Chapter 15
Proficiency Scale Construction

INTRODUCTION

This chapter discusses the methodology used to develop the PISA reporting scales. These describe levels of proficiency in the different PISA domains, and presents the outcomes of the development process for reading literacy, the major domain in the PISA 2018 assessment.

The reporting scales are called "proficiency scales" rather than "performance scales" because they describe what students *typically* know and can do at given levels of proficiency, rather than how individuals who were tested *actually* performed on a single test administration. This emphasis reflects the primary goal of PISA, which is to report general population-level results rather than the results for individual students. PISA uses samples of students and items to make estimates about populations. A sample of 15-year-old students is selected to represent all 15-year-olds in a country/economy and a sample of test items from a large pool is administered to each student. Results are then analysed using statistical models that estimate the likely proficiency of the population, based on this sampling.

The PISA test design makes it necessary to use techniques of modern item response modelling to both, estimate the ability of all students taking the PISA assessment and the statistical characteristics of all PISA items. These are described in Chapter 9.

The PISA data are collected using a rotated test design in which students take different but overlapping sets of items. The mathematical model employed to analyse the PISA data is implemented through test analysis software that uses iterative procedures to simultaneously estimate the distribution of students along the proficiency dimension assessed by the test, as well as a mathematical function that describes the association of student proficiency and the likelihood of a correct response for each item on the test. The result of these procedures is a set of item parameters that represents, among other things, locations on a proficiency continuum reflecting the domain being assessed. On that continuum, it is possible to estimate the distribution of groups of students, and thereby the average (location) and range (variability) of their skills and knowledge in this domain. This continuum represents the overall PISA scale in the relevant test domain, such as reading, mathematics, or science.

PISA assesses students and uses the outcomes of that assessment to produce estimates of students' proficiency in relation to the skills and knowledge being assessed in each domain. The skills and knowledge of interest, as well as the kinds of tasks that represent those abilities, are described in the PISA frameworks (OECD, 2019). For each domain, one or more scales are defined, each ranging from very low levels of proficiency to very high levels. Students whose ability estimate places them at a certain point on a PISA proficiency scale would be more likely to be able to successfully complete tasks at or below that point. Those students would be increasingly *more likely* to complete tasks located at progressively lower points on the scale, and increasingly *less likely* to complete tasks located at progressively higher points on the scale. Figure 15.1 depicts a simplified hypothetical proficiency scale, ranging from relatively low levels of proficiency at the bottom of the figure, to relatively high levels towards the top. Six items of varying difficulty are placed along the scale, as are three students of varying ability. The relationship between the students and items at various levels is described in the figure.

In addition to defining the numerical range of the proficiency scale, it is also possible to define the scale by describing the competencies typical of students at particular points along the scale. The distribution of students along this proficiency scale is estimated, and locations of students can be derived from this distribution and their responses on the test. Those location estimates are then aggregated in various ways to generate and report useful information about the proficiency levels of 15-year-old students within and among participating countries.

The development of a method for describing proficiency in PISA reading, mathematical and scientific literacy occurred in the lead-up to the reporting of outcomes of the PISA 2000 survey and was revised in the lead-up to each of the subsequent surveys. The same basic methodology has again been used to develop proficiency descriptions for the PISA 2018 assessment, however, like in the PISA 2015 assessment, a more general statistical model was used in the scaling procedure compared to assessments before 2015.

The proficiency descriptions that had been developed for the science domain in the PISA 2015 assessment, for the mathematics domain in the PISA 2012 assessment, and for financial literacy in 2012 were used again to report the results of the PISA 2018 assessments.

Reporting for reading, the major domain in the PISA 2018 assessment, was linked back to the 2009 proficiency scale and was based on the detailed proficiency level descriptions developed in 2009, the last PISA assessment in which reading was the major domain. These proficiency level descriptors were reviewed and revised based on the PISA 2018 assessment data in order to incorporate the new reading framework and the performance of the new items, including the multiple source items.

Figure 15.1: Simplified relationship between items and students on a proficiency scale
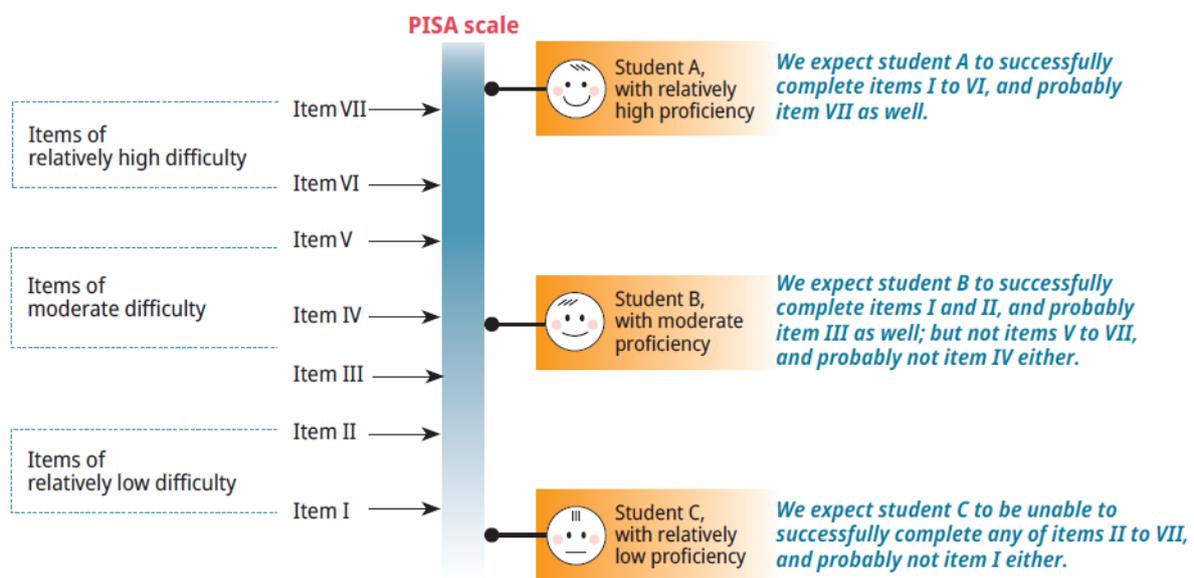
The reading expert group worked with the PISA international contractor to review and revise the sets of described proficiency scales and subscales for PISA reading. Similarly, the international contractor worked with the global competence expert group to develop the described proficiency scale for that domain.

DEVELOPMENT OF THE DESCRIBED SCALES

The development of described proficiency scales for PISA has been carried out through a process that typically involves several tasks conducted by the expert groups and the item development team. The process of developing the described scales involved several iterations as the data were collected and analysed during the PISA 2018 assessment. It should be noted that, as each PISA assessment builds upon the work implemented in previous assessments, the same tasks are not completed for every domain in every assessment. The following description of the development process focuses on the development of described proficiency scales for reading and global competence.

Classification of items

As part of new item development for reading and global competence, test developers classified all items based on the specifications provided in the framework for each domain. Item classifications for the trend reading items were also revised to reflect the PISA 2018 assessment framework. All classifications were reviewed by each of the expert groups and revised as needed.

Defining the overall proficiency scale

As part of its work in developing the assessment framework for reading, the expert group drafted initial descriptors of the factors that drive difficulty for each cognitive process defined for 2018 based on their knowledge of the research literature on reading processes and comprehension. These descriptors, presented as an initial hypothesis, were shared as part of the framework to allow item developers to design items representing the increase in skills and ability reflected across the levels.

Final item parameters were estimated for the trend and new reading items based on analysis of the Main Survey data. Using this information on item performance, the reading expert group met over several days and reviewed representative items, particularly those that were classified with the new cognitive processes and discussed key characteristics that differentiated performance along the proficiency scale. As part of that review process, the initial draft descriptors for each level in the overall proficiency scale were refined and finalised.

Defining the proficiency scale for global competence was more challenging because the domain was newly developed for the PISA 2018 assessment. The experts defined the factors that drive difficulty across the levels of proficiency and the type of thinking that is required as students demonstrate greater levels of proficiency. They also set cut-off points along the scale that defined each level of performance.

Identifying possible subscales

For each domain assessed in PISA, reporting includes an overall proficiency scale based on the combined results for all items within that domain. In addition, the assessment framework may support subscales based on the various dimensions of the framework. Where subscales are included, they must arise clearly from the domain framework, be meaningful and potentially useful for feedback and reporting purposes, and be defensible with respect to their measurement properties. Thus, the first stage in the process involves having the experts articulate possible reporting subscales based on the most recent framework.

As the major domain in the PISA 2018 assessment, work on identifying possible subscales for reading, in addition to the overall reading literacy scale, began with a review of the subscales used in the PISA 2009 assessment, when reading was last a major domain. In the PISA 2009 assessment, volume I of the *PISA 2009 Results* included an overall reading scale and descriptions of subscales that described the types of reading tasks or "cognitive aspects": access and retrieve, integrate and interpret and reflect and evaluate and subscales based on the form of reading material: continuous texts and non-continuous texts (OECD, 2010). For digital reading, a separate, single scale was developed based on the digital reading assessment items administered in 19 countries/economies in the PISA 2009 assessment, as an international option (OECD, 2011). In the PISA 2012 assessment, when reading reverted to minor domain status, a single print reading scale was reported, along with a single digital reading scale. For the PISA 2018 assessment, the reading expert group decided the former distinction of "cognitive aspects" should be updated to "cognitive processes". This terminology better connects the PISA 2018 assessment framework with the literature on reading psychology and better reflects the actual skills and proficiencies assessed. The subscales that correspond to the ways students interact and process text were updated to the following: locate information, understand, and evaluate and reflect. The former subscales that were based on the form of reading material are not included in the PISA 2018 assessment. Instead scales are included corresponding to using a single unit of text or multiple units of texts for answering the questions.

Scales in the minor domains

For science, the subscales selected for inclusion in the PISA 2006 database were the three competency-based subscales based on the scientific dimensions documented in the framework: *explaining phenomena scientifically, identifying scientific issues* and *using scientific evidence*. The 2015 expert group recommended reporting again on the three scientific competencies, as they were defined in the updated framework: *explain phenomena scientifically, evaluate and design scientific enquiry,* and *interpret data and evidence scientifically*. In addition, the expert group recommended that two knowledge subscales be reported: *content knowledg*e and *procedural/epistemic knowledge*. Procedural and epistemic knowledge were combined into a single reporting subscale due to a limited number of epistemic items in some of the administered forms. Finally, for continuity with previous reporting scales, three systems – *physical, living* and *Earth and space* – were recommended as a third reporting scale. For the PISA 2018 assessment, only a single scale representing overall proficiency in the science domain is reported.

In the case of mathematics, a single mathematical scale was developed for PISA 2000. With the additional data available in the PISA 2003 assessment, when mathematics was the major test domain, subscales based on the four overarching ideas – *space and shape, change and relationships, quantity* and *uncertainty* – were reported. In the PISA 2006 and PISA 2009 assessments, when mathematics was again a minor domain, only a single scale was reported. For the PISA 2012 assessment, the expert group carried out a comprehensive revision of the framework at the specific behest of the PISA Governing Board that indicated an interest in seeing mathematical process dimensions used as the primary basis for reporting in mathematics. As well as considering ways in which this could be done, the mathematics expert group also had to consider how the addition of the optional computer-based assessment component included in the PISA 2012 assessment could be incorporated into the reporting for 2012. The outcome of these considerations was, firstly, a decision that the computer-based items would be used to expand the same mathematical literacy dimension that was expressed

through the paper-based items. Secondly, the expert group recommended that three process-based subscales should be reported. These included: *formulating situations mathematically* (or "formulate"), *employing mathematical concepts, facts, procedures and reasoning* (or "employ"), and *interpreting, applying and evaluating mathematical outcomes* (or "interpret"). In addition, for continuity with the PISA 2003 reporting scales, the content-based scales including *space and shape, change and relationships, quantity*, and *uncertainty and data* (formerly "uncertainty"), were also reported. For the PISA 2018 assessment, only a single scale representing overall proficiency in the mathematics domain is reported.

For global competence, the innovative domain in the PISA 2018 assessment, a proficiency description on a single overall reporting scale was developed. The optional assessment of financial literacy used the same proficiency description developed for the PISA 2015 assessment.

Developing an item map

Based on item performance in the main survey, the test items in the study can be ordered from easiest to most difficult and this range of difficulty can be described using an item map. The item map contains a brief description of a selected number of released items along with their scale values. These descriptions explain the specific skills each item is designed to assess and are linked to the descriptions of performance at each level for the overall scale. As a result, the item map provides some insight into the range of skills and knowledge required of students and the proficiencies they need to demonstrate at various points along the scale.

DEFINING THE PROFICIENCY LEVELS

The proficiency levels for each of the PISA domains were defined in previous assessment cycles when each was first a major domain. The goal of that process was to decide how to divide up the proficiency continuum into levels that might have some utility. And, having defined those levels, decisions needed to be made about how to decide on the level to which a particular student should be assigned.

The relationship between the observed responses and student proficiency and item characteristics is probabilistic. That is, there is some probability that a particular student can correctly solve any particular item and each item can be differentially responsive to the proficiency being measured.

One of the basic tenets of the measurement of human skills or proficiencies is this: if a student's proficiency level exceeds the item's demands, the probability that the student can successfully complete that item is relatively high, and if the student's proficiency is lower than that required by the item, the probability of success for that student on that item is relatively low. The rate of change of the probability of success across the range of proficiency for each item is also affected by the sensitivity of the item to the proficiency scale.

This leads to the question as to the precise criterion that should be used to locate a student on the same scale as that on which the items are located. How can we assign a location that represents student proficiency in meaningful ways? When placing a student at a particular point on the scale, what probability of success should we deem sufficient in relation to items located at the same point on the scale? If a student were given a test comprising a large number of items, each with the same item characteristics, what proportion of those items would we expect

the student to successfully complete? Or, thinking of it in another way, if a large number of students of equal ability were given a single test item with a specified item characteristic, about how many of those students would we expect to successfully complete the item?

The answers to these questions depend on assumptions about how items differ in their characteristics or how items function, as well as on what level of probability is deemed a *sufficient probability of success*. In order to define and report PISA outcomes in a consistent manner, an approach is needed to define performance levels and to associate students with those levels. The same basic methodology has again been used to develop proficiency descriptions for the PISA 2018 assessment, however, like in the PISA 2015 assessment, a more general statistical model was used in the scaling procedure compared to assessments before 2015.

Defining proficiency levels for a PISA assessment progressed in two broad phases. The first, which came after the development of the described scales, was based on a substantive analysis of PISA assessment items in relation to the aspects that underpin each assessment domain. This produces descriptions of increasing proficiency that reflect observations of student performance and a detailed analysis of the cognitive demands of PISA assessment items. The second phase involves decisions about where to set cut-off points for levels and how to associate students with each level in order to lay out how a *sufficient probability of success* plays out in these levels. This is both a technical and a very practical matter of interpreting what it means to be at a level, and has significant consequences for reporting national and international results.

Several principles were considered in developing and establishing a useful meaning of being at a level, and therefore for determining an approach to locating cut-off points between levels and associating students with them. For the levels to provide useful information to the PISA assessment stakeholders, it is important to develop a common understanding of what performance at each of those levels means.

First, it is important to understand that the skills measured in each PISA assessment domain fall along a continuum: There are no natural breaking points to mark borderlines between stages along this continuum. Dividing the continuum into levels, though useful for communication about students' development, is essentially arbitrary. Like the definition of units on, for example, a scale of length, there is no fundamental difference between 1 metre and 1.5 metres – it is a matter of degree. It is useful, however, to define stages, or levels along the continua, because they enable us to communicate about the proficiency of students in terms other than continuous numbers. This is a rather common concept, an approach we all know from categorising clothing and portions by size (S, M, L, XL, etc.).

The approach adopted since the PISA 2000 assessment was that it would only be useful to regard students as having attained a particular level if this would mean that we can have certain expectations about what these students are capable of, in general, when they are said to be at that level. It was thus decided that this expectation would have to mean, at a minimum, that students at a particular level would be more likely than not to successfully complete tasks at that level. By implication, it must be expected that they would succeed on at least half of the items on a test composed of items uniformly spread across that level. This definition of being "at a level" is useful in helping to interpret the proficiency of students at different points across the proficiency range defined at each level.

For example, the expectation is that students located at the bottom border of a level would complete at least 50% of items correctly on a test set at the level, while students at the middle and top of each level would be expected to achieve a higher success rate. At the top border of a level would be the students who would be likely to solve a high proportion of the tasks at that level. But, being at the top border of that level, they would also be at the bottom border of the next highest level where, according to the reasoning here, they should have at least a 50% likelihood of solving any tasks defined to be at that higher level.

Furthermore, the meaning of being at a level for a given scale should be more or less consistent for each level and, indeed, also for scales from the different domains. In other words, to the extent possible within the substantively based definition and description of levels, cut-off points should create levels of more or less constant range. Some small variation may be appropriate, but for interpretation and definition of cut-off points and levels to be consistent, the levels have to be about equally broad within each scale. The exception would be the highest and lowest proficiency levels, which are unbounded.

Thus, a consistent approach should be taken to defining levels for the different scales. Their range may not be exactly the same for the proficiency scales in different assessment domains, but the same kind of interpretation should be possible for each scale that is developed. This approach links the two variables mentioned in the preceding paragraphs, and third related variable. The three variables can be expressed as follows:

- the expected success of a student at a particular level on a test containing items at that level (proposed to be set at a minimum that is near 50% for the student at the bottom of the level and greater for students who are higher in the level)
- the width of the levels in that scale (determined largely by substantive considerations of the cognitive demands of items at the level and data related to student performance on the items)
- the probability that a student in the middle of a level would correctly answer an item of average difficulty for that level (in fact, the probability that a student at any particular level would get an item at the same level correct), sometimes referred to as the "RP value" for the scale, where "RP" indicates "response probability".

Figure 15.2 summarises the relationship among these three mathematically linked variables under a particular scenario. The vertical line represents a segment of the proficiency scale, with marks delineating the "top of level" and "bottom of level" for any level one might want to consider, with a width of 0.8 logits between the boundaries of the level (noting that this width can vary somewhat for different assessment domains). The RP62 indicates that students will be located on the scale at a point that gives them a 62% chance of getting a typical item at that same level correct. The student represented near the top of the level shown has a 62% chance of getting an item correct that is located at the top of the level, and similarly the student represented at the bottom of the level has the same chance of correctly answering a question at the bottom of the level. A student at the bottom of the level will have an average score of about 52% correct on a set of items spread uniformly across the level. Of course, that student will have a higher likelihood (62%) of getting an item at the bottom of the level correct, and a lower likelihood (about 42%) of getting an item at the top of the level correct. A student at the top of the level will have an average score of about 70% correct on a set of items spread uniformly across the level. That student will have a higher likelihood (about 78%) of getting a typical item at the bottom of the level correct and a lower likelihood (62%) of getting an item at the top of the level correct.

Figure 15.2: Calculating the RP values used to define PISA proficiency levels

One proficiency level (RP62)
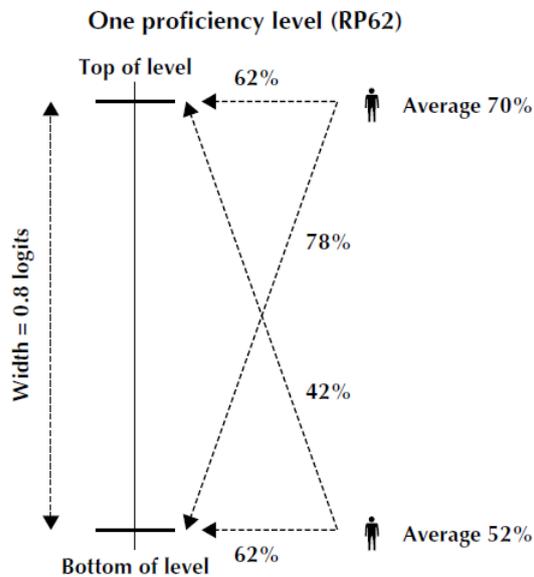
In PISA we have implemented the following solution: Start with the range of described abilities for each bounded level in each scale (the desired band breadth); then determine the highest possible RP value that will be common across domains potentially having bands of slightly differing breadth that would give effect to the broad interpretation of the meaning of being at a level (an expectation of correctly responding to a minimum of 50% of the items in a test comprising items spread uniformly across that level). The value RP = 0.62 is a probability value that satisfied the logistic equations for typical items in that level through which the scaling model is defined, subject to the two constraints mentioned earlier (a width per level of about 0.8 logits and the expectation that a student would get at least half of the items correct on a hypothetical test composed of items spread evenly across the level). In fact, RP=0.62 satisfied the requirements for any scales having band widths up to about 0.97 logits.

The highest and lowest levels are unbounded. For a certain high point on the scale and below a certain low point, the proficiency descriptions could, arguably, cease to be applicable. At the high end of the scale, this is not such a problem since extremely proficient students could reasonably be assumed to be capable of at least the achievements described for the highest level. At the other end of the scale, however, the same argument does not hold. A lower limit therefore needs to be determined for the lowest described level, below which no meaningful description of proficiency is possible. It was proposed that the floor of the lowest described level be set so that it was the same range as the other described levels. Student performance below this level is lower than that which PISA can reliably assess and, more importantly, describe.

REPORTING THE RESULTS FOR THE PISA READING SCALES

In this section, the ways in which levels of reading are defined, described and reported will be discussed. This will be illustrated using a subset of items from the PISA 2018 assessment.

Building an item map for reading

The data from the PISA reading assessment were analysed to estimate a set of item characteristics for the 245 items included in the main survey. During the process of item development, each item was classified to reflect the cognitive process it required. Following data analysis, the items were associated with their difficulty. Table 15.1 shows the item map, which includes this information for the released items from the PISA 2018 reading assessment. Each row in Table 15.1 represents an individual item. The selected items have been ordered according to their difficulty, with the most difficult at the top, and the least difficult at the bottom of the figure. The difficulty estimate for each item expressed in the reporting scale is given in the rightmost column.

Table 15.1: A map for released reading items

Level	Lower score limit	Question	Question difficulty
6	698	(there were no release items at this level)	-
5	626	RAPA NUI – Released item 6 (CR551Q10)	665
		COW'S MILK - Released Item 5 (CR557Q12)	662
		RAPA NUI – Released item 3 (CR551Q06)	654
		RAPA NUI – Released item 4 (CR551Q08)	634
4	553	RAPA NUI – Released item 5 (CR551Q09)	597
		RAPA NUI – Released item 7 (CR551Q11)	588
		RAPA NUI – Released item 1 (CR551Q01)	559
3	480	COW'S MILK - Released Item 3 (CR557Q07)	539
		RAPA NUI – Released item 2 (CR551Q05)	513
		COW'S MILK - Released Item 7 (CCR557Q14)	506
		COW'S MILK - Released Item 4 (CR557Q10)	498
2	407	CHICKEN FORUM - Released Item 7 (CR548Q09)	466
		CHICKEN FORUM - Released Item 3 (CR548Q01)	458
		COW'S MILK - Released Item 2 (CR557Q04)	452
		CHICKEN FORUM - Released Item 6 (CR548Q07)	409
1a	335	COW'S MILK - Released Item 6 (CR557Q13)	406
		CHICKEN FORUM - Released Item 2 (CR548Q03)	357
		CHICKEN FORUM - Released Item 5 (CR548Q05)	347
1b	262	CHICKEN FORUM - Released Item 1 (CR548Q02)	328
		CHICKEN FORUM - Released Item 4 (CR548Q04)	328
		COW'S MILK - Released Item 1 (CR557Q03)	323
		Most reading fluency tasks calling for a "no" response (meaningless sentences, such as "Airplanes are made of dogs")	
1c	189	Most reading fluency tasks calling for a "yes" response (meaningful sentences, such as "The red car had a flat tire")	

Defining levels of reading literacy

The reporting approach used by the OECD has been defined in previous PISA assessments and is based on the definition of a number of levels of proficiency. Descriptions were developed to characterise typical student performance at each level. The levels were used to summarise the

performance of students, to compare performances across subgroups of students, and to compare average performances among groups of students, in particular among the students from different participating countries. A similar approach has been used here to analyse and report PISA 2018 assessment outcomes for reading.

Since the PISA 2000 assessment, when reading as the major domain, results have been reported on a scale with a mean of 500 and a standard deviation of 100. In 2009, the last time reading was the major domain, the scale was extended to include items and descriptions of proficiency at levels below Level 1 and above Level 5. In the PISA 2009 assessment, Level 1 was renamed to Level 1a and Level 1b was added to describe lower levels of observed proficiency. Level 6 was added to describe the proficiencies above Level 5. The reading scale for the PISA 2018 assessment adds to the previous reporting scale by adding Level 1c. This was done to even further describe what readers at the lower levels of proficiency can do. The level definitions on the PISA scale are given in Table 15.2 below.

Table 15.2: Reading literacy performance band definitions on the PISA scale[1]

Level	Score points on the PISA scale
6	Above 697
5	Between 626 and 697
4	Between 553 and 625
3	Between 480 and 552
2	Between 407 and 479
1a	Between 335 and 406
1b	Between 262 and 334
1c	Between 189 and 261

Information about the items in each level is used to develop summary descriptions of the kinds of reading literacy associated with different levels of proficiency. These summary descriptions can then be used to encapsulate typical reading proficiency of students associated with each level. As a set, they describe development in reading literacy.

PISA is administered once every three years, with each of the three core domains the major focus in turn. Reading was the major domain in the PISA 2009 assessment. In the PISA 2018 assessment, therefore, had a set of level descriptors upon which to build. The new items that were developed for the PISA 2018 assessment were considered in relation to the existing level descriptions and in relation to the preliminary descriptions that were included in the PISA 2018 assessment framework for reading. The focus was first on the descriptions for the overall reading scale, presented here in Table 15.3.

---

[1] For exact cutpoints for each of the levels, for each proficiency scale, please refer to a later section in this chapter.

Table 15.3: Summary descriptions of the eight levels on reading reading proficiency in PISA 2018

Level	Characteristics of tasks
6	Readers at Level 6 can comprehend lengthy and abstract texts in which the information of interest is deeply embedded and only indirectly related to the task. They can compare, contrast and integrate information representing multiple and potentially conflicting perspectives, using multiple criteria and generating inferences across distant pieces of information to determine how the information may be used.

Readers at Level 6 can reflect deeply on the text's source in relation to its content, using criteria external to the text. They can compare and contrast information across texts, identifying and resolving inter-textual discrepancies and conflicts through inferences about the sources of information, their explicit or vested interests, and other cues as to the validity of the information.

Tasks at Level 6 typically require the reader to set up elaborate plans, combining multiple criteria and generating inferences to relate the task and the text(s). Materials at this level include one or several complex and abstract text(s), involving multiple and possibly discrepant perspectives. Target information may take the form of details that are deeply embedded within or across texts and potentially obscured by competing information. |
| 5 | Readers at Level 5 can comprehend lengthy texts, inferring which information in the text is relevant even though the information of interest may be easily overlooked. They can perform causal or other forms of reasoning based on a deep understanding of extended pieces of text. They can also answer indirect questions by inferring the relationship between the question and one or several pieces of information distributed within or across multiple texts and sources.

Reflective tasks require the production or critical evaluation of hypotheses, drawing on specific information. Readers can establish distinctions between content and purpose, and between fact and opinion as applied to complex or abstract statements. They can assess neutrality and bias based on explicit or implicit cues pertaining to both the content and/or source of the information. They can also draw conclusions regarding the reliability of the claims or conclusions offered in a piece of text.

For all aspects of reading, tasks at Level 5 typically involve dealing with concepts that are abstract or counterintuitive, and going through several steps until the goal is reached. In addition, tasks at this level may require the reader to handle several long texts, switching back and forth across texts in order to compare and contrast information. |
| 4 | At Level 4, readers can comprehend extended passages in single or multiple-text settings. They interpret the meaning of nuances of language in a section of text by taking into account the text as a whole. In other interpretative tasks, students demonstrate understanding and application of ad hoc categories. They can compare perspectives and draw inferences based on multiple sources.

Readers can search, locate and integrate several pieces of embedded information in the presence of plausible distractors. They are able to generate inferences based on the task statement in order to assess the relevance of target information. They can handle tasks that require them to memorise prior task context.

In addition, students at this level can evaluate the relationship between specific statements and a person's overall stance or conclusion about a topic. They can reflect on the strategies that authors use to convey their points, based on salient features of texts such as titles and illustrations. They can compare and contrast claims explicitly made in several texts and assess the reliability of a source based on salient criteria.

Texts at Level 4 are often long or complex, and their content or form may not be standard. Many of the tasks are situated in multiple-text settings. The texts and the tasks contain indirect or implicit cues. |

Level	Characteristics of tasks
3	Readers at Level 3 can represent the literal meaning of single or multiple texts in the absence of explicit content or organisational clues. Readers can integrate content and generate both basic and more advanced inferences. They can also integrate several parts of a piece of text in order to identify the main idea, understand a relationship or construe the meaning of a word or phrase when the required information is featured on a single page.

They can search for information based on indirect prompts, and locate target information that is not in a prominent position and/or is in the presence of distractors. In some cases, readers at this level recognise the relationship between several pieces of information based on multiple criteria.

Level 3 readers can reflect on a piece of text or a small set of texts, and compare and contrast several authors' viewpoints based on explicit information. Reflective tasks at this level may require the reader to perform comparisons, generate explanations or evaluate a feature of the text. Some reflective tasks require readers to demonstrate a detailed understanding of a piece of text dealing with a familiar topic, whereas others require a basic understanding of less-familiar content.

Tasks at Level 3 require the reader to take many features into account when comparing, contrasting or categorising information. The required information is often not prominent or there might be a fair amount of competing information. Texts typical of this level may include other obstacles, such as ideas that are contrary to expectation or negatively worded. |
| 2 | Readers at Level 2 can identify the main idea in a piece of text of moderate length. They can understand relationships or construe meaning within a limited part of the text when the information is not prominent by producing basic inferences, and/or when the text(s) include some distracting information.

They can select and access a page in a set based on explicit though sometimes complex prompts, and locate one or more pieces of information based on multiple, partly implicit criteria.

Readers at Level 2 can, when explicitly cued, reflect on the overall purpose, or on the purpose of specific details, in texts of moderate length. They can reflect on simple visual or typographical features. They can compare claims and evaluate the reasons supporting them based on short, explicit statements.

Tasks at Level 2 may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge by drawing on personal experience and attitudes. |
| 1a | Readers at Level 1a can understand the literal meaning of sentences or short passages. Readers at this level can also recognise the main theme or the author's purpose in a piece of text about a familiar topic, and make a simple connection between several adjacent pieces of information, or between the given information and their own prior knowledge.

They can select a relevant page from a small set based on simple prompts, and locate one or more independent pieces of information within short texts.

Level 1a readers can reflect on the overall purpose, and on the relative importance of information (e.g., main idea vs.non-essential detail) in simple texts containing explicit cues.

Most tasks at this level contain explicit cues as regards what needs to be done, how to do it, and where in the text(s) readers should focus their attention. |

Level	Characteristics of tasks
1b	Readers at Level 1b can evaluate the literal meaning of simple sentences. They can also interpret the literal meaning of texts by making simple connections between adjacent pieces of information in the question and/or the text.
	Readers at this level can scan for and locate a single piece of prominently placed, explicitly stated information in a single sentence, a short text or a simple list. They can access a relevant page from a small set based on simple prompts when explicit cues are present.
	Tasks at Level 1b explicitly direct readers to consider relevant factors in the task and in the text. Texts at this level are short and typically provide support to the reader, such as through repetition of information, pictures or familiar symbols. There is minimal competing information.
1c	Readers at Level 1c can understand and affirm the meaning of short, syntactically simple sentences on a literal level, and read for a clear and simple purpose within a limited amount of time.
	Tasks at this level involve simple vocabulary and syntactic structures.

CUTPOINTS DEFINING PROFICIENCY LEVELS FOR EACH PROFICIENCY SCALE

The following tables present the cut points used to assign items and students to a proficiency level. Values in the table are the lower bound for the corresponding level. For example, in the reading scale, Level 6 begins with 698.32. Level 5 begins with 625.61 and ends just below 698.32, where Level 6 begins. Below Level 1c are those with values lower than 189.33. In other words, those reaching a level are those with a score or difficulty at or above the given cut point.

Table 15.5: Cutpoints for the Reading Scale

Cut point	Level Name
698.32	Level 6
625.61	Level 5
552.89	Level 4
480.18	Level 3
407.47	Level 2
334.75	Level 1a
262.04	Level 1b
189.33	Level 1c

Table 15.6: Cutpoints for the MathematicsScale

Cut point	Level Name
669.30	Level 6
606.99	Level 5
544.68	Level 4
482.38	Level 3
420.07	Level 2
357.77	Level 1[2]

Table 15.7: Cutpoints for the Science Literacy Scale

---

[2] In PISA-D, Level 1 was further divided into Levels 1a, 1b, and 1c, using 295.47 and 233.17 as cutpoints for levels 1b and 1c, respectively. Level 1a was set using the original cutpoint for level 1 at 357.77.

Cut point	Level Name
707.93	Level 6
633.33	Level 5
558.73	Level 4
484.14	Level 3
409.54	Level 2
334.94	Level 1a
260.54	Level 1b[3]

REFERENCES

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

OECD (2011), PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI), OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264112995-en.

OECD (2010), PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I), OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264091450-en.

---

[3] Level 1b bandwidth is slightly narrower than others.