye and might warrant further analysis nationally.

References

Cunningham, A. and R. Whelan (2014), "Maths for chemists", University of Birmingham, University of Leeds, https://www.birmingham.ac.uk/Documents/college-eps/college/stem/Student-Summer-Education-Internships/Maths-for-Chemists-Booklet.pdf. [16]

Fishbein, B. et al. (2018), "The TIMSS 2019 Item Equivalence Study: Examining mode effects for computer-based assessment and implications for measuring trends", *Large-scale Assessments in Education*, Vol. 6/11, https://doi.org/10.1186/s40536-018-0064-z. [2]

Government of Turkey (2019), *Strategic Plan 2019-2023 (MEB 2019-2023 Stratejik Plan)ı*, Government of Turkey, https://www.meb.gov.tr/stratejik_plan/ (accessed on 27 June 2022). [14]

IEA (2020), *Methods and Procedures: TIMSS 2019 Technical Report*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timssandpirls.bc.edu/timss2019/methods/ (accessed on 6 December 2021). [1]

IEA (2020), *TIMSS 2019*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timssandpirls.bc.edu/timss2019/international-database/ (accessed on 20 December 2021). [10]

IEA (2020), *TIMSS 2019 International Results in Mathematics and Science*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timss2019.org/reports/home-contexts/# (accessed on 6 December 2021). [19]

IEA (2017), *TIMSS 2019 Assessment Frameworks*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/mathematics-framework/. [3]

IEA (2016), *TIMSS 2015*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timss2015.org/timss-2015/about-timss-2015/ (accessed on 21 May 2021). [9]

IEA (2016), *TIMSS 2015 International Reports*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, http://timssandpirls.bc.edu/timss2015/international-results/ (accessed on 21 May 2021). [20]

IEA (2013), *TIMSS 2015 Assessment Frameworks*, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement, https://timssandpirls.bc.edu/timss2015/frameworks.html (accessed on 28 July 2021). [6]

Kitchen, H. et al. (2019), *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, https://doi.org/10.1787/5edc0abe-en. [12]

Martin, M. et al. (2016), *TIMSS 2015 International Results in Science*, TIMSS & PIRLS International Study Center, Boston. [5]

Mullis, I. et al. (2016), *Student Achievement*, International Study Center, Lynch School of Education, Boston College, http://timssandpirls.bc.edu/timss2015/international-results/timss-2015/mathematics/student-achievement/ (accessed on 5 March 2018). [4]

OECD (2021), "PISA: Programme for International Student Assessment", *OECD Education Statistics (database)*, https://doi.org/10.1787/data-00365-en (accessed on 21 May 2021). [18]

OECD (2019), *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/b25efab8-en. [17]

OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/5f07c754-en. [7]

OECD (2013), *Lessons from PISA 2012 for the United States*, Strong Performers and Successful Reformers in Education, OECD Publishing, Paris, https://doi.org/10.1787/9789264207585-en. [8]

Rossdy, M. et al. (2019), "The role of physics and mathematics in influencing science students' performance", in *Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017 - Volume 1*, https://link.springer.com/chapter/10.1007/978-981-10-8730-1_40. [15]

Vincent-Lancrin, S. et al. (2019), *Fostering Students' Creativity and Critical Thinking: What it Means in School*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/62212c37-en. [11]

World Bank (2021), "Turkey's improvements in the quality of learning: TIMSS 2019 results", World Bank, Washington, DC, https://www.worldbank.org/en/country/turkey/brief/turkeys-improvements-in-the-quality-of-learning. [13]

Note

[1] In this report, the terminology of "TIMSS, Grade 4" is used throughout since this is the official name of the assessment. However, the data refer to Grade 5 students in lower secondary education in Türkiye.