PISA 2022 Technical Report

3 Test Development for the Core Domains

Introduction

This chapter describes the processes used by the PISA Core A contractor, Educational Testing Service (ETS), and the international test development team to develop the tests for the core domains in the PISA 2022 cycle.

The tests for the PISA 2022 cycle included the following:

- a mathematics test, the major domain in PISA 2022
- a reading and a science test, the two minor domains
- a creative thinking test, the innovative domain for this cycle, and
- a financial literacy test, the international option for this cycle.

Test design and development for the Creative Thinking domain is presented and discussed in the Chapter 4 [*Development of the PISA 2022 Innovative Domain Assessment*] of this technical report.

In the PISA 2015 cycle, PISA moved from a primarily paper-based delivery survey that included optional computer-based modules, to a fully computer-delivered survey. A paper-based version of the assessment that included only trend units was developed for the small number of participants that chose not to implement the computer-delivered survey. The PISA 2018 cycle retained this same paper-based option, using the same paper-based materials as the PISA 2015 cycle. The PISA 2022 cycle retained this paper-based option as well; however, only one participant used the same paper-based materials as in the 2015 and 2018 cycles. The other paper-based participants administered a “new” instrument that was first used in the PISA for Development (PISA-D) assessment. This “new” paper-based instrument, which contained a substantial amount of material that overlapped with computer- and paper-based trend material administered by other participants, was comprised of clusters of units assessing mathematics, science, reading, and reading components.

The computer-based delivery mode allows PISA to measure new and expanded aspects of the domain constructs. In mathematics, new material for PISA 2022 included items developed to assess mathematical reasoning as a separate process classification, and items that leveraged the use of the digital environment (e.g., spreadsheets, simulators, data generators, drag-and-drop, etc.). A mixed-design that included a computer-based multistage adaptive testing was also adopted for the mathematics literacy domain to further improve measurement accuracy and efficiency, especially at the extremes of the proficiency scale. In financial literacy, some new units were developed based on an updated framework and to ensure adequate coverage of the domain following the release/removal of several units following the 2018 administration.

As noted in the list above, the core domains in PISA rotate between being a major or a minor domain. Table 3.1 shows the number of items in the main survey for the core domains for each PISA cycle since PISA 2000. Under this approach for measuring trends, each domain goes through a domain rotation that begins with a new or revised framework and continues with the two subsequent cycles in which it becomes a minor domain. The rotation concludes, and starts again, with becoming a major domain three cycles later. The third cycle- after alternating with the other two main domains - then involves another revision of the framework to reflect the current thinking about assessment for the new data collection as a major domain. For example, the revised framework for mathematics as the major domain in PISA 2022 and the introduction of computer-based items broadened the construct beyond what was measured in PISA 2012, the last time that mathematics was a major domain. Under the current design, the mathematics framework and instruments are expected to remain constant for the next two PISA cycles, with the next revision of the mathematics assessment and items expected for the PISA cycle to take place after PISA 2029, when mathematics will again be the major domain. Note that over time, the number of items included for minor domains has increased, which has helped stabilize and improve the measurement of trends for the minor domains by making the construct coverage for each minor domain comparable to that of a major domain. However, there has been a reduction in the number of student responses per item for the minor domains.

In addition to the three core domains (science, mathematics, and reading) and the innovative domain (creative thinking), the PISA 2022 assessment also included an optional assessment of financial literacy, which was administered only as a computer-based assessment.

Table 3.2 and Table 3.3 present the domain coverage for the computer- and paper-based assessments, respectively. All new items for mathematics were developed as computer-based items. The mathematics field trial design included seven clusters of trend items and twelve clusters of new items to study unit order effects. This was carried out in preparation for the introduction of the multistage adaptive testing design in the main survey. Then, in the main survey, the mathematics items were assigned according to the multistage adaptive design described in Chapter 2 [*The PISA 2022 Integrated Assessment Design*] of this report.

As shown in Table 3.2, there was no new item development for science or reading in PISA 2022. Both financial literacy and creative thinking were administered only as part of the computer-based assessment and therefore all item development was done for computer delivery, although most of the trend items for financial literacy were originally developed for a paper-based administration.

As shown in Table 3.3, there was a paper-based instrument that was used in the PISA 2015 and the PISA 2018 cycles, which contain only items taken from cycles prior to PISA 2015. Only one of the participants administered these instruments. The other three paper-based participants used a “new” paper-based instrument that was first used in PISA for Development.

Une Heure (UH) form

Consistent with previous cycles, a special one-hour test, referred to as the “Une Heure” (UH) form, was prepared for students with special needs. The selected items were among the easier trend items (i.e., items developed prior to PISA 2015) in each core domain and had a reduced reading load. The UH form contained about half as many items as the other forms, with each cluster including from seven to nine items. In PISA 2022 the UH form was comprised of about 53% mathematics, 21% reading, and 26% science items. The UH form included two 15-minute clusters of mathematics (MU1 and MU2), one 15-minute cluster of reading (RU1) and one 15-minute cluster of science (SU1).

The UH form was accompanied by a special UH student background questionnaire that included only a subset of items from the regular background questionnaire (primarily trend items) in a single form design that was administered in CBA only. No PBA participants chose to administer the UH Form.

Assessment of financial literacy

The assessment of financial literacy was again offered as an international option in PISA 2022. The financial literacy instrument included trend items from the PISA 2012, PISA 2015, and PISA 2018 assessments, plus a few new units that were developed for PISA 2022. Financial literacy was administered only as a computer-based assessment.

Like in PISA 2018, the financial literacy assessment was administered to a separate sample of PISA-eligible students who took, in addition to the financial literacy assessment, reading or mathematics items. As with students sitting PISA as part of the main sample described in Chapter 2, the total testing time for each student was two hours (120 minutes) for the cognitive test.

The 2022 mathematics assessment framework[1]

For each PISA domain, an assessment framework is created to guide instrument development and interpretation in accordance with the policy requirements of the PISA Governing Board. The frameworks define the domains, describe the scope of the assessment, specify the structure of the test – including item format and the target distribution of items according to important framework dimensions – and outline the possibilities for reporting results. For PISA 2022, a subject matter expert group (SMEG) was convened to develop a framework for mathematical literacy under the guidance of RTI International and with input from the PISA Governing Board and Core A (ETS). A separate expert group, convened by ACT (Core B3), worked on creative thinking.

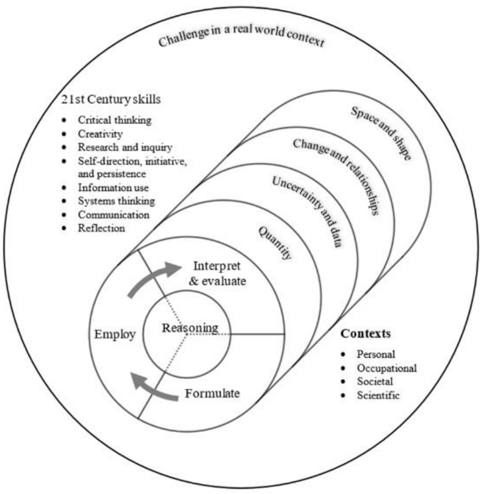
Mathematical literacy, for PISA 2022, is defined as follows:

> *Mathematical literacy is an individual's capacity to reason mathematically and to formulate, employ, and interpret mathematics to solve problems in a variety of real-world contexts. It includes concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to know the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective 21st century citizens.*

Additionally, the definition of mathematical literacy for PISA 2022 can be considered with respect to three interrelated concepts, which are represented in Figure 3.1 and will be explained in this section. These interrelated concepts are:

1. **Cognitive Processes**: mathematical reasoning and the problem-solving model
2. **Content Knowledge**: how the domain is organized into categories
3. **Contexts**: the real-world "setting" in which items are presented, including select 21st Century skills that are supported and developed as part of being mathematically literate.

Figure 3.1. Mathematical literacy for PISA 2022



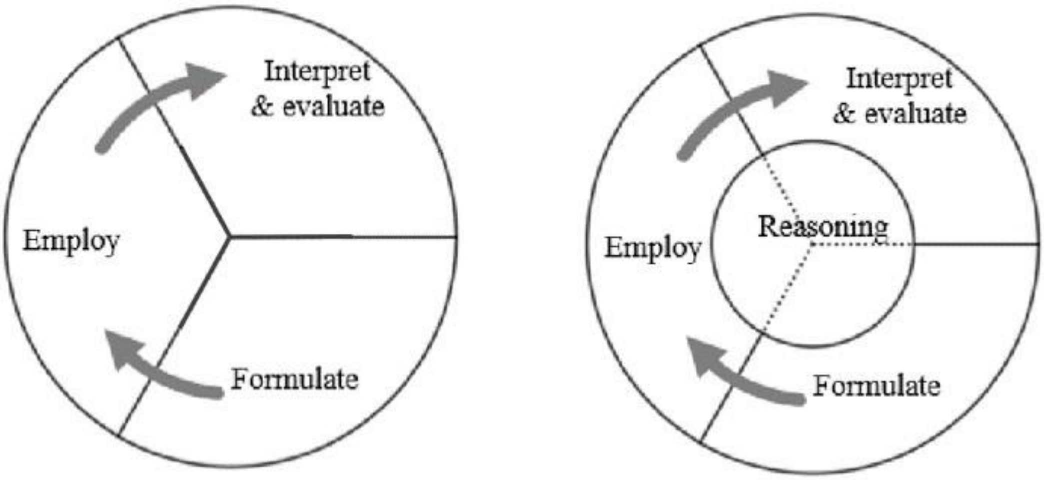
Mathematical Literacy – Cognitive Processes

For PISA 2022, the mathematical literacy domain describes mathematics in terms of four cognitive processes: reasoning, formulating, employing, and interpreting/evaluating.

Previous PISA mathematics frameworks used three cognitive processes (formulating, employing, and interpreting/evaluating), which formed the basis of the mathematical problem-solving model. For PISA 2022, reasoning was included as a separate cognitive process, but it is not a new concept in PISA mathematics. Reasoning – including both deductive (i.e., mathematical) and inductive (i.e., statistical) reasoning – has always existed as an underlying element to the problem-solving model and is considered a core aspect of being mathematically literate; therefore, the updated mathematics framework sought to highlight reasoning as both a central component underlying the processes in the problem-solving model, and as its own process.

Figure 3.2 shows the mathematical problem-solving model used in previous cycles and in the current cycle with reasoning as a fourth process. Note that even though the problem-solving model is comprised of multiple processes, each PISA mathematics item was written specifically towards one of the processes and students were not expected to utilize the full model to respond to each item. For example, a formulate item might assess if a student can write an equation to model a situation without requiring application of any processes/procedures (i.e., employing) or reflection on the result (i.e., interpreting/evaluating). The cognitive processes within each category are briefly defined below.

Figure 3.2. Cognitive processes and the mathematical problem-solving model: prior to 2022 (left) and for 2022 (right)



Reasoning Mathematically

Reasoning mathematically (both deductively and inductively) involves evaluating situations, selecting strategies, drawing logical conclusions, developing and describing solutions, and recognising how those solutions can be applied. Students reason mathematically when they:

- Identify, recognise, organise, connect, and represent
- Construct, abstract, evaluate, deduce, justify, explain, and defend
- Interpret, make judgements, critique, refute, and qualify

Formulating situationsmMathematically

Formulating situations mathematically refers to individuals being able to recognise and identify opportunities to use mathematics and then providing mathematical structure to a problem presented in some contextualised form, including reasoning about the constraints and assumptions in the problem, which may involve:

- Selecting an appropriate model from a list
- Identifying the mathematical aspects of a problem situated in a real-world context and identifying the significant variables
- Recognising mathematical structure (including regularities, relationships, and patterns) in problems or situations
- Simplifying a situation or problem in order to make it amenable to mathematical analysis (for example by decomposing)
- Identifying constraints and assumptions behind any mathematical modelling and simplifications gleaned from the context
- Representing a situation mathematically, using appropriate variables, symbols, diagrams, and standard models
- Representing a problem in a different way, including organising it according to mathematical concepts and making appropriate assumptions
- Understanding and explaining the relationships between the context-specific language of a problem and the symbolic and formal language needed to represent it mathematically
- Translating a problem into mathematical language or a representation
- Recognising aspects of a problem that correspond with known problems or mathematical concepts, facts or procedures
- Choosing among an array of and employing the most effective computing tool to portray a mathematical relationship inherent in a contextualised problem
- Creating an ordered series of (step-by-step) instructions for solving problems.

Employing mathematicalcConceptsf Facts, apd Procedures

Employing mathematical concepts, facts, and procedures refers to individuals being able to apply mathematical concepts, facts, procedures, and reasoning to solve mathematically formulated problems to obtain mathematical conclusions, including:

- Performing a simple calculation
- Drawing a simple conclusion
- Selecting an appropriate strategy from a list
- Devising and implementing strategies for finding mathematical solutions
- Using mathematical tools, including technology, to help find exact or approximate solutions
- Applying mathematical facts, rules, algorithms, and structures when finding solutions
- Manipulating numbers, graphical and statistical data and information, algebraic expressions and equations, and geometric representations
- Making mathematical diagrams, graphs, simulations, and constructions and extracting mathematical information from them
- Using and switching between different representations in the process of finding solutions
- Making generalisations and conjectures based on the results of applying mathematical procedures to find solutions

- Reflecting on mathematical arguments and explaining and justifying mathematical results
- Evaluating the significance of observed (or proposed) patterns and regularities in data

Interpreting, applying andeEvaluatinm Mathematicol Outcomes

Interpreting, *applying* and *evaluating* mathematical outcomes refers to individuals being able to reflect upon mathematical solutions, results or conclusions and interpret them in the context of the real-life problem that initiated the process, including:

- Interpreting information presented in graphical form and/or diagrams
- Evaluating a mathematical outcome in terms of the context
- Interpreting a mathematical result back into the real-world context
- Evaluating the reasonableness of a mathematical solution in the context of a real-world problem
- Understanding how the real world impacts the outcomes and calculations of a mathematical procedure or model in order to make contextual judgments about how the results should be adjusted or applied
- Explaining why a mathematical result or conclusion does, or does not, make sense given the context of a problem
- Understanding the extent and limits of mathematical concepts and mathematical solutions
- Critiquing and identifying the limits of the model used to solve a problem
- Using mathematical thinking and computational thinking to make predictions, to provide evidence for arguments, to test and compare proposed solutions.

Mathematical literacy – content knowledge

The content of the PISA mathematics assessment is divided into the same four categories that were used in previous PISA cycles: quantity, uncertainty and data, change and relationships, and space and shape. Even though PISA is not a curriculum-driven assessment, these four categories reflect content that is common to many school curricula (i.e., content that most 15-year-olds are likely to have encountered in school) and cover a range of topics that are considered central to the study of mathematics.

A brief description of each of the four content categories is given below.

- **Quantity**: number sense and estimation; quantification of attributes, objects, relationships, situations and entities in the world; understanding various representations of those quantifications, and judging interpretations and arguments based on quantity.
- **Uncertainty and data**: recognising the place of variation in the real world, including having a sense of the quantification of that variation, and acknowledging its uncertainty and error in related inferences. It also includes forming, interpreting and evaluating conclusions drawn in situations where uncertainty is present. The presentation and interpretation of data are also included in this category, as well as basic topics in probability.
- **Change and relationships**: understanding fundamental types of change and recognising when they occur in order to use suitable mathematical models to describe and predict change. Includes appropriate functions and equations/inequalities as well as creating, interpreting and translating among symbolic and graphical representations of relationships.
- **Space and shape**: patterns; properties of objects; spatial visualisations; positions and orientations; representations of objects; decoding and encoding of visual information; navigation and dynamic interaction with real shapes as well as representations, movement, displacement, and the ability to anticipate actions in space.

Below is a list of content topics based on the results of an analysis of desired learning outcomes from a sample of eleven countries from around the world. These topics can be applied to one or more of the four content categories, and this list is not intended to be exhaustive, but rather reflective of content that is deemed important for students preparing to either enter the workforce or pursue higher levels of education. Additionally, mathematics experts have added a few focus topics pertinent to the updated framework.

- **Growth Phenomena**: different types of linear and non-linear growth
- **Geometric Approximation**: approximating the attributes and properties of irregular or unfamiliar shapes and objects by breaking these shapes and objects up into more familiar shapes and objects for which there are formulae and tools
- **Computer Simulations**: exploring situations (that may include budgeting, planning, population distribution, disease spread, experimental probability, reaction time modelling etc.) in terms of the variables and the impact that these have on the outcome
- **Conditional Decision-Making**: using basic principles of combinatorics and an understanding of interrelationships between variables to interpret situations and make predictions
- **Functions**: the concept of function, emphasising but not limited to linear functions, their properties, and a variety of descriptions and representations of them. Commonly used representations are verbal, symbolic, tabular and graphical.
- **Algebraic Expressions**: verbal interpretation of and manipulation with algebraic expressions, involving numbers, symbols, arithmetic operations, powers and simple roots
- **Equations and Inequalities**: linear and related equations and inequalities, simple second-degree equations, and analytic and non-analytic solution methods
- **Co-Ordinate Systems**: representation and description of data, position and relationships
- **Relationships Within and Among Geometrical Objects in Two and Three Dimensions**: static relationships such as algebraic connections among elements of figures (e.g., the Pythagorean theorem as defining the relationship between the lengths of the sides of a right triangle), relative position, similarity and congruence, and dynamic relationships involving transformation and motion of objects, as well as correspondences between two- and three-dimensional objects
- **Measurement**: quantification of features of and among shapes and objects, such as angle measures, distance, length, perimeter, circumference, area and volume
- **Numbers and Units**: concepts, representations of numbers and number systems (including converting between number systems), including properties of integer and rational numbers, as well as quantities and units referring to phenomena such as time, money, weight, temperature, distance, area and volume, and derived quantities and their numerical description
- **Arithmetic Operations**: the nature and properties of these operations and related notational conventions
- **Percentages, Ratios and Proportions**: numerical description of relative magnitude and the application of proportions and proportional reasoning to solve problems
- **Counting Principles**: simple combinations
- **Estimation**: purpose-driven approximation of quantities and numerical expressions, including significant digits and rounding
- **Data Collection, Representation and Interpretation**: nature, genesis and collection of various types of data, and the different ways to analyse, represent and interpret them
- **Data Variability and its Description**: concepts such as variability, distribution and central tendency of data sets, and ways to describe and interpret these in quantitative and graphical terms
- **Samples and Sampling**: concepts of sampling and sampling from data populations, including simple inferences based on properties of samples including accuracy and precision

- **Chance and Probability**: notion of random events, random variation and its representation, chance and frequency of events, and basic aspects of the concept of probability and conditional probability

Mathematical literacy –cContexts

Context is the aspect of an individual's world in which a problem is set. All PISA mathematics items are set in a real-life context; however, that does not mean all the items are based on real-life events or scenarios. Some units are based on fictional but plausible scenarios where mathematics can be applied in various ways towards solving problems. The strategies used to solve a problem can be dependent on the context in which the problem is set, but care is taken to ensure that context-specific knowledge is not needed to solve any problem. The PISA 2022 mathematics assessment uses the same four contexts as in previous cycles, which are: personal, occupational, societal, and scientific. Note that there is no reporting by context but having these different classifications helped ensure that the items reflected a broad range of situations where mathematics could be encountered in real life. A brief description of each context follows.

- **Personal**: problems classified in the personal context category focus on activities of one's self, one's family or one's peer group. The kinds of contexts that may be considered personal include (but are not limited to) those involving food preparation, shopping, games, personal health, personal transportation, recreation, sports, travel, personal scheduling, and personal finance.
- **Occupational**: problems classified in the occupational context category are centred on the world of work. Items categorised as occupational may involve (but are not limited to) such things as measuring, costing and ordering materials for building, payroll/accounting, quality control, scheduling/inventory, design/architecture and job-related decision making either with or without appropriate technology. Occupational contexts may relate to any level of the workforce, from unskilled work to the highest levels of professional work, although items in the PISA survey must be accessible to 15-year-old students.
- **Societal**: problems classified in the societal context category focus on one's community (whether local, national, or global). They may involve (but are not limited to) such things as voting systems, public transport, government, public policies, demographics, advertising, health, entertainment, national statistics, and economics. Although individuals are involved in all of these things in a personal way, in the societal context category, the focus of problems is on the community perspective.
- **Scientific**: problems classified in the scientific category relate to the application of mathematics to the natural world and issues and topics related to science and technology. Particular contexts might include (but are not limited to) such areas as weather or climate, ecology, medicine, space science, genetics, measurement and the world of mathematics itself. Items that are intra-mathematical, where all the elements involved belong in the world of mathematics, fall within the scientific context.

Role of the mathematics expert group in item development

As the contractor for mathematics instrument development, Core A was responsible for working with the Mathematics Expert Group (MEG) to understand their vision for the range and types of items to be developed for PISA 2022. To facilitate the transition from the work of Core B1 (framework development) to the instrument development activities, Core A retained the members of the MEG who met under Core B1 to begin work on the updated mathematics framework in 2017, and which continued into 2018.

Core A's work with the MEG began in February 2018 and focused on the following tasks:

- describing the kinds of items needed to assess the skills and abilities in the domain, particularly defining the behaviours associated with mathematical reasoning
- reviewing and understanding the proposed assessment design to determine the distribution of mathematics content across the major components of the framework
- defining the intersection between the kinds of functionality that might be desirable for measuring the construct and the functionality that was practical to implement in the assessment
- developing illustrative examples of tasks that reflect some of the new content and possible functionality of the platform.

Work with the subject matter experts continued beyond the initial meetings and went through instrument development and data analysis. For mathematics, MEG members reviewed assessment tasks as they were developed, provided input into the analysis of the field trial data, approved the set of items for the main survey, and worked with development and analysis staff to develop the described proficiency scales used for reporting the PISA 2022 results.

PISA 2022 test development

Test development for the PISA 2022 cycle began in early 2018 and focused on the development of mathematics items for a computer-based assessment. For example, the following list from the updated mathematics framework presents a few possible ways in which the computer platform was leveraged to assess mathematical literacy:

- Simulation in which a mathematical model has been established and students can change the variable values to explore the impact of the variables to create “an optimal solution”.
- Fitting a curve (by selecting a curve from a limited set of curves provided) to a data set or a geometric image to determine the “best fit” and using the resulting best fit curve to determine the answer to a question about the situation.
- Budgeting situations (e.g., online store) in which the student must select combinations of products to meet achieve a range of objectives within a given budget.
- Purchase simulation in which the student selects from different loan and associated repayment options to purchase an item using a loan and meeting a budget. The challenge in the problem is to understand how the variables interact.
- Problems that include visual coding to achieve a given sequence of actions.

However, it is important to note that not every new unit or item was developed requiring the use of some type of computer-based functionality. Item development efforts strove to maintain a balance between purposeful uses of the available technology, but the focus was always on assessing mathematical literacy and not information and communications technology (ICT) skills. To help with this last point, in addition to the general orientation, which provided students with an overview of the platform and standard functionality (e.g., navigating the interface, using drag-and-drop, selecting *vs* entering a response, etc.), item-specific tutorials and practice opportunities were built-in to every unit/item that used “novel” functionality (e.g., spreadsheets) before students could advance to the actual items. Even after students advanced past the requisite practice screens, instructions for using the specific tool in a unit were always available via drop-down menu at the top of each screen in the unit.

Computer-based assessment: Screen design and interface

The screen design and interface developed for the PISA 2015 cycle, and which was used for the PISA 2018 cycle, was again used for the PISA 2022 cycle.

Navigation

As in PISA 2015 and PISA 2018, students could navigate through the items as needed. For most units, students were able to move back and forth between items *within* a unit. They were not, however, able to move back and forth *between* units. Once students clicked on the “NEXT” arrow on the final item in a unit, a dialog box displayed a warning that the student was about to move on to the next unit and it would not be possible to return to previous items. At this point, students could either confirm that they wanted to go on or cancel the action and continue with the unit on which they had been working. There were a few exceptions regarding navigation *within* units where students were not permitted to return to a previous item. These within-unit restrictions were primarily used when information in a later item might help with answering an earlier item or in instances where it was desired that the students either have access or no longer have access to a tool. When students would click on the “Next” arrow a message would pop up indicating that it, “…will not be possible to return to this work.”, and students would have to click on “Yes” or “No” to indicate if they were ready to continue to the next question in the unit.

Response modes

Across all domains, PISA 2022 included items requiring one of five different response modes:

- **Selection items**: single-selection multiple choice; multiple-selection multiple choice (click on one or more options); complex multiple choice (table with statements and typically several yes/no or true/false options); data selection (selecting rows of student-generated data to support or refute a claim); or click on an image
- **Numeric entry**: only numbers, commas, periods, dashes, and backslashes could be entered
- **Text entry**: a scrolling text box that did not constrain the length of a student response (consistent with what was possible for paper-and-pencil items); or certain mathematics items that used the equation editor
- **Drop-down menus**
- **Drag-and-drop** (including use of a slider).

Orientations

A general orientation introduced students to the screen design and those response modes that were common across most domains. Students received this orientation before beginning the test. Prior to beginning each section of the test, students received a very short domain-specific orientation with instructions specific to the domain in that section. For example, before beginning the mathematics section of the assessment, students were introduced to the calculator and the equation editor and given an opportunity to practice using each of these tools.

Trend items

The computer-based trend reading item pool contained 197 items (152 developed in PISA 2018 and 45 developed prior to PISA 2015), in addition to the 60 reading fluency items. Of the 197 trend reading items, 64 were human coded.

The computer-based trend science item pool contained 115 items (76 developed in PISA 2015 and 39 developed prior to PISA 2015) in six clusters. For science, these were the same trend clusters that were used in PISA 2018 and which remained intact for the PISA 2022 field trial and main survey. Of the 115 trend science items, 32 were human coded.

The computer-based trend mathematics item pool contained 74 items, 16 of which were human coded. The financial literacy item pool contained 46 items (five items developed in 2022 and 41 items developed prior to 2022). There were 16 human-coded items in financial literacy.

For the “new” paper-based assessment there were: 66 science items (nine human coded), 66 reading items (37 human coded), and 62 mathematics items (40 human coded). For the one country taking the older paper-based assessment, there were: 85 science items (32 human coded), 87 reading items (51 human coded), and 71 mathematics items (38 human coded).

New items

For PISA 2022, test development occurred for the domains of mathematics, creative thinking, and financial literacy. To prepare for the implementation of the multistage adaptive design in the main survey, twelve 30-minute clusters of new items were developed for mathematics. In total, 61 new units with 182 new mathematics items were selected and included in the field trial. For financial literacy, three new units were developed with five total new items, all of which were retained for the main survey.

For information on the development of creative thinking, refer to chapter 4 of this technical report.

International test development team

Test development efforts for the mathematics assessment were coordinated by ETS as the Core A Contractor. As is the case with any large-scale international survey, it is important that the material used in PISA reflect the range of contexts and experiences of students across participating countries/economies. One way to meet this goal was by convening an international team of item developers. For PISA 2022, the international test development team included individuals from the University of Luxembourg and the University of Liège. A second way to meet this goal was to work with countries/economies on development of materials. Core A provided countries/economies with a range of opportunities for participation during the development process.

National submissions

The active involvement of countries/economies in the development process is important for the instruments to be internationally valid and representative. Thus, it was important to ensure that the final item pool reflected the international context of an assessment such as PISA. For example, Core A offered two item-development workshops, as well as accepted item submissions via the PISA Portal. This phase of the item-development process primarily occurred between April and September of 2018.

Item development workshops and submissions

Two item-development workshops were offered as part of the PISA 2022 efforts to involve countries/economies in the test development process. These took place in May and June of 2018 in Princeton, NJ, USA and in Liège, Belgium, respectively. Fifty-three participants from 29 countries/economies attended these workshops. From the test developers’ point of view, the workshops made the development process more efficient because of the in-person training and collaboration, which was reflected in the quality of items that came out of the workshop and the items that were submitted subsequently. These workshops allowed representatives from countries/economies to interact and share ideas and expertise with members of the test development teams. Participants in the workshops wrote and reviewed items during the workshop and received some “real-time” feedback from the test development teams. The workshops also provided a venue to exchange ideas for ways to assess the content in the updated framework.

Overall, the item writing workshops and item submission process were extremely successful and ultimately resulted in 44 units with 130 new mathematics items that were used on the main survey. Additional new units were developed internally by experienced mathematics assessment specialists at ETS.

Item reviews

Newly developed units were submitted for translatability review at the same time they were released for country/economy review. Linguists representing different language groups provided feedback on potential translation, adaptation and cultural issues arising from the initial wording of items. Experts at cApStAn and the Translation Referee for the PISA 2022 cycle alerted item developers to both general wording patterns and specific item wording that was known to be problematic for some translations and suggested alternative wordings. This provided item developers with the opportunity to make wording revisions at an early stage. In some cases, revisions were performed by simply using the alternatives provided and in others by working with cApStAn to explore a suitable wording that would lend itself to being translated without compromising what was being assessed.

All newly developed mathematics and financial literacy items were released for country/economy review prior to the field trial. Countries/economies had two weeks to preform reviews and submit feedback on all draft items. Mathematics items were released in four batches between September and December2018. Test developers received review forms from 40 countries/economies for Batch 1, 54 countries/economies for Batch 2, 53 countries/economies for Batch 3, and 54 countries/economies for Batch 4. The newly developed financial literacy items were released in one batch, which was reviewed by 19 countries/economies.

Preparation of the French source version for all new mathematics units provided another opportunity to identify issues with the English source version related to content and expression. Development of the two source versions helped ensure that items were as culturally neutral as possible, identified instances where wording could be modified to simplify translation into other languages, and specified where translation notes would be needed to ensure the required accuracy in translating items to other languages.

In addition, cognitive labs were conducted by the University of Luxembourg and by the University of Liège. A total of 11 new mathematics units (five at the University of Luxembourg and six at the University of Liège) were evaluated as part of these cognitive labs. The 11 units contained a mixture of new content and/or new functionality. These cognitive labs provided useful information to test developers concerning students' understanding of the content and what the items were assessing, response formats, the clarity of instructions and introductions, how the interactive elements functioned, and timing. The results led to improvements in the 11 items used in the cognitive labs, as well as provided test developers with some general guidelines to apply to all new units.

Selection of new items for the field trial

The PISA 2022 item-development process produced a total of 61 new mathematics units with 182 items that were selected for use in the field trial. Items were selected for inclusion in the field trial based on country/economy reviews, feedback from the mathematics expert group and the distribution of items across the key categories as defined in the framework. Of these 182 new mathematics items, 74% were submitted by participating countries/economies (from the item-development workshops and item submissions via the Portal), and 26% were developed by ETS's test development team.

Field trial

The PISA 2022 field trial data collection timeline began in March 2020 but was quickly disrupted by the COVID-19 global pandemic. Even though 17 participants were able to complete and limited field trial in

2020, most participants postponed the field trial until 2021. Of the 17 participants that administered the limited field trial in 2020, six participants chose to readminister the field trial in 2021. In total, 83 countries/economies (79 that administered on computer and four that administered on paper), consisting of 142 language versions, participated in this cycle of PISA. Assessment materials were prepared and released based on the field trial testing dates for each country.

Preparation of field trial instruments

As part of the quality control procedures for PISA 2022, the Core A contractors continued to assume responsibility for assembling the assessment instruments for both paper- and computer-based countries/economies. Countries/economies were responsible for translating all new material and performing both linguistic and layout quality control checks for trend and new items.

Computer-based trend items

Countries/economies that participated in the PISA 2015 and/or PISA 2018 computer-based assessment, were given access to the existing XLIFFs (XML Localization Interchange File Format) files from the previous administration and had the opportunity to review their materials for any errors or necessary updates.

For countries/economies switching from a paper- to a computer-based assessment, the Core A contractors copied their material into the computer-readable XLIFF that was used for the computer-based instruments. This was done both as a quality control process and to reduce the number of tasks assigned to countries given the short development timeline. Once the XLIFF files were created, the Core A contractors asked the countries/economies to perform a review comparing the new computer versions against PDF files of their paper-based items.

In both cases, countries/economies were asked to document any errors, which included typographical mistakes or text errors introduced in the process of copying and pasting across formats. All content issues identified by countries were reviewed by verifiers on the linguistic quality control team and, if approved, the verifiers made the needed change in the computer files. If countries identified any serious layout issues, those were reviewed and corrected by the Core A technical team. As an additional quality control check, the Core A contractor also performed layout checks of all items in all languages to identify errors that may have been missed.

Computer-based new items

All new mathematics and financial literacy items needed to be translated following the translation and reconciliation processes defined in the PISA standards. Following verification of the translations and the correction of any remaining errors, countries/economies were asked to sign off on their cognitive materials and those files were then considered locked for use.

Preparing the field trial national student delivery systems (SDS)

The Student Delivery System (or SDS) was again used for PISA 2022 and was a self-contained set of applications for delivery of the computer-based cognitive assessments and computer-based student background questionnaires. A master version was assembled first for countries to test within their national IT structure. This allowed countries/economies to become familiar with the operation of the SDS and to check the compatibility of the software with the computers being used to administer the assessment.

Once all the cognitive and background materials were approved and locked, the SDS was assembled and tested first by the Core A technical team. The SDS was then released for national testing. Countries/economies were asked to check their SDS following a specific testing plan provided by Core A and to identify any residual content or layout issues. If issues were identified, they were corrected by the Core A technical team, and a second SDS was released. Once countries/economies signed off on their SDS, their instruments were released for the field trial.

Paper-based instruments

National versions of the paper-based trend clusters were again prepared by the Core A contractor. To better ensure comparability of the paper-based assessment materials across countries/economies and languages, digital files of the booklets were centrally created by Core A and then reviewed and approved by countries/economies. Those countries/economies who were new to PISA needed to translate those materials following the standard translation and verification process. Existing paper-based countries/economies needed to update the common booklet parts (which included the cover, general instructions, formula sheet for mathematics, and the acknowledgements page), while new countries/economies had to translate these materials.

The approved clusters were then assembled into field trial paper booklets by the contractors using a centralised process that ensured comparability of layout. As a final step, the assembled booklet files were released and participants performed a final review and Core A implemented any changes, as needed. This process continued until National Centres had approved, print-ready files.

Field trial coding

Coding guides for trend items were compiled by Core A based on previous national versions. Continuing a practice that started in the PISA 2018 cycle, separate guides were updated/prepared for computer-based and paper-based participants.

The English master versions of the new mathematics and new financial literacy coding guides were released in draft form prior to the coder training meeting in January 2020. Based on discussions at that meeting, the coding guides were finalised and the updated English versions, along with the French source version (for new mathematics), were released to countries/economies in March 2020, prior to the beginning of the field trial data collection period. For the trend domains, a similar process was followed but with corrections to the guides restricted to correcting outright mistakes or providing some additional examples for clarification purposes.

Field trial coder training

The international field trial coder training was held in-person in January 2020 with sessions for all domains, including separate sessions for paper-based participants. The goals of the training included having attendees (master coders) develop an in-depth understanding of the coding rules for each item, so they would be prepared to train coders in their countries/economies and reaching consensus about the coding rules to better ensure consistency of coding both within and between countries/economies and across cycles. Trainers reviewed the content of the coding guides, general coding principles, common problems, and guidelines for applying special codes. Sample student responses were provided, and attendees were required to code them. When there were disagreements about coding for an item, they were discussed so that all attendees understood the specific coding rules for that item.

Due to the postponement of most field trials in 2020, field trial coder trainings were held virtually in January and February of 2021 for new mathematics, creative thinking, and financial literacy (new items only). The virtual training also included a recap of general coding principles and procedures, as well as a refresher training on the open-ended item coding system (the OECS).

Field trial coder queries

As was the case during previous cycles, Core A set up a coder query service for the PISA 2022 field trial. Countries/economies were encouraged to send queries to the service so that a common adjudication process was consistently applied to all coder questions about human-coded items. Queries were reviewed, and responses were provided by domain-specific teams that included item developers, and for trend items, members of the response team from previous cycles. For the new items, the coder query service was particularly valuable as it provided item developers with a better sense of the "range" of responses that could be expected, which in turn led to refinements of the coding guide.

In addition to responses to new queries, the queries report included the accumulated responses from previous PISA cycles. This helped foster consistent coding of trend items across cycles. The report was updated and posted weekly on the PISA Portal for National Centres.

Field trial outcomes

The PISA 2022 field trial was designed to yield information about the quantity and quality of data collected as well as to prepare the multistage adaptive testing design for the main survey. More specifically, general goals of the field trial included collecting and analysing information regarding:

- the quantity of data and the impact, if any, that survey operations had on that data
- the functioning of the computer-delivery platform
- the quality of the items including both those items that were newly developed for computer-based delivery and those that were adapted from earlier cycles
- the use of the data to establish reliable, valid, and comparable scales based on item-response theory (IRT) models in both the paper- and computer-based versions.

Overall, the field trial achieved all the stated goals. This information was crucial for the selection and assembly of the main survey instruments and for refining survey procedures where necessary. Furthermore, the field trial results confirmed the feasibility of introducing multistage adaptive testing in the main survey as unit order effects were found to be negligible.

The field trial analyses were conducted in batches based on data submission dates. Most of the analyses implemented to evaluate the goals noted above were based on data received from countries by 31 July 2021. That batch included data from 52 countries/economies, of which 41 carried out the field trial in 2021 and 11 in 2020. Of those, one participant administered the paper-based assessment, 51 administered the computer-based assessment, and one conducted data collection in 2020 and in 2021. The field trial analyses were updated after receiving additional data, which increased the number of participating countries/economies to 80 by the end of 2021. Of these, three participants implemented the field trial as a paper-based survey and 77 that implemented it as a computer-based survey.

Main survey

The PISA 2022 main survey was conducted between March and December 2022. Most countries/economies completed the main survey data collection by May 2022. In preparation for the main survey, countries reviewed items based on their performance in the field trial and were asked to identify any serious errors with the items still in need of correction. The Core A contractors worked with countries/economies to resolve any remaining issues and prepare the national instruments for the main survey.

National item review following the field trial

The item feedback process began in September 2021 and was conducted on a rolling basis based on main survey start dates.

Following release of the field trial data, countries/economies completed item feedback forms that included flags for any items that had been identified as not fitting the international trend parameters. Flagged items were reviewed by national teams and participants were asked to provide comments about these specific items where they could identify serious errors. Requests for corrections were reviewed by Core A, and if approved, implemented.

Item selection for Mathematics

The initial selection of mathematics items for the main survey was a collaborative effort between the test development team and psychometricians based mostly on item statistics from the first batch of field trial data. The first step was to generate a list of flagged items based on the following statistics and associated criteria:

- Scoring reliability rater agreement (below 0.92%)
- Percentage of omitted responses (above 20% in each country/economy)
- IRT discrimination and difficulty parameters ($a < 0.1$ or $|b| > 5$)
- IRT MD and RMSD fit statistics (0.15 for new items and 0.20 for trend items)
- Item-level and unit-level response time (more than three minutes per item)

Next, the list of flagged items was reviewed from a content perspective with an aim towards removing any items with possible content flaws or items that were not able to be scaled appropriately. Another factor influencing main survey item selection was feedback from National Centres. Participants were asked to rate each item from the field trial with regards to how common the content was to their national curriculum using the following values: 1 = not in curriculum, 2 = in some curriculum, or 3 = standard curriculum material. They were also asked to rate each item on how relevant each item was to “preparedness for life” using the following values: 1 = not relevant, 2 = somewhat relevant, or 3 = highly relevant. The final step was a review of the remaining items, based on the degree to which they had been flagged (i.e., items that had stronger statistics were kept over those with weaker statistics), but also to determine if removing certain items would lead to an imbalance in domain representation (according to the target construct distributions in the framework), and to check for any changes to how a unit would function if an item or items were removed (e.g., if an item was removed that introduced or built on the scenario which the unit was written about, so that a subsequent item became unclear because it referenced information no longer present in the unit).

Once this review process was completed, a total of 30 mathematics items (22 new items and eight trend items) were dropped from across 20 units (15 new units and five trend units). A total of seven units (five new units and two trend units), which consisted of 17 items (12 new items and five trend items), were dropped completely. The remaining dropped items came from units where one or more items were retained for the PISA 2022 main survey. The resulting computer-based mathematics item pool for the main survey contained 99 total units (56 new units and 43 trend units) and 234 total items (160 new items and 74 trend items). For the paper-based designs, no items or units were dropped following the field trial.

Assigning mathematics units to the multistage adaptive design

The multistage adaptive design for mathematics expanded and enhanced what was accomplished with the adaptive design for reading in PISA 2018. Test assembly for PISA 2022 was implemented in four steps:

1. Assemble non-overlapping parallel item sets.

2. Assemble core and adaptive testlets from each item set.
3. Assemble multistage paths using the core and adaptive testlets.
4. Assemble linear forms using the core testlets.

Also, for PISA 2022 automated test assembly (ATA) was employed to assemble the test paths and forms via mixed-integer linear programming. This was done using commercial software. The software provided a principled design approach and was able to much more efficiently handle the large number of decision variables and constraints at each step of the assembly process. Note that there was some flexibility with constraints when creating the core and adaptive testlets as long as all constraints were met in the full path or form. A summary of some key features – framework distributions and psychometric properties – of the four steps follows.

Non-overlapping parallel item sets

Each of the three item sets contained 78 items and 33 units. Each unit only appeared in one item set. The maximum score of each set was either 99 or 100 points. Each set contained approximately 27% trend items. Approximately 85% of the items in each set were machine coded, and across all sets there were approximately equal numbers of items for each of the four major item types used in PISA (simple multiple choice, complex multiple choice, computer-scored open response, and human-coded open response). Each set contained approximately 24% of items from change and relationships, 32% from quantity, 18% from space and shape, and 26% from uncertainty and data. Each set contained approximately 32% employ items, 21% formulate items, 24% interpret/evaluate items, and 23% reasoning items.

Core and adaptive testlets from each item set

Each of the core testlets in the three item sets contained from three to five, three to six, or four to five units, and nine to 10 total items. The maximum score per core testlet, across all items sets, was from 12 to 14 points, of which human-coded items contributed from two to four points (the number of human-coded items in each core testlet ranged from one to two or one to three across all item sets). The maximum number of common items was set at six, so the percent overlap was either 27% or 28% depending on the item set. Percent overlap is the number of test pairs with overlap divided by the total number of test pairs. The core testlets had a percent connectedness of either 20% or 21%, depending on the item set. Percent connectedness is the number of unit pairs in tests divided by the total number of unit pairs. The median total response times for the core testlets ranged between 11 and 13 minutes across all item sets.

Each of the stage 1 adaptive testlets in the three item sets contained from three to five or three to six units, and nine to 10 total items. The maximum score per stage 1 testlet, across all items sets, was from 12 to 14 points, of which human-coded items contributed from two to three or two to four points (the number of human-coded items in each stage 1 testlet ranged from one to two or zero to three across all item sets). The percent overlap ranged from 25% to 27%, depending on the item set. The stage 1 testlets also had a percent connectedness of either 20% or 21%, depending on the item set. The median total response times for the stage 1 testlets also ranged between 11 and 13 minutes across all three item sets.

Each of the stage 2 adaptive testlets in the three item sets contained from three to five or from three to six units, and nine to 10 total items. The maximum score per stage 2 testlet, across all items sets, was from 12-13 or 11-14 points, of which human-coded items contributed from one to two, two to three, or zero to five points (the number of human-coded items in each stage 2 testlet ranged from one to two, one to three, or zero to three across all item sets). The percent overlap ranged from 23% to 26%, depending on the item set. The stage 2 testlets had a percent connectedness of either 19% or 20%, depending on the item set. The median total response times for the stage 2 testlets again ranged between 11 and 13 minutes across all item sets.

Multistage paths using the core and adaptive testlets

A total of 192 adaptive paths in the mathematics assessment were implemented for the PISA 2022 main survey. The number of units per path ranged from 10 to 16 with a median of 13 units. The number of items per path ranged from 28 to 30 with a median of 30 items. The number of trend mathematics items ranged from 3 to 16 with a median of 9, while the number of new mathematics items ranged from 14 to 27 with a median of 20. The median number of items by content area for each path was seven for change and relationships, 10 for quantity, five for space and shape, and seven for uncertainty and data. The median number of items by process for each path was nine for employ, six for formulate, seven for interpret/evaluate, and seven for reasoning. For both the content areas and the process classifications, the percentage distributions in each testlet mirrored the distributions of the entire mathematics item pool. Each unit appeared on average in 24.5 paths. The overlap percentage across all 192 paths was 75% (i.e., 75% of the possible pairs of paths have at least one unit in common). The percentage of observed unit pairs was 78% (i.e., 78% of the possible pairs of units were observed). For comparison, in PISA 2018, the percentage of observed unit pairs in the reading MSAT design was only 55%.

Linear forms using the core testlets

A total of 48 linear forms were in the PISA 2022 main survey mathematics assessment. The linear forms were comprised of the 48 core testlets. The number of units per form ranged from 11 to 15 with a median of 13 units. The number of items per form ranged from 29 to 30 with a median of 30 items. The number of trend mathematics items ranged from 1 to 19 with a median of 10, while the number of new mathematics items ranged from 11 to 29 with a median of 20. The median number of items by content area for each form was six for change and relationships, 10 for quantity, five for space and shape, and eight for uncertainty and data. The median number of items by process for each form was nine for employ, five for formulate, seven for interpret/evaluate, and seven for reasoning. For both the content areas and the process classifications, the percentage distributions in each linear form mirrored the distributions of the entire mathematics item pool.

After the four steps above were completed by the psychometrics team, all the proposed testlets were reviewed by the mathematics development team to look for any potentially problematic unit pairings (e.g., having multiple units within a testlet that assess the same construct) and to propose recommended changes. The development team then worked closely with the psychometricians to determine the effect the proposed changes would have on the design, and to make additional changes if needed. Once the unit pairings in each testlet were finalized, the development team made recommendations for how to order the units within each testlet.

Review by the Mathematics Expert Group

Once the item selection was complete and the units were assigned to the multistage adaptive design, Core A psychometricians performed simulation studies to assess the performance of the design using the preliminary item parameters obtained from the field trial. The details of these simulation studies are described in Yamamoto, Shin and Khorramdel (2018[1]). In short, the simulation studies suggested that the item parameters could be recovered well with minimal errors and that the proposed multistage adaptive design would improve the measurement precision for all ranges of skill distribution, particularly at the lower and higher ends of distribution. Specifically, the simulation study showed a gain in measurement precision of 10.6% at the lowest proficiency level, and a 13% gain at the highest proficiency level.

Given that the multistage adaptive testing design consisted of 192 possible paths, it was not possible for the mathematics experts to review all those combination of item sets and make recommendations for the selection. Instead, at the MEG meeting following the field trial, a thorough explanation of the item selection process and the characteristics of the main survey item pool were presented and discussed. The item pool was evaluated at a holistic level, considering the representation of the content areas and cognitive processes across the entire pool, including the distributions of difficulty and construct representation within each stage of the multistage adaptive design. At the end of the meeting, the experts signed off on the main survey item pool and the multistage adaptive design.

Construct coverage

The set of mathematics items for the main survey was relatively well balanced in terms of construct representation, based on the overall distributions recommended in the frameworks.

A total of 234 items – 74 trend and 160 new items – were selected for the computer-based mathematics assessment, and those 234 items represent a total of 253 possible score points. Table 3.5 shows the item counts, score points and percentage of score points by cognitive process and by content area for the main survey CBA mathematics items.

Of the 160 new items retained for the main survey, 74% were originally submitted by countries/economies (from either the item-development workshops or item submissions) and 26% were created by test developers at ETS.

Financial literacy

Item selection for financial literacy was based on classical item analyses. All five new items were retained for the main survey and two trend items – one from each cluster – were recommended by the PISA Technical Advisory Group (TAG) to be dropped based on concerns over the amount of time that students were spending on those two items. A total of 46 items (41 trend and 5 new) were used in the main survey financial literacy assessment. Table 3.6 shows the distributions of the 46 financial literacy items across the two aspects of the framework: process and content.

The paper-based and computer-based item counts for reading, mathematics, science, creative thinking, and financial literacy in both the field trial and main survey are presented in Table 3.7.

Preparation of data collection instruments

Preparing the main survey national student delivery systems (SDS)

The process for creating the main survey student delivery system (SDS) followed the approach used during the field trial, beginning with assembly and testing of the master SDS followed by the process for assembling national versions of the main survey SDS.

After all components of the materials were agreed upon, they were digitally locked, and it was not possible to edit or change them. This included the questionnaires and cognitive instruments. The student delivery system was then assembled and tested first by Core A. Countries/economies were then asked to check their SDS and identify any remaining content or layout issues. Once countries/economies signed off on their national SDS, their final systems were released for the main survey.

Preparing main survey paper-based instruments

As in the field trial, national versions of the main survey paper-based booklets were centrally prepared by the Core A contractor to better ensure comparability of the paper-based assessment materials across participants and languages. Once the workflow for reviewing field trial data and requesting changes to items was completed, and the common booklet parts (i.e., cover page, formula sheet, general instructions) were updated as needed, the approved materials were assembled into main survey booklet files by Core A. The booklet files were then sent to the countries/economies for review. If any changes were needed, Core A would implement them, and the process for reviewing the files would repeat until the National Centre approved all files for printing.

Main survey coding

Coder training for the main survey was conducted virtually for all domains. For mathematics and creative thinking, full trainings were offered for all main survey items (trend and new). The trainings for reading and science were targeted on items that were typically more challenging to code (e.g., items with low reliability rates or items with a high number of coder queries). The training for financial literacy covered all the new items but was targeted for the trend items, using the same criteria that reading and science used to identify items.

The coder query service was again used in the main survey, as it had been in the field trial, to assist countries in clarifying any uncertainty around the coding process or particularly challenging responses. Queries were reviewed, and responses were provided by domain-specific teams including item developers and members of the response team from previous cycles. Revisions were made to the coding guides for mathematics and creative thinking, and to the new financial literacy items following the field trial. The coder queries helped test developers see response categories that were not anticipated during the initial development of the coding guide. Thus, based on the queries received, test developers made some coding guides clearer and added sample responses to the guides to better illustrate the range of, and different types of, responses. Workshop examples were also enhanced by adding more authentic student responses that better illustrated the boundaries between full credit, partial credit (if applicable) and no credit. Following the international coder trainings, additional revisions were made to the mathematics, creative thinking, and financial literacy (new items only) coding guides in response to discussions that took place during the trainings.

Released items to illustrate the framework

As has been the case in previous PISA cycles, several items were released to the public domain at the time of publication of the PISA 2022 results to illustrate the kinds of items included in the assessment. Following the field trial, a list of proposed units to release was reviewed by the MEG and the OECD, and after the main survey, another list of proposed units to release was reviewed by the MEG and the OECD. The following four new mathematics units were approved for release after the field trial: *Car Purchase* (2 items), *DVD Sales* (3 items), *Moving Truck* (2 items), and *Spinners* (3 items). After the main survey, the following four new mathematics units were approved for release: *Solar System* (2 items), *Triangular Pattern* (3 items), *Points* (1 item), and *Forested Area* (4 items). These units are available at *www.oecd.org/pisa*.

References

Yamamoto, K., H. Shin and L. Khorramdel (2018), “Multistage adaptive testing design in international large-scale assessments”, *Educational Measurement: Issues and Practice*, Vol. 37/4, pp. 16–27. [1]

Notes

1. For a complete description of the PISA 2022 Mathematics Framework, please visit the site https://pisa2022-maths.oecd.org.

Chapter 3 tables

Table 3.1. Number of PISA items by core domain and across cycles in the main survey

	2000	2003	2006	2009	2012	2015	2018	2022
Reading	129	28	28	131	44	103	245	197
Mathematics	43	84	48	35	109	83	83	234
Science	45	34	103	53	53	184	115	115

Note: Red font color = Major domain for that cycle.
For the 2015 and 2018 cycles, the computer-based mathematics instrument contained 82 items, while the equivalent paper-based instrument contained 83 items. This is because there was one item that was not able to be transitioned to a computer-based delivery in 2015 (the item requires students to draw on a map).
The number of mathematics items in the 2022 cycle includes 74 "trend" items (i.e., items developed prior to this cycle) and 160 "new" items (i.e., items developed this cycle).

Table 3.2. Domain coverage for PISA 2022: CBA

	Field trial		Main survey		Total items – 2022 MS
Domain	New	Trend	New	Trend	
Reading Literacy	No new item development for 2022	Adaptive design: 197 items	No new item development for 2022	Same as Field Trial Trend	197
Scientific Literacy	No new item development for 2022	6 clusters: 115 items (76 from the 2015 cycle; 39 used prior to 2015)	No new item development for 2022	Same as Field Trial Trend	115
Mathematical Literacy	12 clusters: 182 items	7* clusters: 82 items (All items used in 2018 and taken from the 2012 cycle)	Adaptive design: 160 items	Adaptive design: 74 items	234
Creative Thinking	5 clusters: 38 items	New domain – no trend items	5 clusters: 36 items	New domain – no trend items	36
Financial Literacy	3** units: 5 items	2 clusters: 43 items	5 items	41 items	46

Note: Each cluster was designed to take approximately 30 minutes of testing time.
* For the PISA 2022 cycle field trial, there were actually 7 trend mathematics clusters because all computer-based participants administered the units from clusters M6a ("standard items") and M6b ("easier items"). In previous administrations, participants administered either M6a or M6b but not both.
** There are two financial literacy clusters - F1 and F2 - used in both the field trial and main survey this cycle. However, only 3 new units (5 total items) were developed for this cycle, and they were distributed across the two existing clusters (two new units in cluster F1 and one new unit in cluster F2).

Table 3.3. Main survey domain coverage for PISA 2022: PBA

PBA Instrument Used by One Participant this Cycle	
Domain	Field trial and main survey
Reading	6 clusters: 87 items Same set of items that all PBA participants used in 2018 and 2015 Prior to 2015, these items were last used in 2012 and 2009
Science	6 clusters: 85 items Same set of items that all PBA participants used in 2018 and 2015 Prior to 2015, these items were last used in 2012, 2006 and 2003
Mathematics	6 clusters: 71 items Same set of items that PBA participants used in 2018 and 2015 These items were all taken from the 2012 cycle
New Instrument Used by All Other PBA Participants this Cycle	
Domain	Field trial and main survey
Reading	4 clusters: 66 items*
Science	4 clusters: 66 items
Mathematics	4 clusters: 63 items*

Note: * There are 64 items in the new PBA mathematics assessment; however, one of the items is actually a reading item (it is in a set that contains a mathematics and a reading item), so there are only 63 items that contribute towards the mathematics scale.

Table 3.4. Main survey computer-based UH form design

Form	Cluster 1	Cluster 2	Cluster 3	Cluster 4
99 (UH)	MU1	MU2	RU1	SU1

Note: Where M = mathematics, R = reading, and S = science.

Table 3.5. Item counts and score points of the main survey CBA mathematics items by framework categories

	Trend Items	New Items	Combined (Trend + New)	Dichotomously Scored Items (1 point each)	Polytomously Scored Items (2 points each)	Total Score Points*		Framework Recommendation
Cognitive process	**Count**	**Count**	**Count**	**Count**	**Count**	**Points**	**%**	**%**
Formulating situations mathematically	11	37	48	47	1	49	19%	25%
Employing mathematical concepts, facts and procedures	24	51	75	72	3	78	31%	25%
Interpreting, applying and evaluating mathematical outcomes	10	47	57	55	2	59	23%	25%
Reasoning	29	25	54	41	13	67	26%	25%
Total	74	160	234	215	19	253	100%	100%
Content area	**Count**	**Count**	**Count**	**Count**	**Count**	**Points**	**%**	**%**
Change and relationships	17	38	55	50	5	60	24%	25%
Space and shape	17	26	43	39	4	47	19%	25%
Quantity	21	55	76	71	5	81	32%	25%
Uncertainty and data	19	41	60	55	5	65	26%	25%
Total	**74**	**160**	**234**	**215**	**19**	**253**	**100%**	**100%**

Note: *The total score points are based on one point for each dichotomously scored item and two points for each polytomously scored item.

Table 3.6. Main survey financial literacy item counts by framework categories

			Framework Recommendation
Process	**Number**	**%**	**%**
Identify financial information	7	15%	15-25%
Analyse information in a financial context	14	30%	15-25%
Evaluate financial issues	15	33%	25-35%
Apply financial knowledge and understanding	10	22%	25-35%
Total	**46**	**100%**	**100%**
Content	**Number**	**%**	**%**
Money and transactions	11	24%	30-40%
Planning and managing finances	16	35%	25-35%
Risk and reward	12	26%	15-25%
Financial landscape	7	15%	10-20%
Total	**46**	**100%**	**100%**

Table 3.7. Item counts in the field trial and main survey by domain and delivery mode

Domain	Field trial		Main survey	
	Paper-based (Design 1 / Design 2)	Computer-based	Paper-based (Design 1 / Design 2)	Computer-based
Reading	(87 / 66)	197 (+ 65 fluency items)	(87 / 66)	197 (+ 65 fluency items)
Mathematics	(71 / 63)	264	(71 / 63)	234
Science	(85 / 66)	115	(85 / 66)	115
Creative thinking	NA	38	NA	36
Financial literacy	NA	48	NA	46