PISA 2025 CORE 4: Summary of tasks and project timeline

Learning in the Digital World assessment

Contents

This document has been prepared as a supporting document for the Core 1 Call for Tender for PISA 2025. It summarises the tasks and project timeline of Core 4, as specified in the Core 4 Call for Tender for PISA 2025, which has concluded. Unless otherwise specified, in this document "The Contractor" refers to the Core 4 Contractor.

Guiding principles of the assessment development and data analysis

1. The Core 4 Contractor shall develop the instrument for the PISA Learning in the Digital World innovative domain assessment. In doing so, the Contractor shall:

I. Guarantee alignment between the assessment instruments and the assessment framework through close collaboration with the OECD Secretariat and the Expert Group;

II. Develop instruments that measure both students' achievement on the task (i.e. proficiency in selected learning practices), and the strategies they follow to complete the task, exploiting the evidence that can be gathered in the form of process or log data;

III. Ensure that the assessment environment has advanced interactive features, for example providing intelligent feedback to the students on the progress or results of their work;

IV. Ensure the test items and associated measurement tools are developed in a software that is compatible with the PISA platform;

V. Guarantee the assessment instrument is valid, fair and accessible across cultures, languages and countries

VI. Develop internationally valid scoring procedures and supporting materials.

2. The Core 4 Contractor shall also undertake the data analysis related to the Learning in the Digital World assessment (analysis of all item-level data in the Field Trial and Main Survey). The successful Core 4 Contractor shall develop an analysis plan for the analysis of the Field Trial and Main Survey data and deliver the results via a database to the OECD following an agreed timeline.

Introduction to the PISA Innovative Domain: Management, stakeholders and expectations

3. The purpose of the PISA innovative domain assessments is to provide insights into how important skills and competences beyond reading, mathematics and science (e.g. creative thinking) are being developed within education systems. The innovative domain assessments also represent an opportunity to pilot new approaches to large-scale assessment.

4. The assessment will be co-developed by two entities:

- The OECD Secretariat, in close collaboration with the Learning in the Digital World Expert Group (hereafter Expert Group), will develop the assessment framework. The OECD Secretariat will also be responsible for creating and managing the Expert Group, including the organisation of its meetings.

- The Contractor, in consultation with the Expert Group and the OECD Secretariat, and taking into account feedback from participating countries/economies, will design, develop and validate the assessment instruments, as well as develop the associated scoring materials. The Contractor will also be responsible for the analysis of the pilot studies, Field Trial and Main Survey data.

5. The PISA 2025 innovative domain assessment, *Learning in the Digital World*, aims to include several innovations to the task design and tools provided to the students to learn on the task, as well as innovations to how the assessment constructs will be measured:

- Gathering information on how students self-regulate when solving problems using digital tools;

- Developing and incorporating open-ended tasks that provide learning opportunities for the students, in the form of worked examples and automated feedback;

- Developing analytical approaches to derive a measure of 'learning on the task' that account for initial levels of knowledge/ability of the students in each country.

6. The role of the Expert Group is to advise the OECD Secretariat and the Contractor on the development and validation of the assessment framework and items, the scoring materials and the data analysis. This will include regular reviews of the tasks that will be designed by the Contractor, in addition to their plans for measurement and data analysis. The Expert Group is currently constituted by six experts in the fields of science, computer science, learning science, education technology and psychometrics, with the possibility to include one or two additional experts (see Annex C of this document for the composition of this Expert Group). The Expert Group will support the work of the Core 4 Contractor on the development of the instrument, and in particular, the work related to defining and analysing the measures of self-regulation in alignment with the competency model. The members of the Expert Group will be contractually engaged by the OECD for the duration of this project.

7. Given the Expert Group's central role in the development of the PISA 2025 *Learning in the Digital World* assessment, the Core 4 Contractor is expected to fully consider and implement the suggestions of the Expert Group throughout the process.

8. The assessment development process will also require the Core 4 Contractor to collaborate with other key stakeholders in the PISA project, in particular the Core 1 Contractor. The *Learning in the Digital World* assessment will be implemented within the overall survey design of PISA developed by the Core 1 Contractor. Core 1 will also conduct a considerable part of the day-to-day oversight and management of the project, including:

- Facilitating the implementation of the project management approach agreed to by all PISA International Contractors;

- Co-ordinating the work of the PISA International Contractors by developing and maintaining an integrated project plan and timeline;

- Ensuring the OECD Secretariat is kept fully informed on the progress of the project by organising progress meetings between representatives of all PISA International Contractors and providing quarterly reports.

9. While the Core 1 Contractor will have a considerable oversight and management role, the OECD Secretariat remains ultimately responsible for the international management of PISA 2025 and for facilitating collaboration between all of the PISA International Contractors.

10. In developing the assessment instruments for a computer-based assessment of *Learning in the Digital World*, the Core 4 Contractor shall be required to:

- Identify, in close collaboration with the Expert Group, a feasible and meaningful approach to operationalising an international assessment of Learning in the Digital World, following the indications provided in the assessment framework;

- Develop assessment instruments that reflect this approach and that provide information on the cognitive practices that students can master, as well as information about their self-regulation processes when working through complex tasks. The assessment instruments should deliver comparable international data, in the form of unidimensional, multidimensional, or multiple scales, on the relevant cognitive skills identified in the assessment framework, using cost-efficient and cross-culturally valid methods. The integration of measurement tools in the design of the assessment that can allow profiling students' self-regulation and measuring learning on the task is intended to advance PISA's capacity to measure thinking processes and engagement in the task.

The Contractor shall work closely with the Expert Group to define and develop these methods for analysing and reporting on measures of self-regulation and learning on the task.

- Liaise with the Core 1 Contractor to ensure that the test items developed can be rendered in the assessment delivery platform and are able to collect all the relevant data, including process data (see Annex B for a description of example prototype tasks that will be similar to those administered in the PISA 2025 innovative domain test). Work with the Core 1 Contractor to integrate the entire assessment into the PISA computer platform for the Field Trial and Main Survey;

- Develop test items and conduct small-scale validation exercises, including two rounds of cognitive laboratories and two pilot studies, respectively;

- Incorporate inputs from the PISA participating countries, that will be invited to review the results of the pilot studies and to review all of the developed instruments;

- Collaborate productively with the Core 1 Contractor during the entire development process, given the responsibility of the Core 1 Contractor for the overall survey design, platform development, test delivery and project management for PISA 2025;

- Coordinate the delivery of all instruments to the Core 1 Contractor in a timely manner, as agreed, and support the linguistic verification team, countries and other PISA International Contractors on content-related questions during the translation and verification process of the *Learning in the Digital World* assessment.

Instrument development process, test design, data analysis and responsibilities of the Contractor

11. This section describes the responsibilities broken down by task, for the instrument development, test design, data analysis and project management aspects of the PISA 2025 Learning in the Digital World assessment that will be carried out by the Core 4 Contractor. These responsibilities will span the contract period from 2020-2026.

12. The description and deliverables for each of the eight (8) Tasks for which the Core 4 Contractor is responsible can be found below. The Core 4 Contractor will hereafter be referred to as ‘the Contractor’. The eight (8) tasks are as follows:

- Task 1: Conceptual framework review
- Task 2: Instrument development
- Task 3: Authoring and platform compatibility
- Task 4: Cognitive laboratories and pilot studies
- Task 5: Test design plan
- Task 6: Item scoring procedures
- Task 7: Data analysis
- Task 8: Project management

Task 1. Conceptual framework review

The framework for the PISA 2025 Learning in the Digital World assessment provides the conceptual underpinning of the assessment. The Contractor shall use the framework to identify a feasible and meaningful way to operationalise an international assessment of Learning in the Digital World and to inform the design and development of the assessment instruments.

The development of the framework is managed by the OECD Secretariat, and each draft version of the framework will build upon the work of the Expert Group, as well as the feedback and input received from participating countries. The Contractor will also review and provide feedback on the first full draft version of the conceptual framework.

The OECD Secretariat will provide the Contractor with a more advanced working draft of the framework concepts in September 2020. The Contractor will provide a review and written feedback on the first draft of the framework by May 2022. A second revised draft of the conceptual framework will be presented to the PGB in 2022.

Task 2. Instrument development

2a. Design and development of test items and scoring procedures

The Contractor is responsible for developing the test items and associated scoring materials for the Learning in the Digital World assessment. Student performance will be measured through different types of computer-based items that can be either human- or machine-scored. Given the challenges and costs of ensuring reliable and comparable human scoring in a large-scale, international assessment, automated scoring methods are recommended whenever feasible.

The assessment will require innovative item design and scoring solutions that can capture information about students': (1) capacity to engage in the relevant cognitive practices identified in the framework; and (2) self-regulated learning processes when working through complex tasks. The Contractor is responsible for conceptualising and designing the items and the affordances of the assessment environment, as well as defining the evidence rules and item scoring methods.

Given the complex, open-ended nature of tasks, each assessment unit must be based on an evidence model describing how relevant student responses and interactions within the task will be parsed into observable indicators useful for analysis.

The test units will have an approximate duration of 30 minutes. The Contractor shall develop nine (9) learning units, as part of the development of the initial pool of units. Eight (8) units (four hours of material) will be administered in the Field Trial, and the six (6) best performing units (three hours of material) will be retained for administration in the PISA 2025 Main Survey. The Contractor will also develop a method to measure levels of initial knowledge/ability, either through discrete items in the first cluster of items administered to each student or through items integrated in each unit. The Contractor will also develop tutorial(s), demonstrating the affordances of the environment to students and providing them with acclimation time, either across or within units. Each unit will provide resources for students to learn on the task, either as help/hint functions that students can query or as worked examples.

2b. Verification of platform compatibility

The Contractor is responsible for authoring one full unit into the PISA platform by July 2021, with the support of the Core 1 Contractor, to ensure its full integration and functioning in the platform. The Contractor is required to work closely with the Core 1 Contractor throughout the development process to test and validate that all items, especially new item-types introduced in this assessment, and the collection and use of process data, can be supported and delivered by the chosen PISA computer platform.

---

[1] This timeline can be adjusted in case of delays in the finalisation of the contract with the Core 1 Contractor.

In addition to verifying the compatibility of the items with the PISA platform, the Contractor is responsible for developing, validating and testing the interactive components of the tasks (e.g. automated feedback on correctness of the response), and the scoring methods of all items, including those requiring the extraction, parsing and coding of process data. The Core 1 Contractor will provide advice and support to the Contractor for this process. This support from Core 1 might include the implementation of add-ons to the platform that are functional to the integration of the instruments for the assessment of Learning in the Digital World. The terms of reference in the Core 1 Call for Tender require the Core 1 Contractor to dedicate sufficient resources to support the incorporation of more complex assessment tasks authored by other parties in the PISA platform, including the provision of a simplified authoring tool, or the provision of an appropriate alternative. This includes making developments in the platform that allow more interactive and data-heavy assessment content.

2c. Validation activities

Given the international scope of the assessment, it is important that the test items are cross-culturally, cross-linguistically and cross-nationally valid. The Contractor is expected to address these validity issues during the development of the items, as well as through various small-scale validation exercises (e.g. cognitive laboratories, pilot studies) and statistical analyses of the data following the Field Trial and Main Survey.

2d. Preparation of pool of units for country review

Participating countries/economies in PISA have the opportunity to review all new items developed for PISA. This review process is an important mechanism through which to ensure that the test items are internationally valid, culturally relevant and appropriate, and representative of the domain framework constructs. The Contractor must ensure that the initial pool of units can be packaged in a suitable format for country review, in coordination with the Core 1 Contractor. The Contractor will submit the initial pool of units to the Core 1 Contractor by November 2022. The Core 1 Contractor will be responsible for sharing the items for review with the PISA participating countries/economies, and for receiving and sharing the feedback with the Contractor.

2e. Translatability assessment by Core 1

The Core 1 Contractor will conduct a translatability assessment to ensure that the test items developed by the Core 4 Contractor can be successfully translated into all languages included in the PISA 2025 assessment. The Core 4 Contractor will liaise with Core 1 Contractor to ensure proposed linguistic changes resulting from the translatability assessment do not influence the desired meaning or purpose of items.

2f. Finalisation of pool of units for Field Trial and Main Survey

The Contractor will oversee the finalisation of the pool of units for administration in the Field Trial by reviewing the responses of countries during the country review of the initial pool of units. The Contractor will also oversee the finalisation of the pool of units for administration in the PISA 2025 Main Survey by reviewing the outcomes of the Field Trial, proposing the most suitable items for inclusion in the Main Survey, and developing the achievement level descriptors that describe students' capabilities at different achievement levels of the thinking practices. These tasks will be carried out in close collaboration with the Expert Group and the OECD Secretariat.

Task 3. Authoring and platform compatibility

3a. Unit authoring process

The Contractor is responsible for all of the front-end authoring work for the 2025 Learning in the Digital World assessment. Front-end operations include: (1) the initial design of the content and structure of the unit; (2) the programming of the graphic and textual content displayed on screen and of the available interactions using HTML, CSS and JavaScript; and (3) the programming of the automated feedback and scoring rules for each item.

The Core 1 Contractor is responsible for the back-end operations of the authoring process, which include: (1) packaging the units produced by the Contractor into test clusters; (2) ensuring that all the content included in the unit can be processed for translation; and (3) verifying the consistency of the layout and of the user experience. The Core 1 Contractor will provide the Contractor with detailed specifications on the conventions and guidelines that need to be followed during the front-end authoring process to ensure that all units can be packaged for delivery in the PISA platform, translated, and provide a consistent test-taking experience to students.

3b. Testing the integration of units within the PISA platform
On the basis of the specifications provided by the Core 1 Contractor, the Contractor will produce one sample unit and work with the Core 1 Contractor to test the integration of the unit within the PISA platform by July 2021. The Call for Tender for the Core 1 Contractor includes an obligation to support the engineers/programmers of Core 4 during this testing process. This process is expected to provide the Contractor with all of the necessary information to undertake the front-end programming of all of the units.

The main tests required to validate the compatibility of the unit include, but are not limited to:

- Support platform integration: verifying the unit works as expected in the PISA platform;
- Support scoring process: verifying the unit collects the answers, builds the score and sends it to the platform (database encryption);
- Support log information;
- Support basic translation: verifying the unit can handle the translation process;
- Support long translation: verifying the unit can handle long text and maintain a structured display;
- Support right-to-left languages: verifying the unit can display translations, such as Hebrew and Arabic translations.

3c. Content protection
The content of the unit must be protected. It will not be possible to expose the item content to non-authorised individuals.

3d. Delivery of units to OECD

Given that several or all the units will be released to the public after the administration of the test and the publication of an international report, the Contractor will deliver the units in a format (e.g. HTML and JavaScript) that can be easily embedded within the PISA website and/or other platforms using web technology, so that users can visualise and interact with the material.

Task 4. Cognitive laboratories and pilot studies

To inform item development, the Contractor shall conduct initial validation studies, including two rounds of cognitive laboratories and two pilot studies, between 2021 and 2022.

4a. Cognitive laboratories
Cognitive laboratories will be conducted in two rounds (Round 1 in one country by September 2021; Round 2 in three countries by June 2022). The Contractor will be responsible for selecting one country per round. For the second round, the OECD Secretariat will select and manage the other two countries that will voluntarily undertake the work related to these validation studies. These studies should gauge the extent to which the assessment units address the intended cognitive and self-regulatory processes as defined in the framework. It is therefore important that participants take the assessment units on a fully functional platform (i.e. PISA platform) that allows for the collection of response and process data. Given they will be conducted in multiple countries, the second round of cognitive laboratories should also provide some evidence about the functioning of the tasks in different cultures.

Cognitive laboratories will inform plans for developing indicators based on process data in that they will provide information about the extent to which students undertake the assessment tasks as expected by the assessment designers and the Expert Group.

The Contractor is responsible for conducting the cognitive laboratories in one country in each round; the OECD Secretariat will be responsible for the logistics and organisation of the cognitive laboratories in the two additional countries during the second round. The cognitive laboratories and pilot studies will likely be conducted from the same grouping of countries. Either the participating countries or the OECD Secretariat will transcribe and translate into English (if necessary) transcripts and students' responses from the cognitive laboratories, and deliver these to the Contractor. The Contractor will be responsible for analysing the data, in collaboration with the Expert Group, and providing the OECD Secretariat with a short report (for each cognitive laboratory) on their findings based on the analyses, including how assessment units will be modified based on the findings. To ensure cohesion between the experiences of participating students between rounds and among countries, the Contractor will produce a cognitive laboratory protocol document before the second round of cognitive laboratories and will conduct a training session in English to guide those responsible for administering the cognitive laboratories in the additional two countries in 2022.

4b. Pilot studies

The Contractor will conduct two pilot tests before the finalisation of the initial pool of units for the Field Trial. The Contractor is responsible for developing a fully-functional version of the units in one or more test forms for the pilot studies, packaged in a software of their choosing (i.e. for either online or offline delivery), and for providing instructions to countries for administering the test forms to students. By November 2021, the Contractor shall conduct the first pilot study, administering two (2) fully functional assessment units to students in one country. By August 2022, the Contractor shall conduct a second pilot study, administering fully functional versions of all nine (9) units developed for the assessment in two countries. The Core 1 Contractor will provide support for the integration of units into test forms that can be run on the PISA platform, and will provide the Contractor with a dataset after the data collection. In both rounds, data for 200 students for each item will be collected.

The Contractor is responsible for analysing the data collected during the pilot tests and producing a report of the findings after each round that specifies how assessment units will be modified on the basis on the findings. These reports will be shared with the OECD Secretariat, the Expert Group and the PISA Governing Board. Pilot test data will provide quantitative evidence essential for refining the tasks, parsing process data, and developing indicators. The pilot studies will also verify the technical and operational functioning of the test items. The Contractor should plan to report on the frequency of scores on each indicator (for response items and indicators based on process data), the extent to which students progressed through tasks in the predicted manner, and should identify any unexpected patterns that emerge from the data. In addition, indicator parameters (difficulty and discrimination) should be estimated and misfitting indicators should be identified.

The OECD Secretariat will be responsible for all of the logistics relating to the second round of the pilot studies in one additional country, including identifying and liaising with the participating country.

Task 5. Test design plan

The PISA 2025 Main Survey assessment has a duration of two hours. It has a rotational design, meaning that different students take a different combination of domains (e.g. one hour reading and one hour science, or one hour math and one hour innovative domain, etc...). This effectively means that some students may take the innovative domain test for two hours, whereas others may take it for only one hour (in past designs, the maximum length per student was one hour per domain so that the design would support the identification of pairwise inter-correlations among all domains; this requirement could be relaxed in the case of Learning in the Digital World). A set of discrete test items that gauge the relevant knowledge and skills that students bring to the assessment (level of initial knowledge/ability) might be administered at the outset of testing. In addition, a small number of task-specific items on initial knowledge/ability might also be included at the beginning of each unit.

Testing units should include enough items and data points to ensure that the test as a whole adequately measures the construct as described in the framework, and can produce reliable scales. Dimensionality analysis will determine whether student performance on the cognitive practices identified in the framework is best summarised by a single scale, sub-scales, or a multi-dimensional scale.
It is also anticipated that responses to situated self-report items, indicators based on process data, and questionnaire items will be pooled across tasks to generate distinct indicators for different components of self-regulatory processes.

Prior to the implementation of the Main Survey in 2025, the Field Trial (in 2024) will test the functioning of the units in a large-scale, international scenario. The Field Trial is undertaken in all countries, with samples of about 2000 students per country, in which around 750 of those students per country will sit the innovative domain assessment. For the Field Trial of the *Learning in a Digital World* assessment, there should be four (4) hours of assessment material. Only the units included in the Field Trial can be implemented in the Main Survey.

The Contractor, seeking input from the Core 1 Contractor on test design, will submit an initial test design plan for the Field Trial to the OECD Secretariat for approval by June 2021. The test design plan will specify how many units will be developed and in which format for the country review, and how many units will be included in the Field Trial and in the Main Survey. The test design plan should also indicate the expected duration of each unit, the number of items included to gauge initial levels of knowledge/ability, and provide a mapping of the items and process-data indicators to the components of the framework that demonstrate adequate coverage. An update to the test design plan should be submitted to the OECD Secretariat after the analysis of Field Trial data, by October 2024.

The Main Survey will include a minimum of six (6) units (3 hours of assessment material, if the average length of each unit is 30 minutes) for the Learning in the Digital World assessment from the pool of units tested in the Field Trial. Revisions of Field Trial units that affect translations or would require further validation should be avoided as much as possible at this stage.

It is possible that some units will need to be modified, redesigned or dismissed altogether as a result of the feedback provided by countries during the review process. It is therefore recommended that the Contractor plan to develop a larger number of units for the country review than is planned for administration during the Field Trial.

Task 6. Item scoring procedures

The international comparability of item scoring is paramount in the PISA assessment. The Contractor is responsible for developing rigorous, internationally comparable scoring procedures for the assessment. In this case, scoring procedures will include methods for parsing log data and transforming them into observable indicators.

The Contractor will need to produce and programme scoring rules, including specifications or code for parsing raw data and generating indicators from these parsed data, and revise them on the basis of data analyses. These scoring rules should be described in the documentation that is submitted to the OECD and to the Expert Group every three months for review, so that they can provide feedback. Explanations of scoring rules should also be provided in the documentation for the initial pool of units that is distributed by the Core 1 Contractor to the countries for review.

Task 7. Data analysis

7a. Analysis plan
The Contractor shall prepare an analysis plan, in collaboration with the OECD Secretariat and the Expert Group and upon guidance from representatives of the PISA Governing Board (particularly the Analysis and Dissemination Group). When developing the analysis plan, the Contractor shall follow the timeline established by the Core 1 Contractor. The Contractor shall submit an initial analysis plan by February 2021 for the PISA Governing Board to review and approve at its meeting in the first quarter of 2021.

The initial analysis plan should outline the analyses that will be possible for the Field Trial and Main Survey of the Learning in the Digital World assessment following the below guidelines, in addition to how findings from the second rounds of the cognitive laboratories and pilot studies will be used to finalise the selection and operationalisation of the measures. The plan will include information about the construction of scales and indices, plans for parsing and analysing log file data, and a discussion of the analytical and psychometric techniques that:

- Reflect state of the art approaches to generating scores from embedded assessments;
- Address potential multi-dimensionality and dependence in item scores;
- Focus on ensuring valid, reliable country-level indicators generated from data collected from incomplete test designs (such as matrix sampling or adaptive test designs); and
- Make best use of the available data while ensuring that the analyses remain accessible to policy makers and researchers.

The analysis plan should also advise on the limitations of, and possible ways of controlling for, any lack of full equivalence of the items across cultures and languages.

The analysis plan will define the indicators to be built and included in the database and will be updated after the analysis of the Field Trial data. The plan will also indicate how the members of the Expert Group will contribute to the analysis. Once approved by the PISA Governing Board, the analysis plan will guide the OECD Secretariat in preparing and designing the reporting plan for the Learning in the Digital World assessment. The kinds of scores that should be generated and specified in the analysis and reporting plan are described next.

Within the analysis plan, the Contractor should include specific sections that describe their methods and plans for developing the following elements, which are described in further detail below:

- Proficiency scales
- Achievement levels
- Indicators of self-regulation
- Measure of learning in the task

Secondary analysis of the data might produce other indicators for reporting, such as an indicator on students' basic digital literacy or more nuanced reporting on self-regulated learning processes. Such additional analysis will be conducted by the Expert Group and the OECD Secretariat.

Proficiency scales
The Contractor will need to build proficiency scales. The analysis plan should address the generation of student-level plausible values by combining process data and traditional item responses, and aggregation of plausible values into country-level scale score means and standard errors.

Achievement levels
Achievement levels, which characterise overall levels of performance associated with scale score bands, will also be developed by the Contractor. Achievement levels will be reported in terms of the percentage of students at each performance level, by country and overall. The Expert Group will be heavily involved in writing achievement level descriptors, but the Contractor will conduct analyses to set the cut scores for each achievement level. The analysis plan should thus describe the processes used to develop achievement level descriptors.

Indicators of self-regulation

Indicators of students' capacities to self-regulate their learning processes will be developed through self-report items (situated in the task and in the background questionnaire) and through the analysis of process data. The analysis plan should specify which (and how many) indicators will be used to gauge each construct. For situated self-report indicators, the approach to embedding these questions in the task should also be specified with sufficient justification (for example, how frequently these questions will be asked and where they will be placed—at the beginning, at the end, or in the middle of the task). The Expert Group will support the Contractor in this design. The Contractor will be responsible for analysing and reporting measures developed from self-report data.

It is expected that the methods used for developing indicators of self-regulation based on process data might not yield measures that have the same level of reliability and international comparability of traditional PISA scales. The Contractor should design the units so that the methods established with the Expert group can be implemented, and can collect and parse the relevant process data so that the Experts can conduct the analyses. This aspect of the work, while very important, is considered to be more experimental than the rest of the measurement work. The research and analysis for producing these indicators based on process data will be conducted by the Expert Group and OECD Secretariat: the main responsibility of the Contractor is to incorporate the interactive elements that allow the collection of relevant process data in the units, and to parse these process data in the database according to the indication provided by the Expert Group.

Measure of learning on the task

Test performance reflects both the level of student preparation as of the testing date (initial ability) and student effectiveness in using digital learning resources to learn. Given the importance of distinguishing the reporting of these two drivers of performance in this assessment, the analysis plan should outline a method for generating indicators/scales of initial ability. As with the measures of regulation, the Contractor is responsible for incorporating in the design of the test those elements (e.g., additional discrete items) that would make it possible to implement the methodology defined by the Expert Group.

7b. Data preparation, analysis and handling

The Core 1 Contractor will provide to the Contractor all of the data collected during the international administration (both for the Field Trial and the Main Survey) following the specifications provided by the Contractor and agreed on by the Core 1 Contractor. The Contractor will then produce and provide to the OECD an international database, including all derived variables, scales and indices for the Learning in the Digital World assessment. The international database will also include a complete set of log-file data for secondary analysis, both in a raw data format and cleaned and organised in the format specified in the analysis plan.

The international database shall include the following types of electronic files:

- Two distinct and widely adopted data formats for statistical software (e.g. SAS, STATA, SPSS) as well as additionally flat file format such as ASCI or CSV, including weights. The final file types to be provided to the OECD will be determined after discussions with the OECD. Where applicable, each file will contain appropriate linking information supported by linking instructions.
- An electronic codebook for each file. This file contains a brief listing of the name, position, format, and description of each data variable, and an expanded listing for the discrete variables that have coded information, including their code values, descriptors and (unweighted) frequencies within each data set.
- A user guide document file in PDF format. This document will contain information about the contents and structure of the data file, information about how to use the other resource files in the product, and enough technical information on how to conduct analyses of the data, including the names and usages of the key variables in each data set.

Following the Field Trial data analysis, the Contractor shall work with the Expert Group, and the OECD Secretariat to develop and submit a Field Trial analysis report by October 2024. The Contractor will then submit, in collaboration with the Expert Group, the proposed Main Survey unit selection to the OECD Secretariat by October 2024.

Following the Main Survey data collection, the Contractor will be responsible for processing the clean data that they receive from the Core 1 Contractor (as agreed by the Contractor and the Core 1 Contractor), and producing a scale (single, multidimensional or with sub-scales) of student performance on the learning tasks. The Contractor will also proceed to build additional indicators (of learning on the task, self-regulatory processes) that are included in the analysis plan, collaborating as much as possible with the Expert Group.

Once the Contractor receives the data collected by the Core 1 Contractor for the Learning in the Digital World assessment and the Contractor performs the necessary analysis of these data (scaling and indicator production), the Contractor will send the national data files to the OECD Secretariat and the Core 1 Contractor on an agreed upon timeline. It is the responsibility of the Core 1 Contractor to then send the national data files to the participating countries.

Achievement level descriptors will need to be developed by the Contractor, in close collaboration with the Expert Group, by February 2026.

The Contractor shall provide a fully documented database to the OECD and the Core 1 Contractor in interim form by the end of April 2026, and in a second more complete form in June 2026. The final database, incorporating all amendments or corrections, shall be completed no later than the end of October 2026. In addition, an earlier initial dataset shall be compiled and delivered to the OECD Secretariat in January containing data that have been processed by that time, to allow the OECD and the Expert Group to carry out initial data exploration. These preliminary dates will be confirmed when a timeline is developed by the international Contractors and agreed to by the OECD.

All international Contractors, including the Contractor, shall contribute to the sections or chapters relevant for the methodology of their work in the PISA 2025 Technical Report. Submissions related to Core 4, drafted by the Contractor, should be submitted to the Core 1 Contractor by March 2026, according to the agreed timeline.

Task 8. Project management

8a. Collaboration with Core 1

The OECD Secretariat is responsible for the management of the Expert Group, the PISA Governing Board meetings and, ultimately, the international management of the PISA project. However, the Core 1 Contractor shall assume a considerable portion of the day-to-day management of the project and the coordination of the PISA International Contractors. The Contractor and the Core 1 Contractor are therefore required to propose and manage an 'integration and management plan' for the integration of the innovative domain instrument development within the general survey and implementation operations of PISA 2025. This integration and management plan will be submitted to the OECD Secretariat by May 2021 and should establish an agreed timeline to co-ordinate the development processes of the *Learning in the Digital World* assessment with the respective processes and timelines of the other relevant PISA Contractors[2].

[2] See http://www.oecd.org/pisa/pisaproducts/PISA-2018-INTEGRATED-DESIGN.pdf for an example of the PISA survey design.

The Contractor and the Core 1 Contractor are also required to collaborate closely on ensuring all test items are successfully authored and integrated into the chosen PISA platform at various points in the development process. The OECD Secretariat will remain responsible for resolving all contractual issues and will arbitrate in the event of any disagreements between Contractors.

The Contractor will submit progress updates every three months to the Core 1 Contractor and the OECD Secretariat. In addition, the Contractor will attend a monthly Contractor call between all of the PISA International Contractors and the OECD Secretariat. The Core 1 Contractor may also organise additional calls, as necessary and agreed.

All International Contractors will be expected to contribute to the preparation of the consolidated timeline for PISA 2025, of which the Core 1 Contractor is responsible.

8b. Staff Requirements
To facilitate communication between respective entities, the Contractor's Project Lead will have two main responsibilities:

- Coordinating the assessment development work with the relevant teams in their organisation. The majority of the Project Lead's time will be allocated to content-related work for developing the assessment;
- Liaising with the International Survey Director (Core 1), other Core International Contractors when relevant, and the OECD Secretariat on work related to Core 4.

8c. Participation in meetings with the Expert Group and the International Contractors
There are two distinct types of Expert Group meetings: the first, a face-to-face Expert Group meeting, where all members of the Expert Group and the OECD will meet in person, that will occur five (5) times throughout the cycle and span over two days or two and a half days; and the second, a virtual Expert Group update meeting that will last between four (4) to eight (8) hours over the span of two days and be scheduled on an as-needed basis. At least one member of the Contractor team shall attend the meetings of the Expert Group. The Contractor will physically attend at least one of the five (5) face-to-face meetings of the Expert Group across the entire PISA 2025 cycle (2020-2026).

At least one member of the Contractor team shall also attend the International Contractor meetings (normally two per PISA cycle), physically attending at least one. The Contractor is expected to cover all costs relating to the attendance of its necessary personnel to the above-mentioned meetings.

8d. Collaboration with the Expert Group
The Expert Group will be closely involved in processes related to the assessment development carried out by the Contractor to ensure the assessment framework is reliably and validly measured by the test items (e.g. use of process data and design of analysis and reporting). In particular, the Expert Group will lead the research on the development and validation of measures of self-regulation (those based on process data) and of learning on the task. The Contractor will participate in several virtual meetings with the Expert Group, moderated by the OECD Secretariat, to facilitate the item development process. The first virtual meeting between the Expert Group, the OECD Secretariat and the Contractor will take place in September 2020 to discuss the assessment framework and next steps.

8e. Support translation process
After the finalisation of the initial pool of units, the Contractor shall provide all of the assessment instruments and scoring guides in one of the source languages (English or French) to the Core 1 Contractor, as per the agreed schedule. During the translation verification processes, the Contractor shall be available to liaise with the linguistic verification team of the Core 1 Contractor in order to ensure that any proposed linguistic changes do not influence the desired meaning or purpose of the units.

8f. Coordination with Core 1 on timeline

The overall co-ordination and smooth running of the Contractors' schedules shall be the responsibility of the Core 1 Contractor, and it is expected that the Contractor shall co-operate accordingly to ensure that all agreed timelines are adhered to. However, given the iterative instrument design process, all Contractors will also need to demonstrate flexibility in order to respond to the feedback and requests of the PISA Governing Board and the OECD Secretariat. In particular, the Contractor must be prepared to consider and, where appropriate or deemed necessary, undertake revisions to the instruments in response to requests from the PGB or to unforeseen issues that may arise during the instrument development process.

ANNEX A: Estimated schedule of activities

#	From	To	Activity/Milestone
1		01-Oct-20	Contractor to take receipt of working draft of the PISA 2025 *Learning in the Digital World* Conceptual Framework.
2		Feb-21	Contractor to prepare and submit an initial data analysis plan to the OECD Secretariat and Expert Group, outlining the analyses that will be possible with the data collected in the cognitive labs, pilot studies, Field Trial and Main Survey of the PISA 2025 *Learning in the Digital World* assessment.
3		May-21	Contractor to propose, in partnership with the Core 1 Contractor, an integration and management plan detailing the integration of the Contractors' workflow and processes with the overall survey management and implementation operations of PISA 2025, including an agreed timeline for development and integration processes.
4		Jun-21	Contractor to contribute to the consolidated timeline for PISA 2025 drafted by the Core 1 Contractor.
5		Jun-21	Contractor to submit initial test design plan for Field Trial to the OECD Secretariat for approval, seeking input from the Core 1 Contractor on the test design.
6		Jul-21	Contractor to deliver one fully functional test unit to the OECD Secretariat, with documentation describing the evidence model and scoring rules for each task.
7		Jul-21	Contractor to collaborate with Core 1 Contractor to comprehensively test the integration of the first fully functional test unit in the PISA platform.
8		Sep-21	Contractor to conduct first round of cognitive laboratories in Australia (2 fully functioning test units) in the PISA platform.
9		Oct-21	Deliver test units to Core 1 for preparation of test forms in the platform for the first round of pilot studies.
10		Nov-21	Contractor to administer first round of pilot testing (2 fully functional test units) in Australia on the PISA platform.
11	Oct-21	Dec-21	Contractor to analyse the transcripts from the first round of cognitive laboratories and submit a report of the results to the OECD Secretariat.
12		Dec-21	Contractor to take receipt of first full draft of the *Learning in the Digital World* Conceptual Framework.
13	Dec-21	Feb-22	Contractor to analyse the results of the first round of pilot studies and submit a report of the results to the OECD Secretariat.
14	Mar-22	May-22	Contractor to review and provide feedback on the first full draft of the Conceptual Framework, based on findings from the first round of cognitive laboratories and pilot studies.
15		May-22	Contractor to submit a cognitive laboratory protocol document (in English) and conduct training sessions to support administrators of the second round of cognitive laboratories.

#	From	To	Activity/Milestone
16		Jun-22	Contractor to conduct second round of cognitive laboratories (all nine units) in three countries (Contractor responsible for one country).
17		Jul-22	Contractor to submit a Test Administration Guide (in English) to OECD for the second round of pilot studies.
18		Jul-22	Deliver test units to Core 1 for preparation of test forms for the second round of pilot studies.
19		Aug-22	Contractor to administer second round of pilot testing (all nine units) in one country and support administration in one other country.
20	Jul-22	Sep-22	Contractor to analyse the transcripts from the second round of cognitive laboratories and submit a report of the results to the OECD Secretariat.
21		Nov-22	Contractor to submit Unit Model Book to the OECD Secretariat and Expert Group.
22	Sep-22	Nov-22	Contractor to take receipt of second full draft of the Learning in the Digital World Conceptual Framework.
23	Sep-22	Nov-22	Contractor to analyse the results of the second round of pilot studies and submit a report of the results to the OECD Secretariat.
24		Nov-22	Contractor to develop all nine (9) 30-minute test units with associated scoring rules and adequately package/deliver initial pool of units to Core 1 Contractor for country review and translatability assessment.
25	Dec-22	Feb-23	Contractor to liaise with Core 1 Contractor to ensure proposed linguistic changes resulting from the translatability assessment do not influence the desired meaning or purpose of items.
26		Mar-23	Contractor to select and finalise the pool of units for the Field Trial (8 units), and deliver the selected units to the Core 1 Contractor for back-end authoring/integration into the PISA platform.
27		Jun-24	Contractor to submit a Draft Unit Scoring Model Book to the OECD Secretariat and the Expert Group.
28		Jun-24	Contractor to submit a Draft Machine Scoring Guide to the OECD Secretariat and the Expert Group.
29	Mar-24	Sep-24	**Administration of Field Trial (data collection by National Centres coordinated by Core 1 Contractor)**
30	Jul-24	Oct-24	Contractor to analyse the data collected in the Field Trial according to the analysis plan and submit a Field Trial report to the OECD Secretariat, developed in collaboration with the Expert Group.
31		Oct-24	Contractor to update the initial data analysis plan, in coordination with the OECD Secretariat and the Expert Group, outlining the analyses that will be possible for the Main Survey
32		Oct-24	Contractor to provide the OECD Secretariat with the Field Trial dataset and a document (e.g. codebook) that describes the indicators built in the dataset for additional analysis.

#	From	To	Activity/Milestone
33		Oct-24	Contractor to revise test design plan for the Main Survey in collaboration with the Core 1 Contractor, and submit to the OECD Secretariat.
34		Oct-24	Contractor to submit, in collaboration with the Expert Group, the proposed Main Survey unit selection (6 units) to the OECD Secretariat.
35		Jan-25	Contractor to implement corrections needed in the units selected for the Main Survey and to conduct further tests with the Core 1 Contractor in preparation of the Main Survey administration in the PISA platform.
36		Jan-25	Contractor to submit the updated Unit Scoring Model Book to the OECD Secretariat.
37		Jan-25	Contractor to submit the updated Machine Scoring Guide to the OECD Secretariat and Expert Group.
38		May-25	Contractor to provide OECD with files for one unit in a format (e.g. HTML and JavaScript) that enables the reproduction of the units on websites.
39	Mar-25	Oct-25	**Administration of Main Survey (data collection by National Centres coordinated by Core 1 Contractor)**
40	Dec-25	Jan-26	Deliver an initial dataset to the OECD Secretariat containing data that have been processed by that time.
41		Feb-26	Contractor to develop achievement level descriptors in close collaboration with the Expert Group.
42		Mar-26	Contractor to contribute to the sections or chapters relevant for the methodology of their work in the PISA 2025 Technical Report.
43		TBD with Core 1	Contractor to deliver national data files, containing scales and indicators, to the OECD Secretariat and Core 1 Contractor.
44		Apr-26	Contractor to deliver the first preliminary batch of the documented Main Survey database to the OECD Secretariat, with accompanying documentation.
45		May-26	Contractor to provide the OECD Secretariat with files for all six (6) test units in a format (e.g. HTML and JavaScript) that enables the reproduction of the units on the OECD webpage.
46		Jun-26	Contractor to deliver the second preliminary batch of the documented Main Survey database to the OECD Secretariat, with accompanying documentation.
47	Upon reception of Main Survey data from Core 1	Aug-26	Contractor to conduct all analyses of the Main Survey data relating to the *Learning in the Digital World* assessment, in accordance with the analysis plan approved by the PGB.
48		Aug-26	Contractor to deliver the final batch of fully documented, Main Survey database to the OECD Secretariat and the Core 1 Contractor, with accompanying documentation.

#	From	To	Activity/Milestone
49		Until 30-Sep-26	Contractor to respond to queries sent by the OECD Secretariat or the Core 1 Contractor on the analysis and interpretation of the data.

ANNEX B: Example prototype tasks

13. This section includes two prototype tasks developed by the PISA 2025 Learning in the Digital World Expert Group that present possible ways of operationalising the measurement of the assessment construct. The Expert Group will continue to refine these examples over the following months. Bidders are invited to consider these prototype tasks as illustrative, and are welcome to present examples that significantly deviate from these models. Bidders should, however, follow a similar format for the description of the prototype task(s) that they will develop and present as part of their Technical Bid.

Example Task 1: Yum!

Overall Description

14. The purpose of this task is to evaluate the student's ability to engage in a process of self-regulated learning by modelling in a digital environment. The student will need to learn how to use a digital modelling and graphing tool by watching video tutorials and interacting with scaffolded learning tasks. During the model-building process, the student will establish relationships between variables and learn to use models to make predictions. This process is iterative and the student will receive immediate feedback on whether the model they have constructed is an accurate representation of the given issue (based on the data provided in the task).

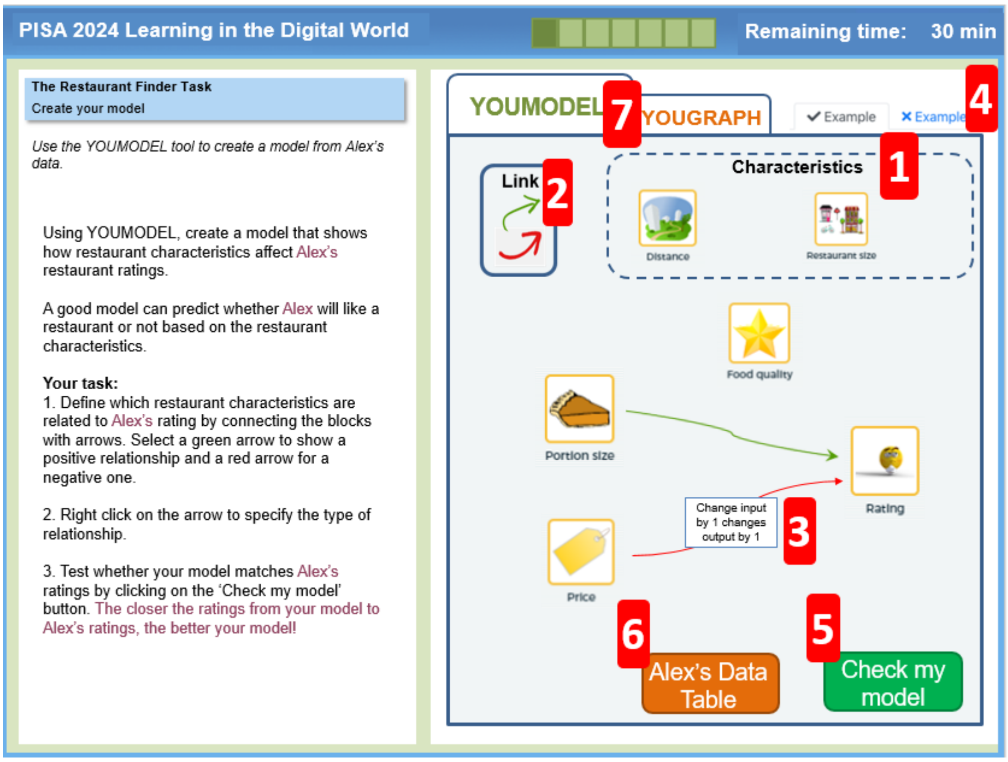
15. This task asks students to investigate the factors that determine whether or not someone likes a given restaurant, in the context of developing an application that suggests restaurants to people based on a rating system. This rating system can be modelled with respect to observable characteristics of the restaurants in order to predict and propose restaurants that fit people's criteria and to guide future dining choices. These types of applications are increasingly popular, and the modelling concepts and methods behind these applications also apply to many other contexts and problems both at the micro-level (e.g. consumer preferences) and macro-level (e.g. interactions between markets or natural phenomena).

16. Student performance on this task will be evaluated on the basis of the completeness and correctness of the model they construct. This learning context was selected because it is relatively accessible, although its cross-cultural relevance would need to be further investigated through validation activities (such activities were beyond the scope of the initial prototyping work of the PISA 2025 *Learning in the Digital World* Expert Group).

Key elements of the learning environment

17. The learning environment is structured as an executable concept map with a data exploration tool. The concept map design is based on existing digital tools: SageModeler (Concord Consortium) and Betty Brain (Vanderbilt University). Students will use the concept map to create a visual representation of their understanding of the conceptual (i.e. which variables affect each other) and quantitative (i.e. to what extent a unit change in one variable affects another variable) relationships between the given variables. The students acquire the information that they need to build their executable model by exploring relationships in the data exploration tool, which includes a graphing tool with limited functionalities.

Figure B1. Visual of the conceptual modelling environment (YOUMODEL tool) and navigation screen



18. Figure B1 shows a possible instantiation of the modelling environment and navigation screen. The environment is "open", meaning that students can go back and forth through the different phases and challenges of the unit. The key user interface elements are enumerated in Figure B1 and described in further detail below:

1. <u>List of variables:</u>
   Students can select the input variables (i.e. restaurant characteristics) to add to their model. Each variable is represented as a block with an image and title that they can drag and drop anywhere in the modelling space of the interface. In the model underlying the current prototype, all input variables but one are associated with the outcome (i.e. the rating).
2. <u>Linking variables:</u>
   By clicking on the 'Link' button, students can connect related variables through arrows. The direction of the arrow indicates which variable affects the other. Students can choose a green arrow to indicate a positive relationship, or a red arrow to indicate a negative relationship. Their choice of arrow colour makes their hypotheses about the relationship between two (or more) variables explicit.
3. <u>Labelling links:</u>
   Executable models require the specification of weights for each link between variables, indicating how a given variation in the input variable affects the output variable. These weights can be expressed through a formal (i.e. functions) or non-formal language (i.e. 'increasing input (price) by one point decreases the output (rating) by one point'). In this prototype, students are instructed to right click on each arrow to specify the functional relationship between the input and output variables. When they click on the arrow, a drop-down menu appears with a list of functional relationships expressed in non-formal language. There is one correct relationship for each pair of characteristics.
4. <u>Worked examples:</u>
   Students can access worked examples that illustrate a similar modelling task in a different context (for example, identifying the variables that affect air quality). They are provided with both a 'good' and a 'bad' example that are meant to complement the technical training on the interface that the students received during the tutorial. The 'good' example demonstrates one

possible approach to solving the modelling problem, while the ‘bad’ example demonstrates a typical mistake (for example, failing to verify whether the relationship changes at different levels of the input variable).

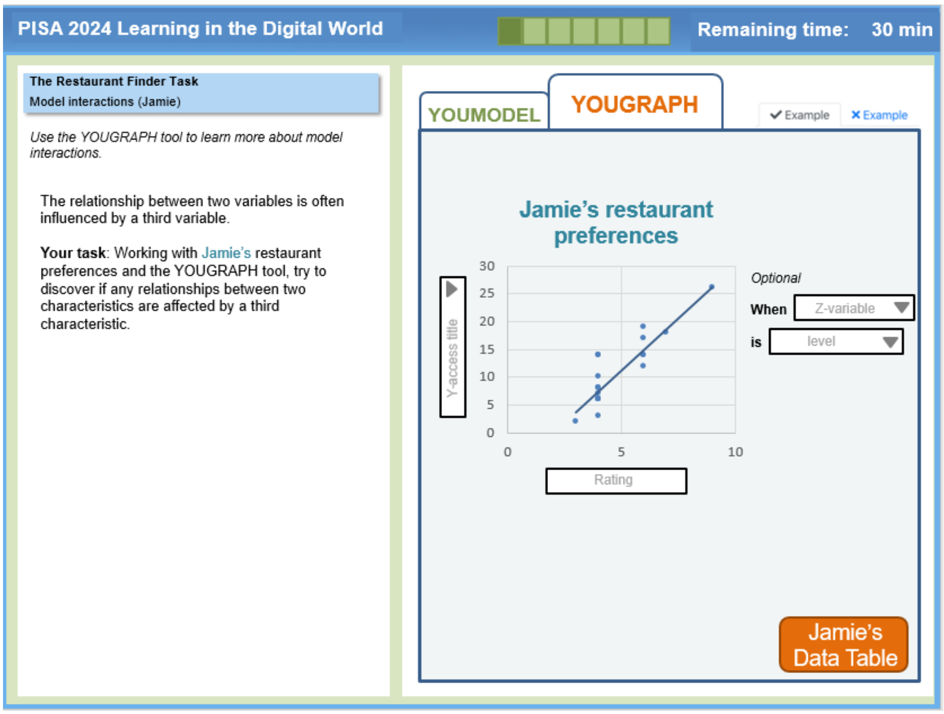
5. <u>Testing the model:</u>

   By clicking on the ‘Check my model’ button, students can check how well their model fits the provided data by examining the comparative results in a spreadsheet. Students are instructed that a good model is one that is capable of accurately predicting the preferences of their team members. We expect that students will iteratively update their concept map (e.g. adding links and/or editing the functional relationships) and observe whether it converges or not to the true model (i.e. whether the difference between the predicted rating and the real rating converges towards zero).
6. <u>Data browsing</u>

   By clicking on the ‘Alex’s data table’ button the students can access the data for either Alex, which are presented in a table. Students can access this table at any point during the task.
7. <u>Graphing tool</u>

   Students can freely navigate between the digital tools provided in this task: the conceptual modelling tool ‘YOUMODEL’ and the graphing tool ‘YOUGRAPH’. In the YOUGRAPH tab (see Figure B2), students are provided with a graphing tool that they can use to explore and visualise relationships between the variables. In the graphing tool, students can select the explanatory variables in the Y-axis, as well as choose to add a third variable and fix its value to analyse interactions. Once the variable(s) are chosen, the graph populates a scatter plot with all the available data and produces a line of best fit to support students’ conceptual understanding of relationships. The use of these tools are explained to students in video

Figure B2. Visual of the graphing tool, YOUGRAPH.



Task structure

19. This task has five components or phases:

1. Introduction of overall learning goal (calibration of motivation)
2. Evaluation of prior knowledge (optional)

3. Tutorial (instructions on the use of the modelling environment)
4. Learning activities
5. Applications

Introduction of overall learning goal (calibration of motivation)

20. The first phase starts with a static page that serves as an introduction to the task and its overall learning goals, contextualised within a real-life scenario (see Figure B3). As a cover story, the students are asked to build a model that allows them to predict whether an individual will like a restaurant, using data about the restaurant's characteristics and the individual's past rating of other restaurants. A gender-neutral computer agent will present the cover story to students to create a more natural learning environment. This computer agent will reappear during the task to ask students questions related to their self-regulatory processes, and to present scaffolded rescue points.

Figure B3. Example of static page introducing the learning task and goals by a virtual guide

Evaluation of prior knowledge (optional)

Purpose

21. Computational modelling tasks might include an 'evaluation of prior knowledge' phase. Prior knowledge in computational modelling tasks refers to several types of 'knowledge' that a student might possess prior to engaging with the learning task, including: (1) knowledge about the (content) issue; and 2) knowledge of data analysis, modelling or algorithmic methods (the latter in the case of more complex, agent-based models that require controlling the agents' behaviour through block-coding). It is important to include questions to measure what students already might know about the phenomenon if we expect substantial differences, both across and within countries, in students' opportunity to learn about that phenomenon. For example, if the task is about coral bleaching then we might expect that some students are given the opportunity to study coral ecosystems as part of their school curriculum. This opportunity could affect how easily they might be able to progress during the learning and modelling phase of the task.

Description

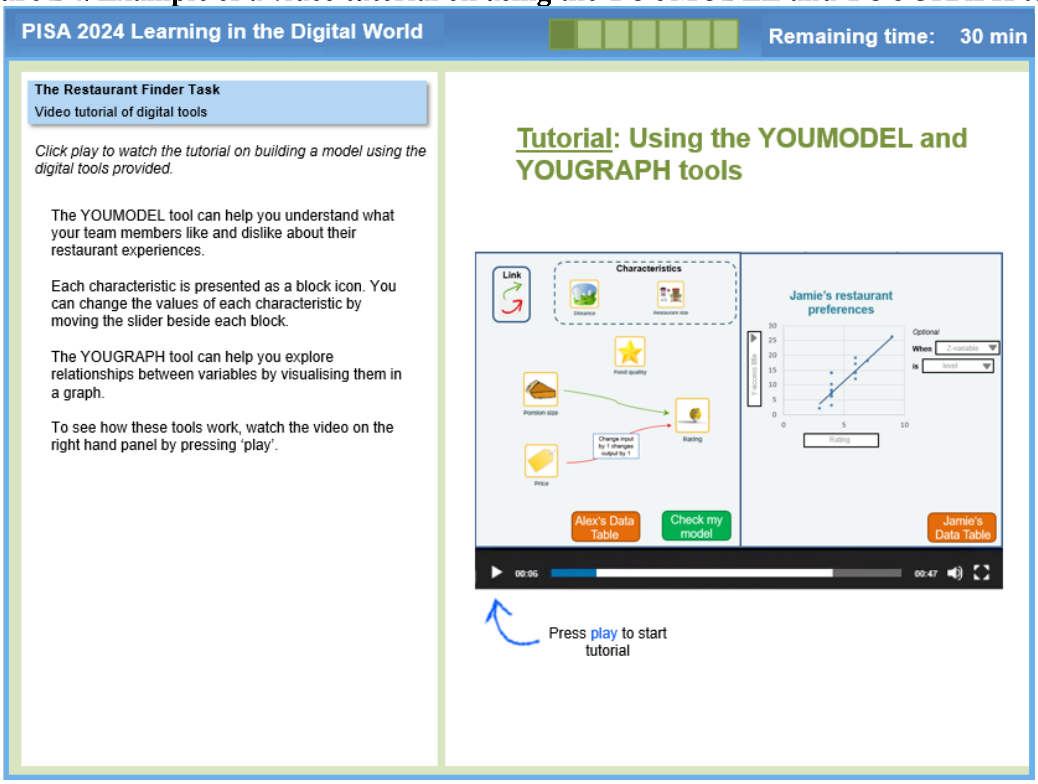
22. In this specific task, it might be less relevant to include such prior knowledge questions given that the underlying model is not based on an 'established, real-world truth' (i.e. the model is based on a formula that it is not possible for students to already know). Students' knowledge of modelling and data analysis methods, however, might be relevant to their performance in the task. Information on what students already know about modelling and data analysis can be collected through specific, stand-alone items included in a test cluster either at the beginning or end of the testing time (for example, asking students to interpret relationships in scatterplots, interpret linear equations, etc.).

Tutorial (instructions on the use of the computational modelling environment)

Purpose

23. The main purpose of the tutorial phase is to familiarise students with the specific computational modelling environment and some basic concepts of data analysis (e.g. how to interpret a scatterplot with line of best fit, see Figure B4). Although the modelling and graphing tools are designed to be intuitive and relatively easy to use for all students, some students may not be familiar with such learning experiences.

Figure B4. Example of a video tutorial on using the YOUMODEL and YOUGRAPH tools



Description

24. In this specific example, the tutorial is delivered in one short video with embedded text that: (1) explains what the digital learning environments are used for; and (2) progressively illustrates the basic features of the digital learning environment (for example, dragging and dropping concept blocks, changing values and labelling relationships in the modelling tool, and visualising data through the graphing tool). One advanced feature of the environment (adding interactions) will be introduced later in the task. Students can watch the video as many times as they want, pausing whenever necessary.

Learning activities

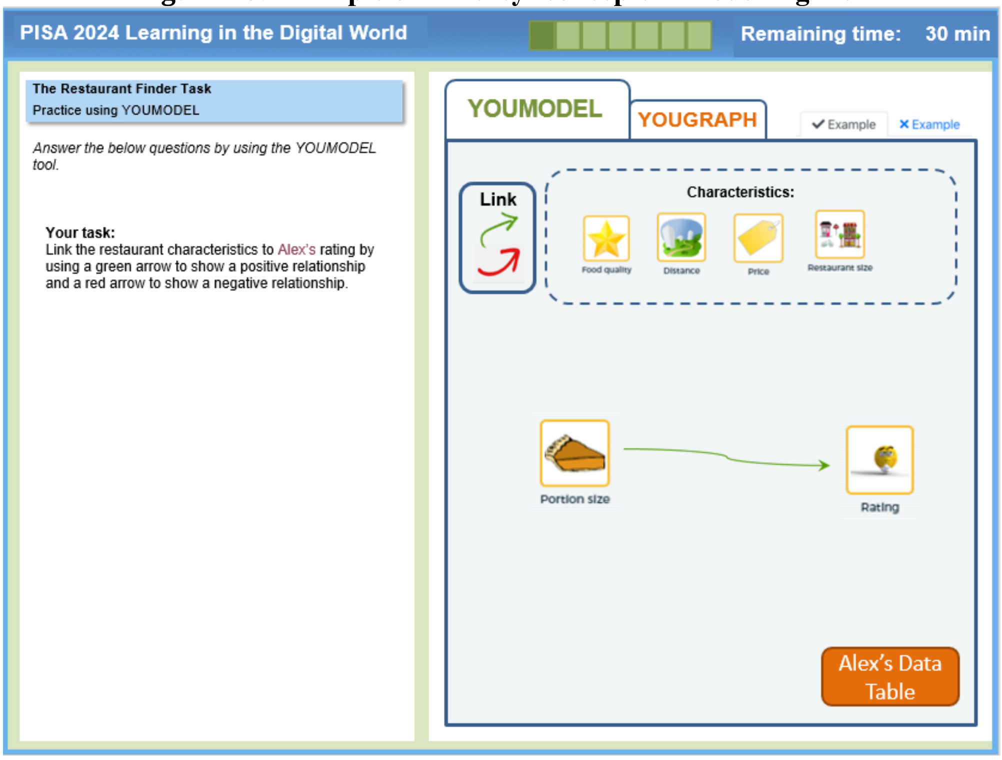
Purpose

25. This central phase of the task provides students with the opportunity to build a model of Alex's and Jamie's preferences for restaurants. The modelling process is scaffolded through a sequence of increasingly complex tasks, during which students progressively acquire more information on the system of relationships and practice and improve their modelling skills. The digital environment supports their learning by providing automated feedback on their model and providing several worked examples that students can access.

Description

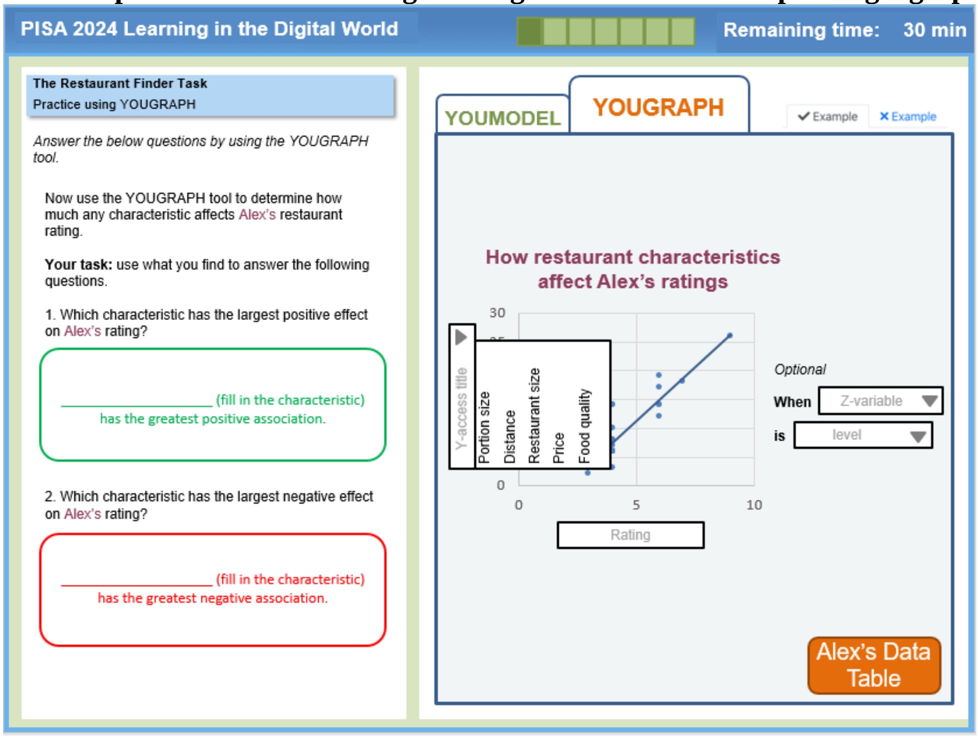
26. The first challenge in the learning activities phase is relatively easy (see Figure B5). Students are asked to specify whether each variable that they include in the concept map is positively or negatively related to the outcome, choosing to connect variables either with a green arrow (to convey a positively relationship) or a red arrow (to convey a negative relationship). Students will investigate the relationships in the data provided to them through the graphing tool (YOUGRAPH), and will represent their findings in the concept map tool (YOUMODEL). The 'good' and 'bad' worked examples provided for this task focus on the interpretation of scatterplots.

Figure B5. Example of an 'easy' conceptual modelling task



27. In the second challenge (see Figure B6), students need to compare the magnitude of the relationships by creating graphical representations of the data, and indicate which variable has the strongest positive relationship with the rating and which has the strongest negative relationship with the rating. The worked examples here focus on how to interpret the steepness of a line of best fit.

Figure B6. Example task for determining the magnitude of relationships using a graphing tool



28. These two challenges function as a 'warm-up' for the main complex learning activity, which consists of creating two models that reflect Alex's and Jamie's preferences respectively. In the first part of the main learning activity, students are asked to define the model for Alex using the concept map tool and the graphing tool (see Figure B7). The correct model includes four relevant variables that are linked to the outcome (Alex's rating): two with a positive relationship, and thus green arrow (food quality and portion size); and two with a negative relationship, and thus red arrow (distance and price).

29. Students can check how well their model for Alex's preferences align with the provided data by clicking on the 'Check my model' button. This button populates a table (see Figure B8) that is automatically updated every time the student adds and labels links in the concept map, and which provides information on Alex's actual rating of 21 restaurants and the ratings produced by the student's constructed model. A perfect model would result in zero difference between the two ratings. For the model to correctly predict Alex's preferences, all of the relationships between the variables must be labelled correctly and indicate the quantitative relationship. The students have the option to review worked examples that both provide clues on ways to conduct the data analysis and modelling processes and highlight common mistakes.

Figure B7. Example of main learning task (constructing a model with the concept mapping tool)

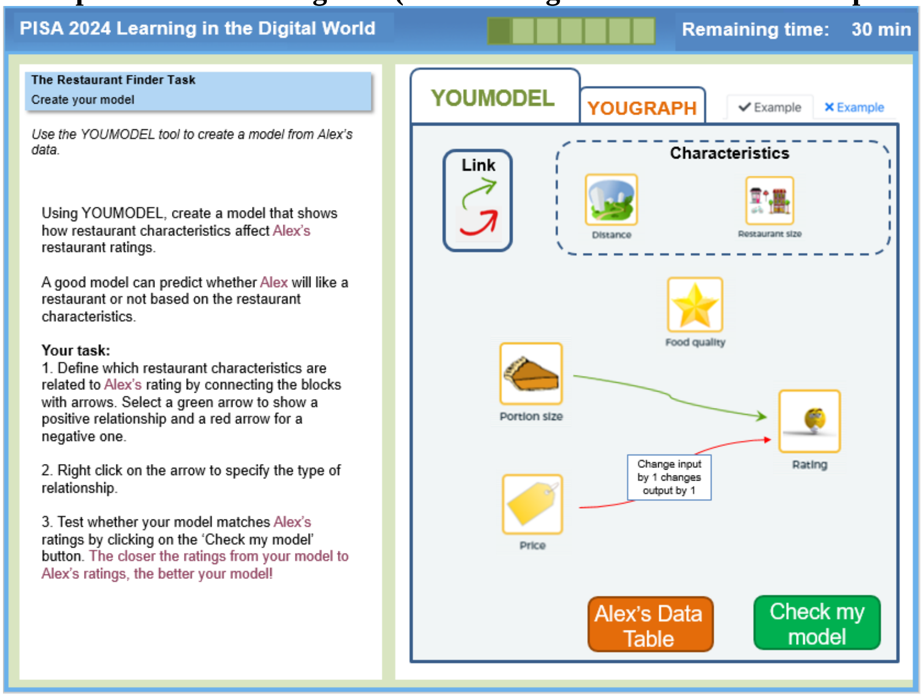
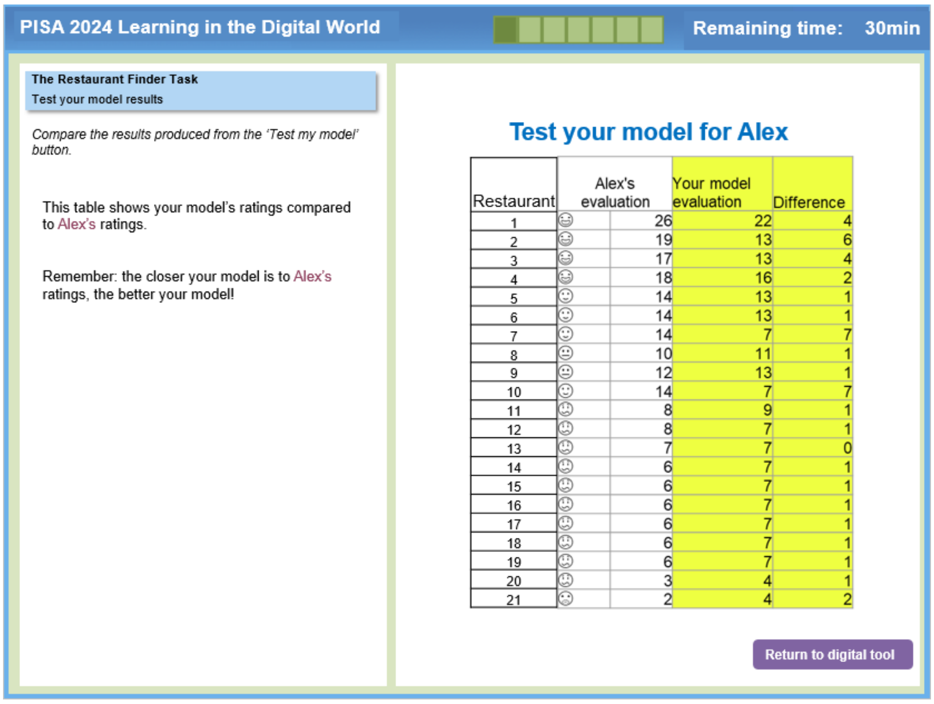


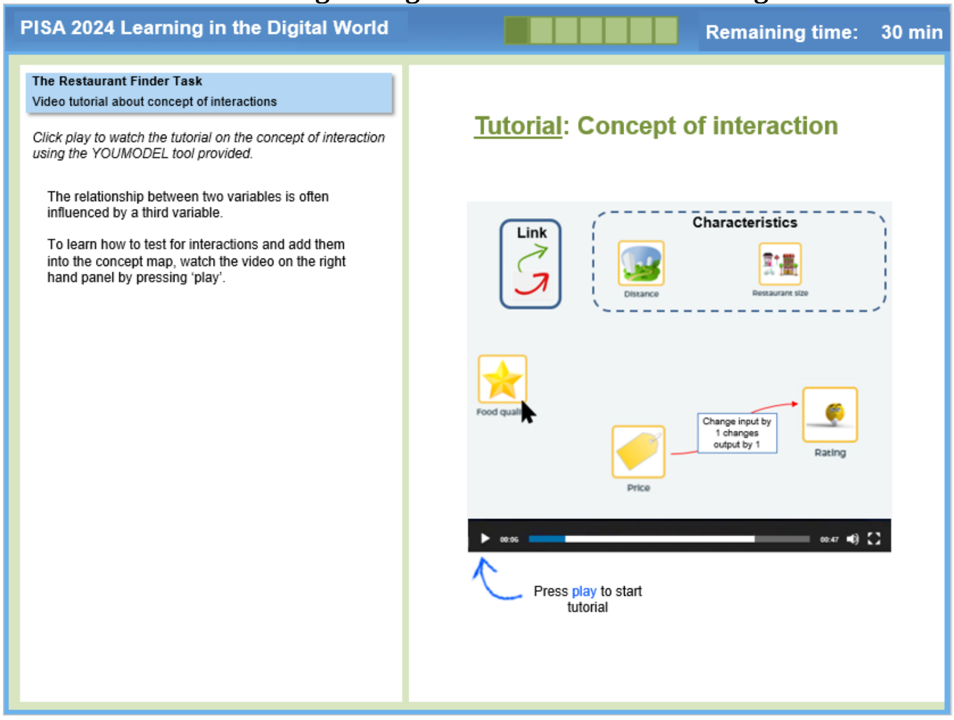
Figure B8. Example of 'Check my model' button interface enabling students to check the correctness of their model



30. The second part of the main learning activity asks students to produce a concept map for Jamie's preferences, working from the correct concept map for Alex which students will be provided through a rescue point. This is to ensure that all students begin the second part of the learning task from the same point. The model underlying Jamie's preference is more complex than the one for Alex, as it incorporates one interaction between price and food quality (the impact of food quality on the rating is lower, if the price increases). It is important that during this assessment students are provided with the opportunity to learn concepts that they do not already master. Before they start developing the concept map for Jamie, they will be provided with some explanation on the concept of interaction (possibly through a video, see Figure B9). They are also provided with two different worked examples of

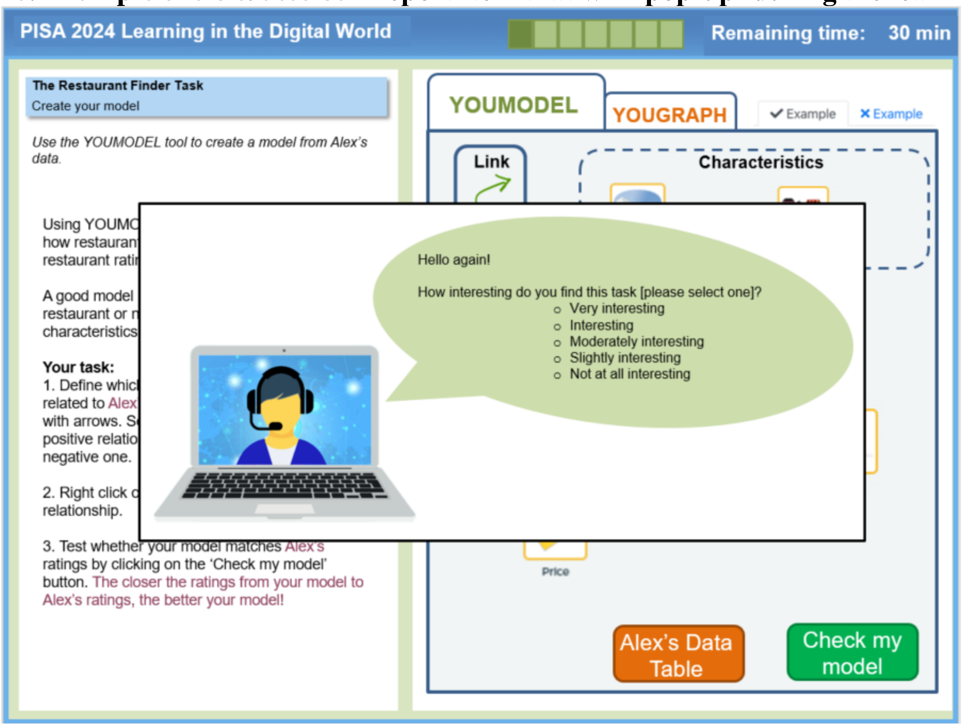
modelling tasks within a different context that incorporate interactions in the models. These worked examples demonstrate how students can identify interactions through the graphing tool.

Figure B9. Tutorial demonstrating testing for interactions and adding them to a concept map



31. Within the complex learning tasks, situated self-report items can be used to provide information on students' self-regulatory processes in addition to process (log-file) data of their interactions with the environment. These self-report items will "pop-up" during a task and ask students to answer some questions related to their motivation and affect. The questions will be presented by the virtual guide, with whom students should be familiar. Figure B10 provides an example of the format and type of situated self-report item that can be implemented within the tasks to measure students' interest in the task they are working through, which is related to their motivation for engaging with the task.

Figure B10. Example of a situated self-report item that will 'pop-up' during the learning phase



Applications

Purpose

32. The last phase of the task is meant to assess whether students understand how they can apply models to make predictions (as the ultimate purpose of the activity is to build an application that can predict which restaurants people will like with reasonable accuracy). In order to reduce dependencies between what students achieved during the main learning activity phase and the application phase, students will be provided with the correct model for Jamie's preferences at the start of this phase. Similar rescue points can be designed across all of the learning tasks in this assessment. The assessment interface could also have a 'time remaining' feature that indicates the time remaining for the learning activities phase, which then graduates students automatically to the application phase once the time limit has expired.

Description

33. One possible item in this phase might ask students to select the restaurant, from a list of three possibilities, that Jamie would prefer given the correct model of Jamie's preferences and the data provided on characteristics of the restaurants (see Figure B11). Other, more complex prediction items can also be included to provide additional data points and target different difficulty levels.

Figure B11. A simple prediction item

Data collected

34. Throughout the entire task, the assessment delivery system will be required to capture several data points. These data points will include both response data and process (or log-file) data.

35. For the closed items (e.g. selecting the right restaurant) in the task, the system records:

- Student answers;
- Timestamps of student answers.

36. The system also logs the following aspects of students' interactions with the environment as 'events':

- Clicks on buttons/tabs;
- Sequence and timestamps of semantically meaningful actions (e.g. including blocks in the concept map, linking blocks, creating scatterplots, browsing through worked examples, watching tutorial);

Measurement and assessment of the construct

Assessment of basic digital literacy

37. Basic digital literacy is roughly defined as the ability to use interactive affordances and interpret the representations of a digital environment. It is a necessary co-variate in the analysis of students' ability to engage in the modelling task with digital tools; if students do not know how to use a digital environment, their ability to model phenomena through the digital environment is greatly impacted.

38. Different factors of students' interactions with the digital environment can be aggregated into a measure of their basic digital literacy. For example, students with sufficient basic literacy can be expected to recognise which elements of the user interface are interactive (e.g. buttons), and which ones are not. A system that logs all mouse clicks can enable inferences on useless actions. Students with sufficient basic digital literacy can also be expected to quickly follow straightforward instructions, such as connecting two blocks with arrows. Measuring these different aspects by examining traces and timestamps throughout the tasks might provide sufficient data to determine students' basic digital literacy.

Assessment of performance (cognitive practices)

39. Student performance on this task is primarily evaluated on the basis of the completeness and correctness of their final computational models, through the use of polytomous scoring. The items included before the main learning activities phase and during the application phase of the task will provide additional data points for estimating student proficiency on the computational modelling component of the competency model.

Assessment of self-regulation processes

40. Students' self-regulatory processes will be assessed by looking at the choices that they make (as reflected in their actions within the environment), as well as by using data from situated self-report items.

41. Several measures of students' engagement in 'metacognition and cognitive strategies' could be constructed using both process data and situated self-report items. For example, sequence analysis of students' actions during the learning phase can be conducted to determine students' use of coherent, cognitive strategies. Process data relating to their time spent on 'productive' actions (e.g. accessing information or seeking help when stuck) could also be used to develop a measure of this component. Situated self-report items could be included at the beginning of the task to evaluate students' self-efficacy with respect to successfully completing the task. Additionally, these items could provide indicators of students' metacognition based on their evaluation of the quality and accuracy of their work.

42. To construct measures of 'motivation and affect', situated self-report items would be included to evaluate how students value the task (e.g. their interest in the task), in addition to asking them to report their affective state(s) as they work through the activities. Process data measures could also provide information on students' willingness to continue exerting effort throughout the task, both when faced with a novel environment and when faced with difficult challenges.

43. All situated self-report items will be carefully drafted to avoid stereotype threat and bias.

Example Task 2: Programming Karel

Overall Description

44. The purpose of the task is to evaluate students' ability to engage in a process of self-regulated learning of computational practices. The students will need to develop algorithmic solutions to computational problems using basic control structures and concepts, such as decomposition and repetition. Students will learn incrementally through a series of scaffolded problems, using a digital environment that provides resources for learning specific control structures and concepts and that gives immediate feedback on students' programs (i.e. algorithmic solutions). The resources for learning consist of worked examples, which show how to use the target concepts on problems similar to the given problem in the task.

45. At its core, the task is about programming a simple robot in the form of a turtle ("Karel") to execute actions in a grid-like world (up-down and left-right). In this world, Karel understands four basic instructions: "move forward", "turn left", "place stone", and "pickup stone". These programmed actions can be seen as abstractions of various navigational problems in the real world; these real-world problems are used as cover stories for the tasks.

46. The proposed digital learning environment is a block-based programming environment. This environment and type of task has been chosen because similar environments/tasks have been implemented and tested in various contexts around the world, and have proven to be both intuitive and easy to engage with even for young children or those with limited programming experience.

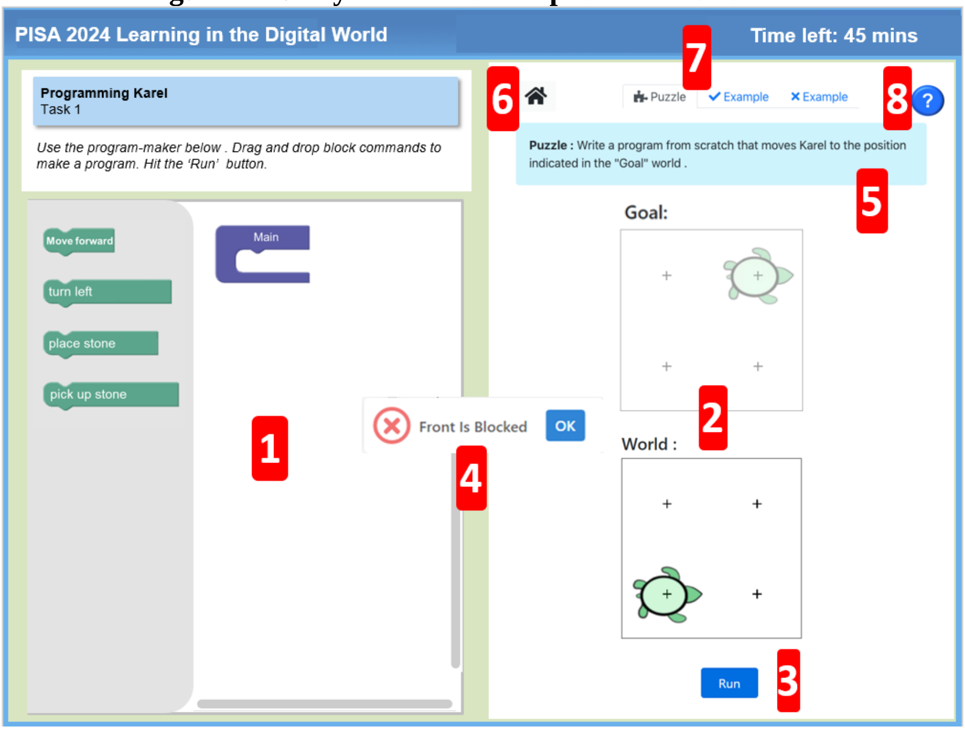
47. The task can include a range of problem types. Examples of different problem types are open-ended items requiring students to generate their own algorithm, to complete a partial program, to fix or modify an existing program, or to interact with a given program and environment (e.g., clicking buttons to add stones). Other types of tasks are multiple-choice items that ask students to make predictions about the execution of a given program.

Key elements of the learning environment

48. The digital learning environment consists of two main interfaces: the problem environment and the "learning centre", which is a dashboard that shows students' current state of progress through the whole task. Figures B12 and B13 show a possible instantiation of these two interfaces.

49. The digital learning environment is "open", meaning that students can move freely back and forth between the two interfaces and various subpages (e.g. worked examples).

Figure B12. Key elements of the problem environment



50. The key user interface elements of the problem environment are enumerated in the Figure B12 and include:

1. <u>Block-based programming editor:</u>
   The editor can be customised dynamically, in terms of the types of blocks and functionalities provided to the students. In addition, hovering the mouse over a block of code opens a text bubble explaining the functionality of that specific block.
2. <u>World:</u>
   This is the grid-like world within which Karel exists and executes actions. Each pixel can hold up to a set number of stones (blue diamonds). Upon program execution, the system models the Karel behaviour in the world.
3. <u>Program controls:</u>
   Code in the block-based programming editor is executed as a simulation of Karel's behaviour in the 'execute' world. Students can execute the code either by running the code in its entirety (clicking 'Run'), or by stepping through the code line by line (clicking 'Step').
4. <u>Pop-up messages:</u>
   These messages appear to communicate errors to the student.
5. <u>Item instructions:</u>
   Each item instruction succinctly states what students have to do on the item and what the desired goal is. The goal can also be represented visually, displaying the desired final state of the world upon problem completion.
6. <u>Home button:</u>
   A 'home' button allows students to access the 'learning centre' (see Figure A13), which contains the various learning problems and the so-called 'Big Challenge Problem'.
7. <u>Worked example tabs:</u>
   Students can switch between three tabs. The puzzle tab contains the current problem to be solved. The two example tabs provide students with two worked examples, one "good" and one "bad", related to the target concept and the problem at hand (or 'puzzle'). The worked examples help students to learn about the target concepts and to make progress on the current problem.

8. Help button:
The help button is accessible to students during the main learning activity phase. This button gives students hints about how to progress through the problem (e.g. directing students' attention towards key blocks, prompting use of interface tools, etc.).

51. The 'learning centre' (Figure B13) is the home navigation screen that allows students to navigate between sub-tasks, learning goals and the 'Big Challenge Problem' by target concept. This navigation screen provides students with more direct control over their learning progress. The learning centre contains sub-tasks that aim to either: (1) teach a specific concept (arranged in rows); or (2) allow students to practice the target concepts within variations of a problem (arranged in columns). In order to unlock a sub-task, students have to solve the preceding sub-tasks. The learning centre also includes the 'Big Challenge Problem,' which is the open-ended, complex problem that incorporates the use of all target concepts.

Figure B13. Learning centre navigation screen

Task structure

52. The task has five phases:

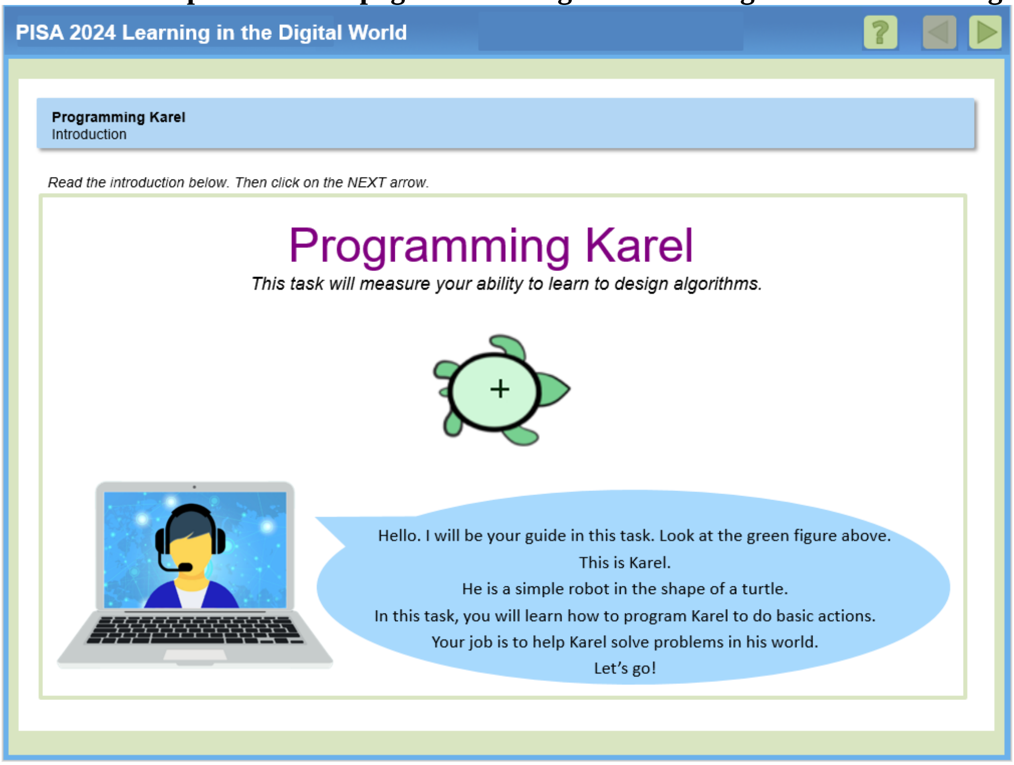
1. Introduction of overall learning goal (calibration of motivation)
2. Tutorial (on the use of the block-based programming environment and learning centre)
3. Evaluation of prior knowledge (optional)
4. Learning activities

Introduction of overall learning goal (calibration of motivation)

53. The first phase starts with a static page that explains the task and overall learning goals, contextualised in a real-life scenario. A computer agent will present this information to the student (see Figure B14). A possible cover story is that students have to program a robot that can collect and deliver

food to different places in a city. The introduction could also involve a video or animations to bolster the story.

Figure B14. Example of a static page introducing the task and goals with a virtual guide



Tutorial (on the use of the block-based programming environment and learning centre)

Purpose

54. The tutorial is designed to familiarise students with the affordances and representations of the digital learning environment, and to make sure that they know how to navigate within it. Although the block-based programming environment is designed to be intuitive and relatively easy to use for all students, some students may not be familiar with such learning experiences. The tutorial might consist of both videos/animations and interactive tasks for which students need to follow and execute simple instructions. This is to ensure that students do not just see how the environment is being used, but actually use it.

Description

55. The tutorial starts with a video introducing the world of "Karel". Students are then led through a sequence of short interactive tasks to learn the basics of how to control and program Karel. The first four sub-pages introduce the basic actions Karel can complete by means of interactive buttons that control Karel in real-time, as exemplified in Figure B15. In this example, students are instructed to click the "move" button, which in turn makes Karel move one step forward in real-time in the 'execute' world. When the state of the world matches up with the goal state, an animation indicates that the student was successful. The remaining three instructions ("turn left", "pickup stone", "place stone") are similarly introduced. Once completed, students will then be instructed to solve a couple of simple items that require them to click a combination of buttons, as exemplified in Figure B16.

Figure B15. Example of basic instruction item using interactive buttons

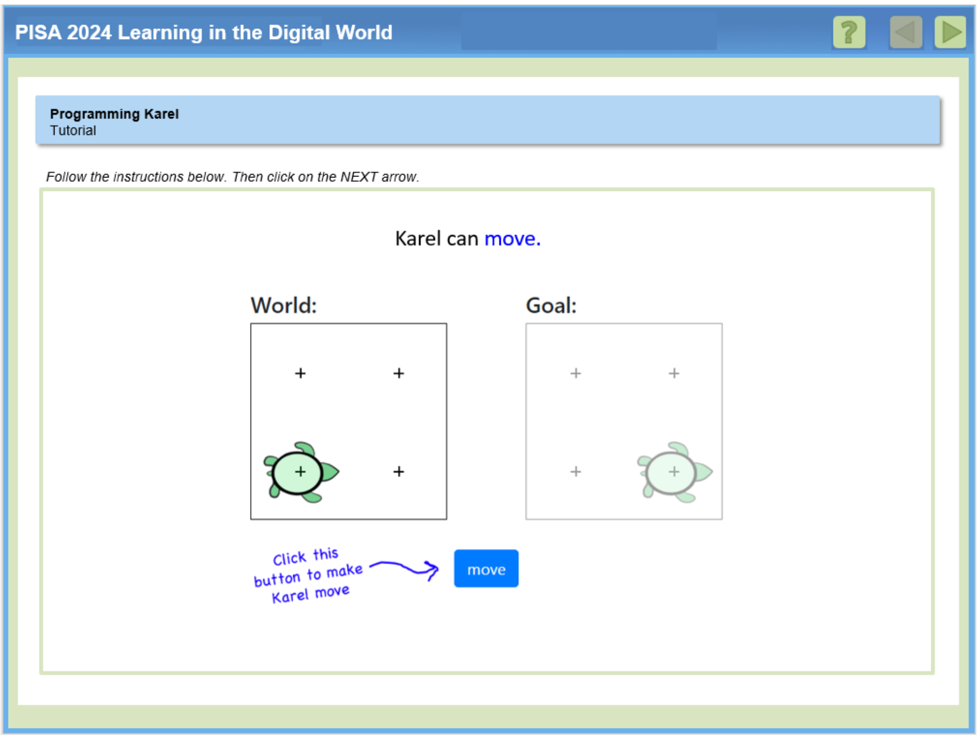
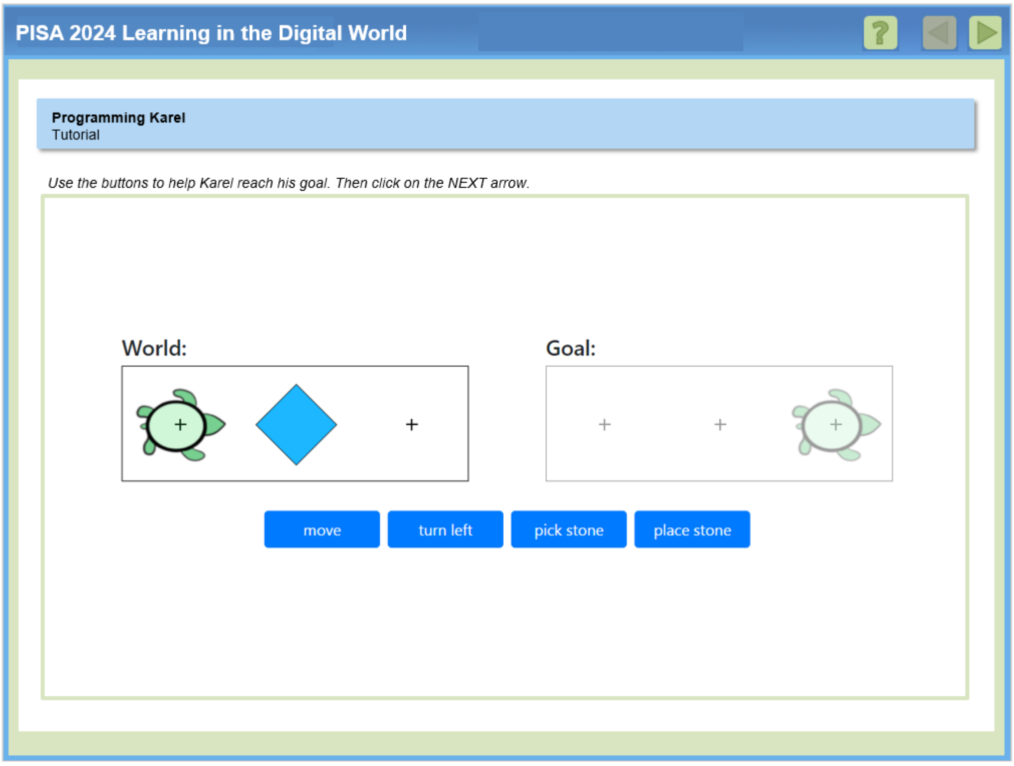


Figure B16. Example of simple item requiring a combination of instructions using interactive buttons



56. The tutorial then transitions from the interactive buttons to the programming environment, which is presented to students as a more powerful way to control Karel (e.g. through a video). A subsequent sequence of interactive sub-pages helps students to learn how to use the programming environment to

build a program to control Karel. For example, they are shown an animation demonstrating how to build code by dragging and dropping blocks into the programming editor, or that clicking the "Run" button executes the program line by line (as exemplified in Figure B17). Students are then given the opportunity to experience building simple programs themselves by dragging and manipulating blocks into the programming editor, as exemplified in Figure B18. An automated feedback system that involves pop-up texts (e.g. "well done") and animations (e.g. "hit the run button", accompanied by an arrow pointing to the "Run" button) will help to guide students through the tutorial sub-pages. Subsequently, the tutorial introduces the "Examples" tab to students, either by means of a video, animation or sequence of interactive instructions. Finally, students will be introduced to the learning centre, where each feature is explained by means of animations of arrows and text.

57. The tutorial is designed to help students understand the key principles of programming Karel and accustom them to the digital learning environment and its resources. Students therefore cannot proceed to the learning activities until they have successfully completed all of the steps of the tutorial. Each step has a time limit to ensure that all students can nonetheless advance to the main learning activities phase; if students reach the time limit, the system automatically shows students the correct solution in order to complete the step.

Figure B17. Example of animation demonstrating the "Run" function

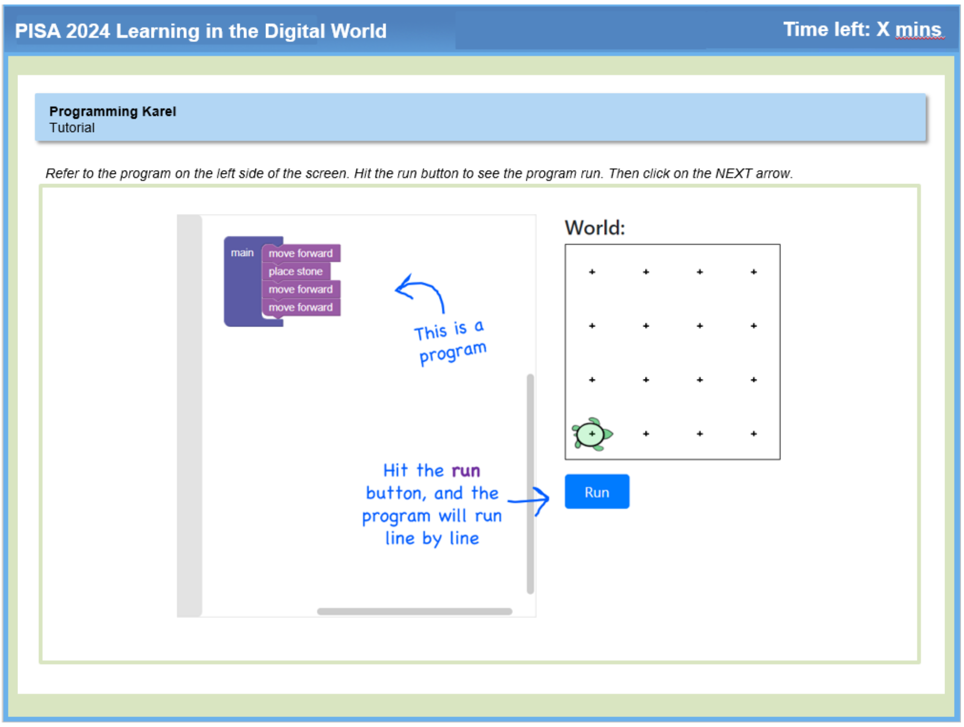
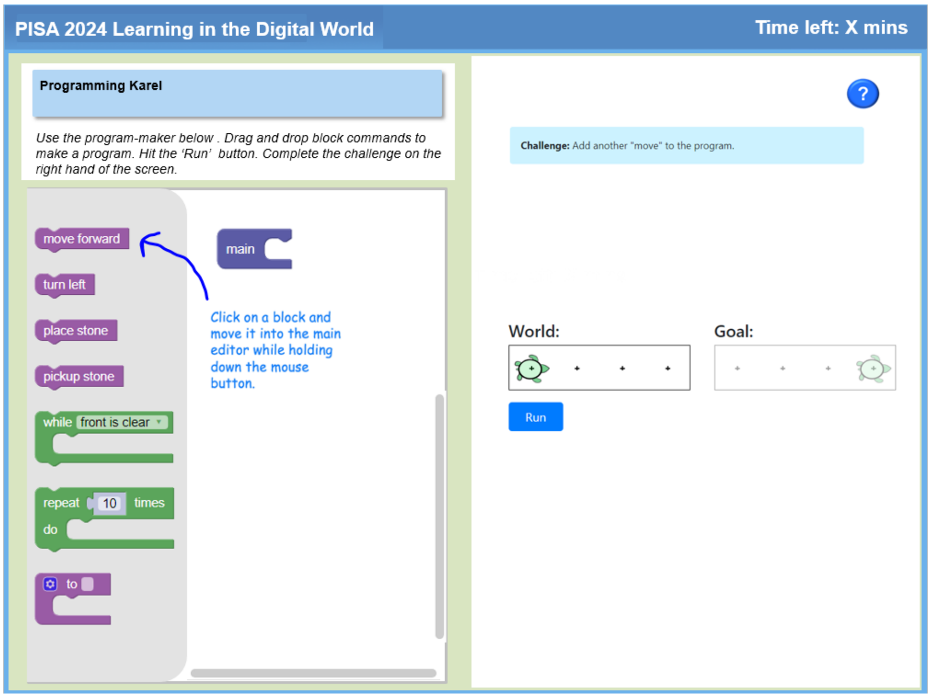


Figure B18. Example item teaching students to build simple programs in the programming editor



Evaluation of prior knowledge (optional)

Purpose

58. Prior knowledge in computational problem solving tasks refers to students' knowledge of computational science concepts (e.g. decomposition and repetition) and their ability to create algorithms (in form of programs) using these concepts. Measuring students' prior knowledge of these aspects is important not just to evaluate how much students learned through the task, but also if we expect substantial differences in students' ability to learn by incoming knowledge. Information on prior knowledge can be collected both in task, for example during the tutorial, and through stand-alone items.

Description

59. Stand-alone, multiple-choice items could evaluate students' ability to reason with or about code. For example, students could be a program accompanied by the initial state of the 'execute' world, and asked to choose the correct final state of the world (e.g. Figure B19). Alternatively, students could be given the initial and final states of the world and asked to choose which code has been executed.

60. Stand-alone, open-programming items could also require students to generate, modify or debug existing code. However, the use of these items will be limited both in quantity and time, so as not to overwhelm or fatigue students before the main learning activity. Depending on the eventual evaluation framework, an open-programming item could be isomorphic to the final Big Challenge Problem (e.g. Figure B20). This could be used both to frame the purpose of the learning activity, namely learning how to program Karel to collect all of the objects in the most efficient way, and to get a gain measure on the isomorphic items. This item will be limited to a few minutes, but students can freely approach it in whichever way they want.

Figure B19. Example of multiple-choice item evaluating prior knowledge

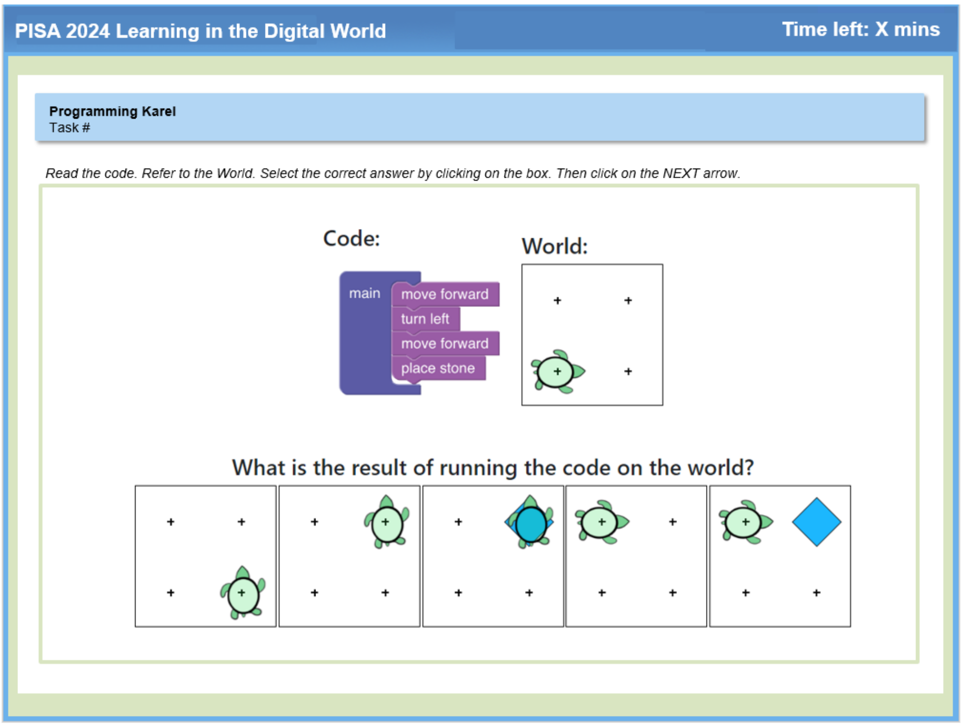
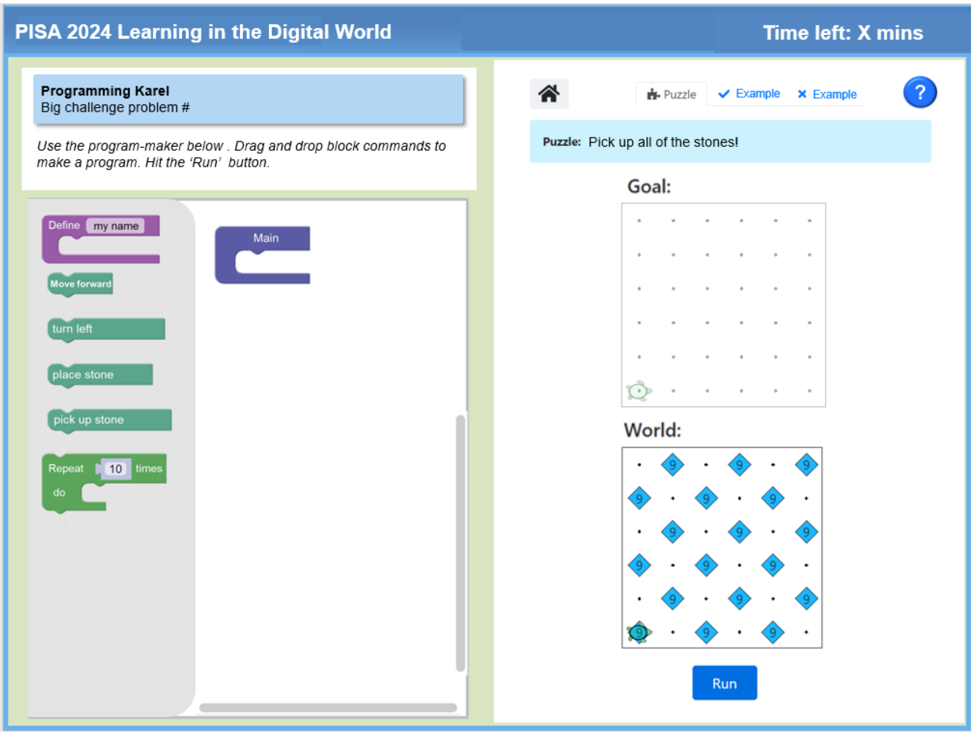


Figure B20. Complex, open-ended item evaluating prior knowledge and similar to the 'Big Challenge Problem' used in the application phase



Learning activities

Purpose

61. This central phase of the task provides students with the opportunity to learn how to design algorithmic solutions for Karel. Whereas the tutorial teaches students how to interact with the specific

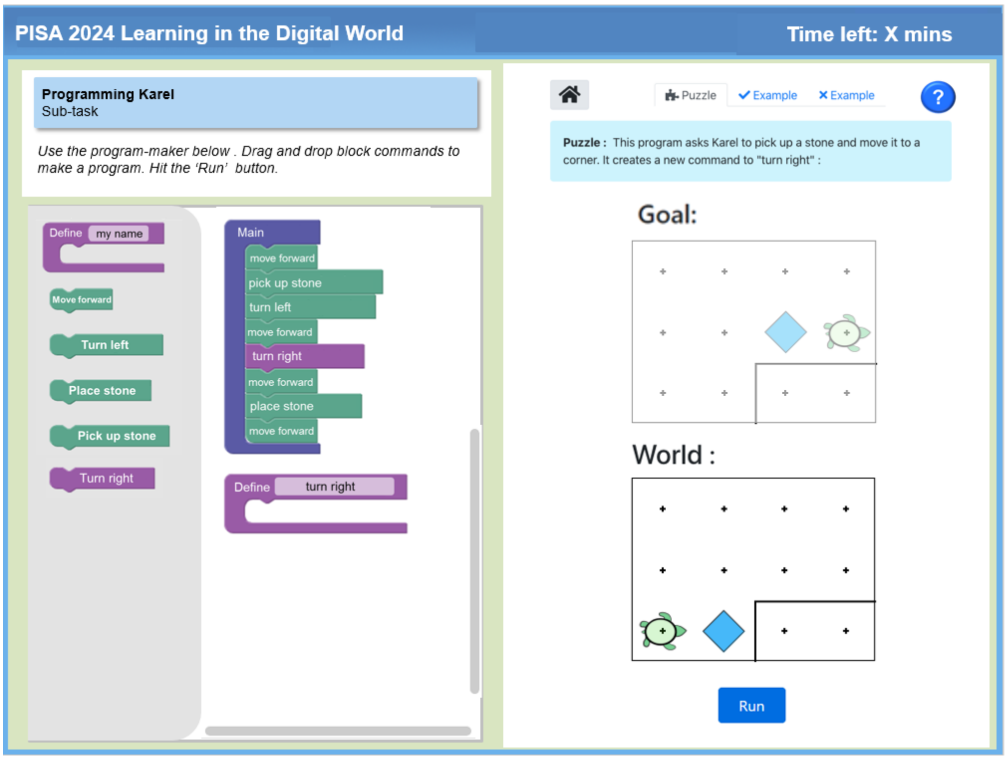
elements in the learning environment, the learning activities phase introduces students to core computational concepts incrementally, through a series of sub-tasks, in the form of control structures such as decomposition (functions) and repetitions. Each sub-task presents a new problem, or modified version of a preceding problem, designed to introduce students to a new concept altogether (e.g. repeat block) or a new use of an already-introduced concept (e.g. movement instructions within repeat block). There are multiple ways to go about solving each sub-task, and students can learn and develop their understanding of concepts by trying to solve the problems, by studying the worked examples, and by asking for hints when needed. At the end of the phase, students are given the opportunity to demonstrate their learning by completing the Big Challenge Problem (a complex, open-ended item).

Description

62. The sub-tasks in the learning activities phase follow a linear progression, from simple to more complex. The progression starts with simple sequences of the basic instruction blocks. For example, students will first learn about encapsulation (i.e. how to combine a sequence of blocks into a new block); this is used for functional decomposition (i.e. breaking a problem into sub-problems that can be solved individually). Students who successfully complete one sub-task unlock other increasingly complex sub-tasks in the learning centre. For example, after the first sub-task students will then progress to learn about including repetition blocks within function blocks to generate complex decompositions. If students are stuck on a sub-task, they have the option to return to the learning centre to consult one of the earlier sub-tasks that they solved. This will be followed by increasingly complex sub-tasks.

63. An example sub-task during this phase of the unit is shown in Figure B21. The goal of the sub-task is to design a “turn right” block to help Karel move from the initial position in the ‘execute’ world to the goal position. The sub-task is designed to introduce students to the concept of encapsulation (for functional decomposition), and ideally follows a sub-task that introduced the “repeat” block. The item instruction bar (in the upper right corner) informs students that they can build new commands (i.e. blocks) to solve problems – in this case, moving Karel around a wall by building a “turn right” block. Students can learn how to build new commands either by seeking situated feedback on their attempt(s) (i.e. interacting with the programming editor, running their program and receiving situated feedback on the success of their attempt) or by studying the provided worked examples.

Figure B21. Example sub-task teaching students the concept of encapsulation



64. The worked example tabs in this example provide students with "good" and "bad" examples of how to build a program using an encapsulation block to reach the goal. Figure B22 presents a "good" worked example: it shows how the "turn around" block is defined with other blocks. In contrast, Figure B23 shows a "bad" worked example: although the program solves the problem, it does not reflect the most efficient way to do so (i.e. using encapsulation to define the "turn around" action).

Figure B22. Example of a "good" worked example to help students learn the concept of encapsulation

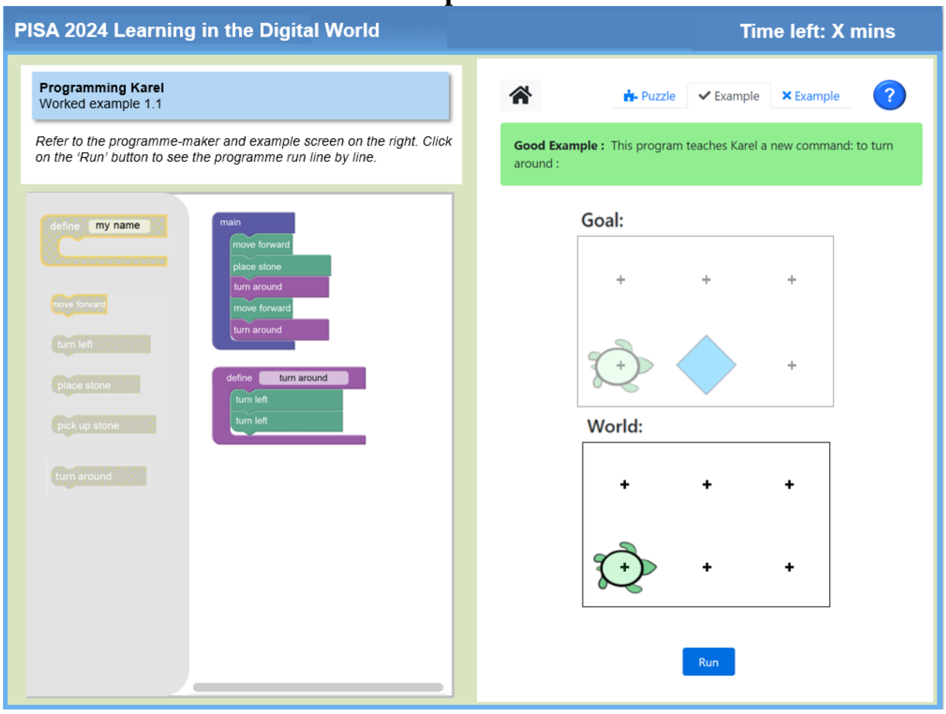
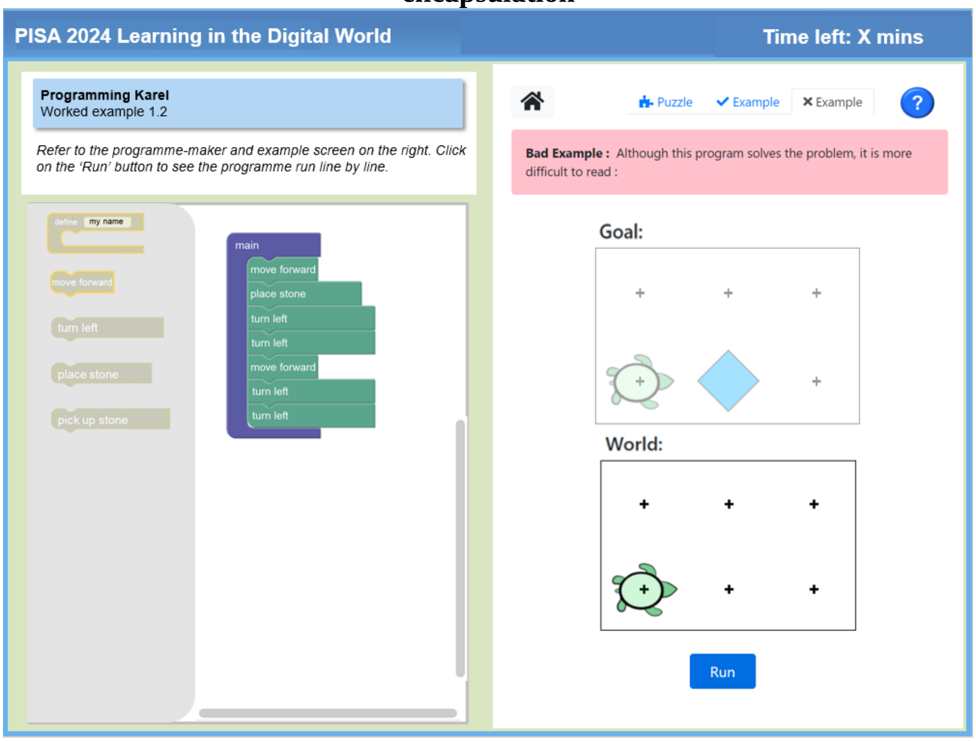


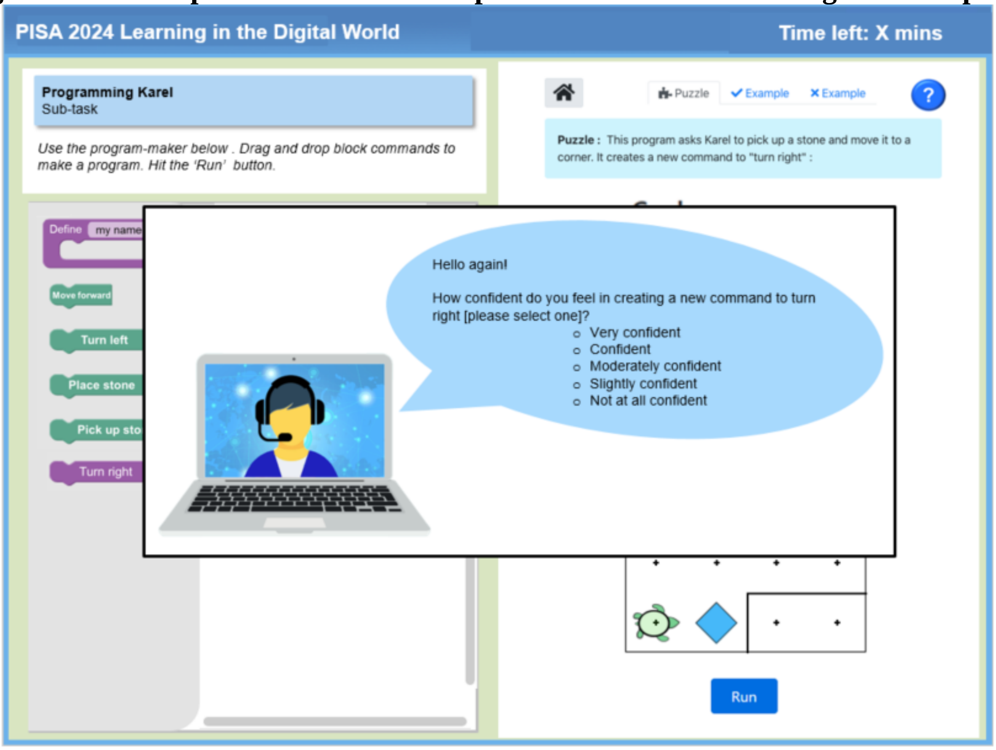
Figure B23. Example of a "bad" worked example to help students learn the concept of encapsulation



65. In addition to the worked examples, students can ask for hints (by means of a 'hint' button) throughout the learning activities phase. For instance, in the encapsulation sub-task above, students might be provided with the following hint: "Look at the blocks that are available to you. Do you recognise which block is new? Look at the example tabs to learn more about this block."

66. Within the main learning activities phase, students may also be presented with situated self-report items that aim to gather information on students' self-regulatory processes related to managing their motivation and affect. These questions will "pop-up" during a task and will be presented by the familiar virtual guide. An example situated self-report item and format is presented in Figure B24, which aims to measure a student's confidence in learning the skills necessary to succeed on the task (related to motivation).

Figure B24. Example of a situated self-report item within the learning activities phase



67. The overall time of the learning activities phase is fixed; proficient students will therefore complete more sub-tasks. At the end of the unit, students are presented with the 'Big Challenge Problem' – a final complex, open-ended item (isomorphic to the one given during the evaluation of prior knowledge phase, if used). All students automatically get to this item after a set amount of time. Students are given a fixed amount of time to complete the item (e.g. 7 minutes).

Data, measurement and assessment

Data collected

68. Throughout the entire task, and within each sub-task or item, the assessment system will be required to capture several data points. These data points include both response (or outcome) data and process (or log-file) data.

69. For closed items (e.g. multiple-choice or self-report items), the system records:

- Student answers;
- Timestamps of student answers.

70. For open-ended items, the system logs the following aspects of students' interactions with the environment as events:

- Clicks on buttons;
- Changes of state of the programming editor (e.g. new block dragged in, block deleted, etc.);
- Sequence and timestamps of semantically meaningful actions (e.g., changing code, running code, clicking on ‘hint’ button, opening resources tab);
- Sequence of programs created and tested;
- Sequence of successful and failed attempts for completing an item;
- Interventions of the automated feedback system;
- Sequence of resources accessed;
- Timestamps of above events (including relational times, e.g. time from instruction to execution of actions).

71. The system also records the number of sub-tasks successfully completed by students during the learning activities phase, as well as the final programs submitted by students for each open-ended item (i.e. response data for those items).

72. In addition to creating measures of the construct, part of the data collected throughout the task will be used to generate adaptive feedback in real-time. This adaptive feedback is designed to ensure students are successfully guided throughout the tutorial steps and ultimately succeed in completing them. Part of the data will also be used to assess students’ basic digital literacy, as explained in the Yum! Prototype example.

Measurement and assessment of the construct

Assessment of basic digital literacy

73. See the description in the Yum! Prototype example for an explanation of how digital literacy can be measured in these tasks. The start of the assessment task will include some items specifically meant to assess students’ basic digital literacy.

Assessment of prior knowledge

74. Data from the evaluation of prior knowledge phase will be used to assess students’ prior knowledge/ability. Students’ performance on the multiple choice questions (i.e. number of correct responses) and on the open-ended task(s) can provide a composite score of prior knowledge/ability. Performance on the open-ended task(s) includes evaluating the quality of the student-generated program, in terms of its code complexity, structure (e.g. functional decomposition) and completeness (i.e. how many objects Karel successfully picks up).

Assessment of performance (cognitive practices)

75. Data from the whole task, with the exception of those extracted in the tutorial phase, contributes to students’ performance score for the ‘computational problem solving’ component of the competency model. Part of the composite performance score derives from the number of sub-tasks that students are able to successfully complete; IRT scaling will account for the level of difficulty of the sub-tasks. The complex open-ended items are expected to be scored as polytomous items, with the score taking into account not only the correctness of the program (i.e. workable solution for a sub-task, or number of stones collected in the complex final item) but also the quality of student-generated programs in terms of their code complexity, structure (e.g. functional decomposition), and generalisation.

Assessment of self-regulation processes

76. Students’ self-regulatory processes will be assessed by looking at the choices that they make throughout the assessment task (as reflected in their actions in the environment), as well as by using data from the situated self-report items.

77. Process data can be used to construct indicators for the ‘metacognition and cognitive strategies’ component of the competency model. For example, asking for feedback or running code to see whether

a solution works or not could be interpreted as evidence of a student engaging in metacognition. Process-based measures will also be used to assess whether students apply certain cognitive strategies, and what kind of strategies they apply. This can be evaluated based on the changes made between subsequent programs that are being tested (number of blocks, abstract syntax tree, etc.), and the sequence of constructing and testing programs. A particularly interesting use of this data is to analyse debugging strategies employed when a program is wrong or doesn't work (e.g. stepping through code, adding and removing blocks after a run). For example, unproductive debugging strategies can involve running the same non-functioning program multiple times without changes, developing and modifying a program without testing it in-between changes, or changing too many parts of an unsuccessful program without changing it for the better. Analysing process data can also indicate, for example, whether students followed up on feedback and/or worked examples that they were shown. Additionally, situated self-report items could contribute to measuring this 'metacognition and cognitive strategies', for example asking students to assess their ability to complete the task before they begin working.

78. Finally, 'motivation and affect' can be assessed by using students' responses to situated self-report items directly before beginning work on a task, during, or directly after finishing the task. Questions can ask students to report how they feel after challenge events (e.g. debugging a faulty program) or their level of interest in engaging with different types of tasks to infer how they regulate their motivational and affective states. This component can also be complemented by process data measures. For example, analysing the time stamps of activities can identify those students who maintain effort throughout the sub-tasks, especially after experiencing some type of difficulty (i.e. avoiding extended periods of non-activity).

ANNEX C: PISA 2025 Learning in the Digital World Expert Group

Table C1: Members of the PISA 2025 Learning in the Digital World Expert Group

Members	University/Institute
Biswas, Gautam	Vanderbilt University
Bumbacher, Engin	Stanford University
Goldhammer, Frank	DIPF, Leibniz Institute for Research and Information in Education
Järvelä, Sanna	University of Oulu
Piech, Chris	Stanford University
Roll, Ido	Israeli Institute of Technology