CHAPTER 4:

Students' computational thinking

Chapter highlights

Computational thinking (CT) achievement can be described as increasing according to the following progression:

- At the lower region of the scale, students demonstrate a functional working knowledge of computation as input and output. They record data from observed outputs and implement complete solutions to simple coding problems.
- At the middle region of the scale, students demonstrate an understanding of computation as enabling practical solutions to real-world problems. They systematically associate inputs with outputs when planning solutions, and implement complete solutions to complex coding using non-linear logic.
- At the upper region of the scale, students demonstrate an understanding of computation as a generalizable problem-solving framework. They infer the relationship between observed inputs and outputs to evaluate solutions. They implement elegant and efficient solutions to complex coding problems using repeat and conditional statements.

Eight countries and one benchmarking participant completed the ICILS CT option. Students' CT varied more within countries than across countries.

- The range between the lowest five percent and the highest 95 percent of students' CT scores within countries varied between 266 scale points (in Portugal) and 371 scale points (in Korea). (Table 4.1)
- The difference between the highest and lowest average CT scores across countries was 76 scale points. (Table 4.1)

CT achievement tended to be higher among male students.

- Across all countries the average CT scale scores of male students was statistically significantly higher than that of female students. (Table 4.2)
- However, statistically significant differences in the average CT scale scores between female and male students were found in only two countries. In one of those countries the difference was in favor of female students and in the other it was in favor of male students. (Table 4.2)

Socioeconomic status (SES), denoted by parental occupation, parental education, and number of books in the home, was significantly positively associated with student CT.

- In all countries, students in the high SES groups scored significantly higher on the CT scale than those in the lower SES groups. (Table 4.3)

Immigrant background and language background were associated with student CT achievement.

- In six of seven countries, students from non-immigrant families had statistically significantly higher CT scale scores than students from immigrant families. (Table 4.4)
- In five of seven countries, students who reported mainly speaking the language of the ICILS test at home had statistically significantly higher CT scale scores than those who reported speaking another language at home. (Table 4.4)

J. Fraillon et al., *Preparing for Life in a Digital World*,
https://doi.org/10.1007/978-3-030-38781-5_4

Access to computers at home and years' experience using computers were associated with student CT.

- In all countries, students who reported having two or more computers at home had statistically significantly higher CT scores than students who reported having fewer than two computers at home. (Table 4.5)
- In all countries, students who reported having five years or more experience using computers had statistically significantly higher CT scale scores than those who reported having less than five years' experience. (Table 4.5)

Student CT achievement was strongly associated with student computer and information literacy (CIL) achievement.

- On average across all countries, the correlation between students' CIL and CT scale scores was 0.82. (Table 4.6)
- The correlation between students' CIL and CT scale scores varied between 0.74 and 0.89 across countries. (Table 4.6)

Introduction

The International Computer and Information Literacy Study (ICILS) 2018 assessment framework defines computational thinking (CT) as "an individual's ability to recognize aspects of real-world problems which are appropriate for computational formulation and to evaluate and develop algorithmic solutions to those problems so that the solutions could be operationalized with a computer" (Fraillon et al. 2019, p. 27).

According to the ICILS 2018 framework, CT comprises two strands, each of which is specified in terms of a number of aspects. The strands are: *conceptualizing problems* and *operationalizing solutions*. The aspects further articulate CT in terms of the main processes applied within each strand. The three aspects that make up the conceptualizing problems strand are: knowing about and understanding digital systems, formulating and analyzing problems, and collecting and representing relevant data. The two aspects that make up the operationalizing solutions strand are: planning and evaluating solutions, and developing algorithms, programs, and interfaces.

The ICILS 2018 CT assessment was an option for countries participating in ICILS. Eight countries and one benchmarking participant participated in the optional CT assessment. In this chapter, we detail the measurement of CT in ICILS 2018 and discuss student achievement across the countries that participated in the ICILS CT option. We begin the chapter by describing the CT assessment instrument and the proficiency scale derived from the ICILS 2018 test instrument and data. We also describe and discuss the international student results relating to CT. The majority of content in this chapter relates to Research Question CT 1, which focuses on the extent of variation existing among and within countries with respect to student CT. In the final sections of the chapter we address aspects of Research Questions CT 3 (the relationships between students' levels of access to, familiarity with, and self-reported proficiency in using computers and their CT), CT 4 (aspects of students' personal and social backgrounds, such as gender and socioeconomic background, related to students' CT), and CT 5 (the association between students computer and information literacy [CIL] and CT).

Assessing CT

The CT test instrument comprised two 25-minute test modules. In countries participating in the CT option, students completed the two CT test modules in randomized order after they had completed the CIL test and the student questionnaire.

One of the CT test modules (automated bus) focused on CT Strand 1: Conceptualizing problems, and the second (farm drone) focused on CT Strand 2: Operationalizing solutions. The automated bus module comprised a set of discrete, thematically related tasks relating to the configuration of the navigation and braking systems in a driverless bus. The farm drone module provided a visual coding interface that students were required to use to complete discrete coding tasks. The code in each task controlled the behavior of a programmable drone that could complete a set of actions on a farm. Students were presented with a work space, draggable commands, and a visual output that showed the outcomes of the executed commands. The complexity of each task related to the number of targets and actions required to solve the problem instance.

Data collected from the two CT test modules were used to measure and describe CT proficiency. In total, the data comprised 39 score points derived from 18 discrete tasks and questions. Student responses to most tasks were automatically scored. The exceptions were some open-response questions that were scored by trained expert scorers in each country. Data were only included where they met or exceeded the International Association for the Evaluation of Educational Achievement (IEA) sample participation requirements. The ICILS 2018 technical report (Fraillon et al. 2020) provides further information on adjudication of the test data.

The two strands of the ICILS CT framework are each specified in terms of several aspects. The strands describe CT in terms of conceptualizing problems and operationalizing solutions. The aspects further articulate CT in terms of the main (but not exclusive) constituent processes. We used this structure primarily as an organizational tool when describing the breadth of content of the CT construct. The structure was not intended to form the basis of analysis and reporting of achievement by sub-dimensions (such as by strand or aspect).

The following list sets out the two strands and corresponding aspects of the CT framework together with the percentages of score points (of the 39 total score points) attributed to each strand and to each aspect within the strands.

- Strand 1: Conceptualizing problems, comprising three aspects, 41 percent:
  - Aspect 1.1: Knowing about and understanding digital systems, 18 percent.
  - Aspect 1.2: Formulating and analyzing problems, 10 percent.
  - Aspect 1.3: Collecting and representing relevant data, 13 percent.
- Strand 2: Operationalizing solutions, comprising two aspects, 59 percent:
  - Aspect 2.1: Planning and evaluating solutions, 31 percent.
  - Aspect 2.2: Developing algorithms, programs, and interfaces, 28 percent.

The CT achievement scale

The structure described for the CT construct (two strands comprising two and three respective aspects) was established to "allow readers to clearly see the different related aspects of CT and to support the auditing of the CT instruments against the full breadth of content in the CT construct" (Fraillon et al. 2019, p. 28). As mentioned, this described structure did not presuppose a sub-dimensional structure for the analysis and reporting of the CT construct. For ICILS 2018 a single scale of CT achievement has been established and described. Further exploration of the potential of sub-dimensions of CT to be reported are planned for future cycles of ICILS.

We used the Rasch item response theory (IRT) model (Rasch 1960) to derive the CT achievement from the 39 score points obtained from the 18 CT tasks. We set the final reporting scale to a metric that had an international mean of 500 (the ICILS average score) and standard deviation of 100 for the equally weighted national samples. We used plausible value methodology with full conditioning to derive summary student achievement statistics. This approach enables estimation of the uncertainty inherent in a measurement process (e.g., von Davier et al. 2009). The ICILS technical report provides details on the procedures the study used to scale test items (Fraillon et al. 2020).

Description of the ICILS CT scale is based on the content and scaled difficulties of the assessment items. As part of the test development process, the ICILS research team wrote descriptors for each item in the assessment instrument. These item descriptors, which also reference the ICILS assessment framework, describe the CT knowledge, skills, and understandings demonstrated by a student correctly responding to each item. An item map similar to the item map for CIL was produced for CT.

In order to describe the underlying characteristics of achievement across the breadth of the scale we divided the items that were ordered in the item map into thirds with equal numbers of items in each third. For ICILS 2018 we refer to these as the lower, middle, and upper regions of the scale. The descriptions of each region are syntheses of the common elements of CT knowledge, skills, and understanding described by the items within each region.[18] The regions

18 The lower and upper regions are unbounded. The descriptions for these regions are based on items with a scaled difficulty that are within a range of 130 scale points below 459 scale points (for the lower region) or above 589 scale points (for the upper region).

of the CT scale should not be directly compared to the levels in the CIL scale, as they have been developed using a different process and the scale metrics are not comparable.

The lower region of the CT scale was defined as the region below 459 scale points, the middle region is that between 459 and 589 scale points (inclusive), and the upper region is above 589 scale points.

The scale is hierarchical in the sense that CT proficiency becomes more sophisticated as student achievement progresses up the scale. We can therefore assume that a student located at a particular place on the scale because of their achievement score will be able to undertake and successfully accomplish tasks up to that level of achievement. Following is a description of the characteristics of each region on the CT scale.

Lower region (below 459 scale points)

Students showing achievement corresponding to the lower region of the scale demonstrate familiarity with the basic conventions of digital systems to configure inputs, observe events, and record observations when planning computational solutions to given problems. When developing problem solutions in the form of algorithms, they can use a linear (step by step) sequence of instructions to meet task objectives.

Students working at the lower region of the scale can, for example:

- Create a complete but suboptimal route from one location to another on a network diagram;
- Partially debug an algorithm that uses a repeat statement by correcting the logic of connected statements;
- Create an efficient algorithm that meets all of the given task objectives for a low-complexity problem (i.e., a problem with a limited set of available commands and objectives); and
- Create an inefficient algorithm that meets all of the given task objectives for a medium-complexity problem (e.g., a problem with multiple objectives best solved using a repeat statement).

Middle region (459 to 589 scale points)

Students showing achievement corresponding to the middle region of the scale demonstrate understanding of how computation can be used to solve real-world problems. They can plan and execute systematic interactions with a system so that they can interpret the output or behavior of the system. When developing algorithms, they use repeat statements effectively.

Students working in the middle region of the scale can for example:

- Adapt information shown in a network diagram to create a complete set of instructions comprising at least five steps;
- Configure a simulation tool;
- Store and compare data collected using a simulation tool;
- Debug, with some redundancy in the solution, an algorithm for a high-complexity problem (e.g., a problem with multiple task objectives best solved using repeat and conditional statements);
- Create an efficient algorithm that meets all of the objectives for a medium-complexity problem (e.g., a problem with multiple objectives best solved using a repeat statement); and
- Create an inefficient algorithm that meets all of the objectives for a high-complexity problem (e.g., a problem with multiple task objectives best solved using repeat and conditional statements).

Upper region (above 589 scale points)

Students showing achievement corresponding to the upper region of the scale demonstrate an understanding of computation as a generalizable problem-solving framework. They can explain how they have executed a systematic approach when using computation to solve real-world problems. Furthermore, students operating within the upper region can develop algorithms that use repeat statements together with conditional statements effectively.

Students working in the middle region of the scale can, for example:

- Explain the value of a digital system for real-world problem solving;
- Complete a simple decision tree with the correct use of both logic and syntax;
- Debug, with the most efficient solution, an algorithm for a high-complexity problem (e.g., a problem with multiple task objectives best solved using repeat and conditional statements); and
- Create an efficient algorithm that meets all of the objectives for a high-complexity problem (e.g., multiple task objectives best solved using repeat and conditional statements).

Example CT tasks

To provide a clearer understanding of the nature of the CT scale, we include in this section of the chapter a set of example tasks. These indicate the types and range of tasks that students were required to complete during the ICILS test of CT. The items also provide examples of responses corresponding to the different regions on the CT scale. The data for each example task included in the analysis (including calculation of the ICILS 2018 average) are drawn only from those countries that met the sample participation, test administration, and coding requirements for that task in ICILS 2018. The example tasks are drawn from each of the two CT test modules.

Farm drone tasks (Example Tasks 1 and 2)

In the farm drone module, students worked within a simple visual coding environment (students had access to drag and drop code blocks each of which performed a specified function) to create, test, and debug code that controls the actions of a drone used in a farming context. The difficulties of the tasks relate to the code functions that were available and the complexity of the sequence of actions required by the drone to complete the task. Students' responses were captured by the assessment system and later scored on the basis of following two characteristics:

The "correctness" with which the drone performs the actions specified in the task. This includes both the degree to which the drone performs required actions and the presence or absence of any unrequired actions.

The "efficiency" of the code. This was measured by comparing the number of code blocks used in the solution with the minimum number required to implement a fully correct solution (with longer code sequences corresponding to lower scores). Each farm drone task included an instruction for students to use as few code blocks as possible.

Ultimately, each coding task received a single score derived by combining the correctness and efficiency scores. For most tasks, the efficiency score was used to moderate the score attributed to completely correct responses. Full details of the scoring for each farm drone coding task are provided in the ICILS 2018 technical report (Fraillon et al. 2020).

The interface design for the Example CT Task 1 module was divided into two functional spaces (Figure 4.1). The test interface space (using the right and bottom of the screen) was the same as that used for the CIL test modules. Unlike in the other CIL and CT test modules, in the farm drone module students could return to previously completed tasks by clicking on the green task box corresponding to the ordinal position of the task. Students could also use a flag toggle to mark tasks that they wanted to go back to if they had sufficient time to review and improve their solutions. The stimulus area comprised three separate parts: the code blocks space (at the bottom left of the screen), the farm drone display space (the 9 × 9 grid at the top left), and the work space (the central space where code blocks could be arranged to form an algorithm).

All tasks in the farm drone module presented students with the same interface design, with variations in the configuration of the farm, the task objectives, the available code block functions, and the state of the work space. The work space was presented as empty (with only the fixed "when run" command present) for tasks that required students to create code sequences. The work space was presented with pre-populated algorithms for tasks that required students to debug code.

Students could drag code blocks into the work space. Code blocks connected to the "when run" code block would send instructions to the drone when the green "run program" button was clicked. Students could reset the state of the drone and the farm by clicking the blue reset button. They could also reset the state of the work space by clicking the orange reset button.

The complexity of the tasks increased progressively through the farm drone module. The complexity of each task is influenced by the following set of key characteristics:

- The task type (code creation or debugging);
- The variety of available code functions (movement, action, repeat, conditional);
- The number of targets (a target is a tile requiring a specific action to be completed over it such as dropping water, seed, or fertilizer);
- The number of different target types (dirt, low, or high crops);
- Whether or not any given target required more than one action to be completed over it;
- The layout configuration of the targets (single or multiple rows); and
- The number of different materials to be dropped on targets (water, seed, fertilizer).

Figure 4.1: Example CT Task 1 with framework references and overall percent correct

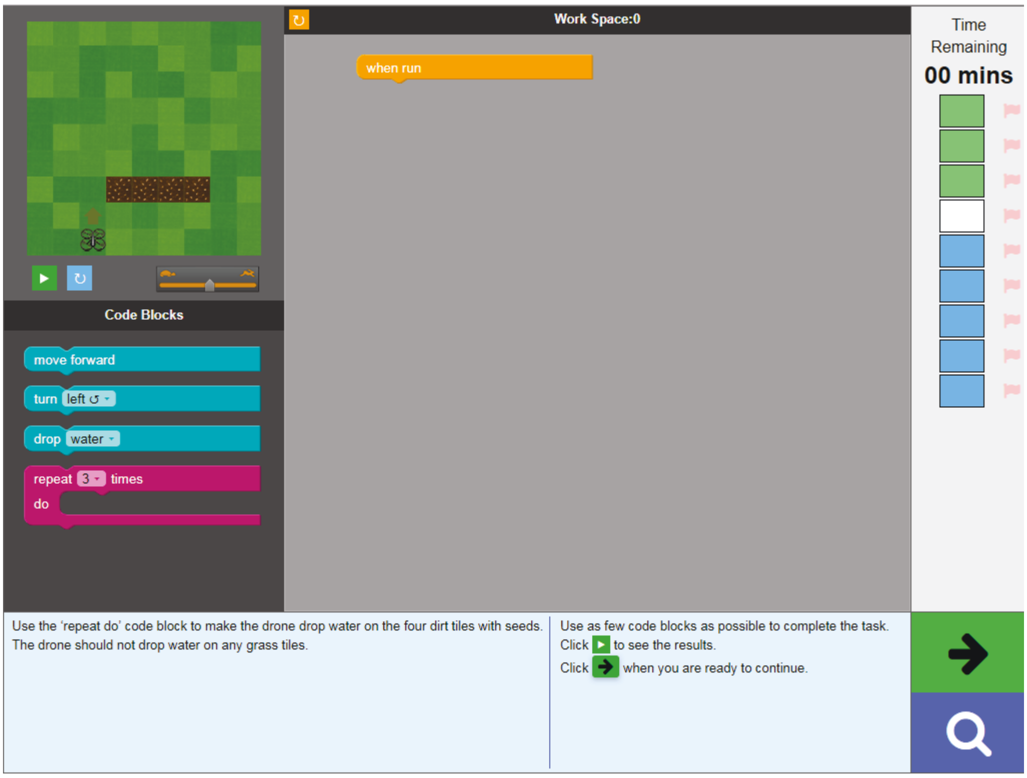


Figure 4.1: Example CT Task 1 with framework references and overall percent correct (contd.)

Score	CT scale region	CT scale difficulty	ICILS 2018 average percentage correct responses
At least one of three points	Lower	353	86 (0.3)
At least two of three points	Lower	396	77 (0.4)
Three points	Upper	613	27 (0.5)
ICILS assessment framework reference			
2.2	Operationalizing solutions		
	Developing algorithms, programs, and interfaces		

Country	Percentage scoring one out of three points	Percentage scoring two out of three points	Percentage scoring three points
Denmark† 1	92 (0.5)	83 (0.9)	26 (1.2)
Finland	87 (1.0)	80 (1.2)	29 (1.0)
France	87 (0.8)	77 (1.0)	40 (1.3)
Germany	83 (1.2)	73 (1.2)	18 (1.2)
Korea, Republic of	90 (0.8)	86 (1.0)	39 (1.9)
Luxembourg	76 (0.5)	66 (0.5)	16 (0.3)
Portugal†† 1	88 (0.8)	78 (1.1)	20 (1.1)
Not meeting sample participation requirements			
United States	86 (0.6)	77 (0.7)	34 (1.1)
Benchmarking participant meeting sample participation requirements			
North Rhine-Westphalia (Germany)	84 (0.9)	73 (1.3)	17 (1.1)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

1 National defined population covers 90% to 95% of the national target population.

Example CT Task 1 (Figure 4.1) is a medium-complexity code creation task that represents the operationalizing solutions strand of the CT construct. The task objectives required students to make the drone drop water on the four dirt tiles with seeds (the targets) without dropping water on any of the grass tiles using the repeat statement.

Students could receive a score of zero, one, two, or three points on this task. Students who completed only some of the task objectives using significantly more code blocks than the minimum necessary were awarded one score point; on average across all countries, 86 percent of students achieved at least this. The percentages across countries and the benchmarking participant varied from 76 percent to 92 percent. Achievement of a score of one on this task was at the lower end of the lower region on the CT scale.

Students who completed all of the objectives and used the repeat statement but included a few code blocks more than the minimum necessary were awarded two score points; on average across all countries, 77 percent of students achieved at least this. The percentages across countries and benchmarking participants varied from 73 percent to 86 percent. Achievement of a score of two on this task was in the upper end of the lower region of the CT scale.

Students who could complete all of the task objectives using the repeat statement and with the fewest number of code blocks necessary were awarded the maximum of three score points; on average across all countries, 27 percent of students received the maximum score in this task. The percentages across countries and the benchmarking participant varied from 16 percent to 40 percent. Achievement of a score of three on this task was in the upper region of the CT scale.

Example CT Task 2 (Figure 4.2) is a high-complexity debugging task. The work space was pre-populated with a five-statement algorithm that students had to modify to complete the task objectives.

Figure 4.2: Example CT Task 2 with framework references and overall percent correct

Score	CT scale region	CT scale difficulty	ICILS 2018 average percentage correct responses
At least one of three points	Lower	456	63 (0.5)
At least two of three points	Middle	552	37 (0.5)
Three points	Upper	733	8 (0.2)
ICILS assessment framework reference			
2.1	Operationalizing solutions		
	Planning and evaluating solutions		

Figure 4.2: Example CT Task 2 with framework references and overall percent correct (contd.)

Country	Percentage scoring one out of three points	Percentage scoring two out of three points	Percentage scoring three points
Denmark† 1	70 (1.3)	39 (1.2)	8 (0.7)
Finland	66 (1.3)	44 (1.3)	9 (0.7)
France	65 (1.2)	41 (1.1)	12 (0.7)
Germany	60 (1.4)	32 (1.2)	7 (0.6)
Korea, Republic of	74 (1.4)	48 (1.6)	12 (0.8)
Luxembourg	48 (0.6)	25 (0.4)	5 (0.3)
Portugal†† 1	61 (1.6)	31 (1.3)	3 (0.4)
Not meeting sample participation requirements			
United States	60 (1.1)	37 (0.9)	8 (0.4)
Benchmarking participant meeting sample participation requirements			
North Rhine-Westphalia (Germany)	61 (1.3)	32 (1.3)	6 (0.6)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

1 National defined population covers 90% to 95% of the national target population.

Example CT Task 2 represents the operationalizing solutions strand of the CT construct and required students to make the drone drop water on the big and small crop tiles and in addition to drop fertilizer on the small crop tiles. The algorithm presented to students in the work space used an *if* statement nested inside a *repeat* statement which included non-linear conditional logic. In the logic of the existing algorithm, the decision to drop fertilizer and water was conditional on the size of the crop. In the simplest correction to the algorithm, students could place the "drop water" command outside the conditional statement after the "move forward" command and reconfigure the conditional command to "if small crop."

Students could receive a score of zero, one, two, or three points on this task. Students who could complete all of the objectives but with many more code blocks than the minimum necessary were awarded one score point. This was typically achieved by removing the *repeat* and *if* statements and using the *move* and *drop* statements using only linear logic. On average across all countries, 63 percent of students achieved at least this. The percentages across countries and the benchmarking participant varied from 48 to 74 percent. Achievement of a score of one on this task was in the lower region on the CT scale.

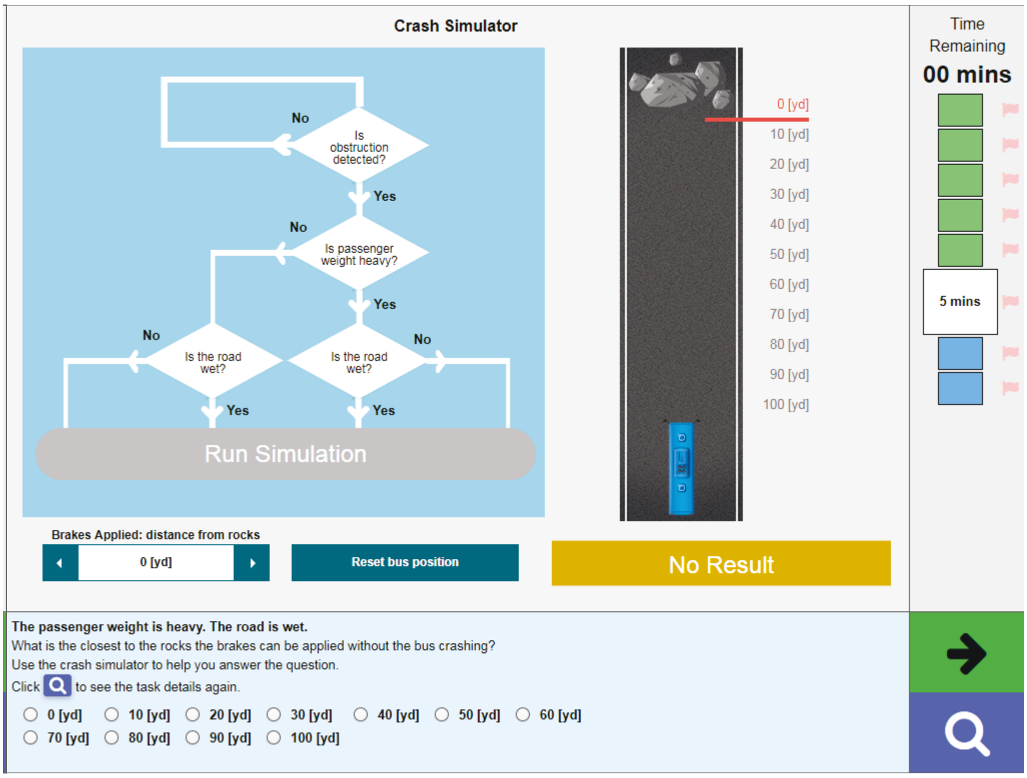
Students who could complete all of the task objectives with only a few code blocks more than the minimum necessary, by using the *repeat* and *if* statements together, were awarded a score of two. On average across all countries, 37 percent of students achieved at least this. The percentages across countries and the benchmarking participant varied from 25 percent to 48 percent. Achievement of a score of two on this task was in the middle region of the CT scale.

Students who were able to correct the algorithm using the minimum number of code blocks were awarded a score of three. These students demonstrated clear control over the non-linear conditional logic of the algorithm. On average across all countries, eight percent of students received the maximum score on this task. The percentages of students who were successful across countries and the benchmarking participant varied from three percent to 12 percent. Achievement of a score of three on this task was in the upper region of the CT scale.

Automated bus tasks (Example Tasks 3 and 4)

Example CT Task 3 (Figure 4.3) represents the conceptualizing problems strand of the CT construct. The task presented students with a stopping distance simulator and the objective of finding the minimum viable braking distance under given conditions. Students were required to configure the flowchart according to the instructions to apply a set of conditions for the simulation. They then had to configure the braking distance and run the simulation to identify whether the bus stopped before, or crashed into, the rocks.

Figure 4.3: Example CT Task 3 with framework references and overall percent correct



Score	CT scale region	CT scale difficulty	ICILS 2018 average percentage correct responses
At least one of two points	Middle	477	58 (0.4)
Two points	Middle	557	36 (0.4)
ICILS assessment framework reference			
1.3	Conceptualizing problems		
	Collecting and representing relevant data		

Figure 4.3: Example CT Task 3 with framework references and overall percent correct (contd.)

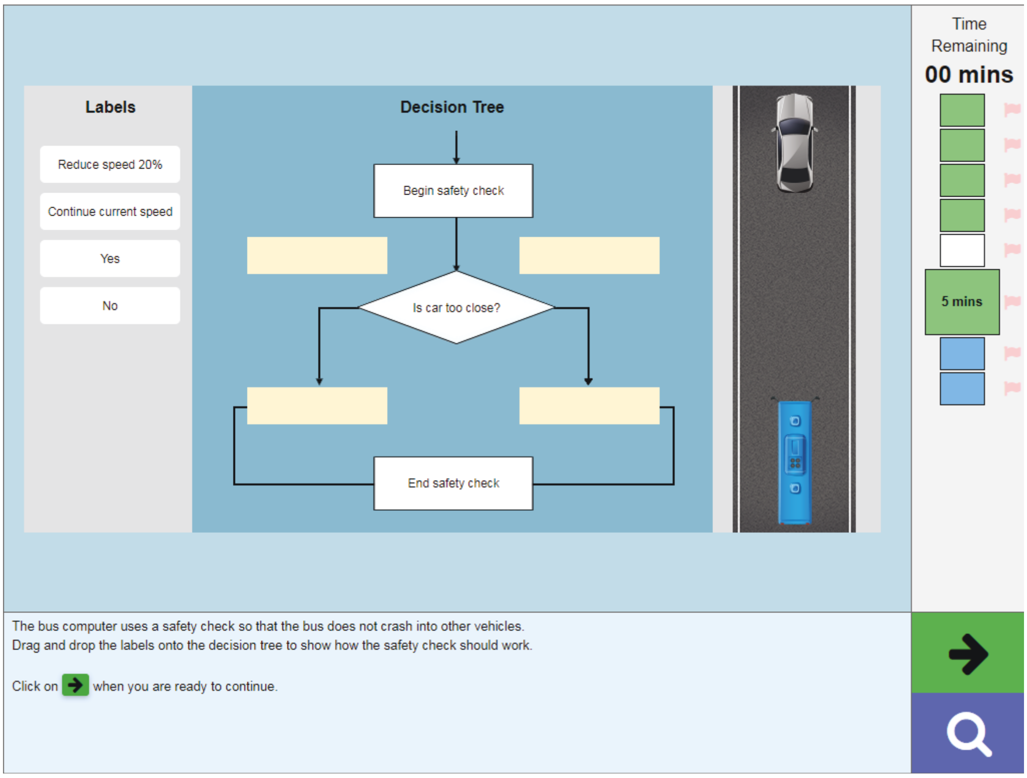
Country	Percentage scoring one out of two points	Percentage scoring two points
Denmark† 1	64 (1.3)	40 (1.3)
Finland	62 (1.3)	37 (1.3)
France	48 (1.1)	27 (1.1)
Germany	56 (1.2)	32 (1.1)
Korea, Republic of	72 (1.2)	58 (1.2)
Luxembourg	50 (0.6)	28 (0.5)
Portugal†† 1	56 (1.4)	28 (1.4)
Not meeting sample participation requirements		
United States	57 (1.0)	34 (0.9)
Benchmarking participant meeting sample participation requirements		
North Rhine-Westphalia (Germany)	55 (1.2)	29 (1.3)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.
† Met guidelines for sampling participation rates only after replacement schools were included.
†† Nearly met guidelines for sampling participation rates after replacement schools were included.
1 National defined population covers 90% to 95% of the national target population.

Students could receive a score of zero, one, or two points on this task. One point was awarded to students who configured the simulator using conditions that were not specified but identified a braking distance that was consistent with the configuration they used. Students who configured the simulator correctly to the conditions and identified the correct braking distance were awarded two score points. Scores of one and two on this task were both indicative of achievement in the middle region of the CT scale, with a score of one near the lower end of the region and a score of two at the upper end of the region. On average across all countries, 58 percent of students achieved a score of at least one on this task and 36 percent of students achieved a score of two. These percentages varied across countries and the benchmarking participant between 50 percent and 72 percent for a score of at least one and between 27 percent and 58 percent for a score of two.

In the CT test, Example CT Task 4 (Figure 4.4) was presented to students as the task preceding Example CT Task 3 above. Example CT Task 4 was a planning task in preparation for configuring the simulator and required students have some understanding of flowcharting conventions to complete the decision tree with the labels provided. Students were required to drag and drop the labels from the left of the screen into the decision tree in a way that was consistent with both the logical sequence of the decision-making and the syntax of the decision tree.

Figure 4.4: Example CT Task 4 with framework references and overall percent correct



Score	CT scale region	CT scale difficulty	ICILS 2018 average percentage correct responses
At least one of two points	Middle	488	56 (0.4)
Two points	Upper	591	28 (0.4)
ICILS assessment framework reference			
1.2	Conceptualizing problems		
	Formulating and analyzing problems		

Figure 4.4: Example CT Task 4 with framework references and overall percent correct (contd.)

Country	Percentage scoring one out of two points	Percentage scoring two points
Denmark† 1	55 (1.3)	29 (1.0)
Finland	56 (1.2)	27 (1.0)
France	55 (1.4)	30 (1.4)
Germany	57 (1.1)	30 (1.0)
Korea, Republic of	65 (1.2)	37 (1.3)
Luxembourg	51 (0.6)	24 (0.6)
Portugal†† 1	50 (1.2)	20 (1.0)
Not meeting sample participation requirements		
United States	52 (0.7)	24 (0.6)
Benchmarking participant meeting sample participation requirements		
North Rhine-Westphalia (Germany)	55 (1.1)	29 (1.0)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent.
† Met guidelines for sampling participation rates only after replacement schools were included.
†† Nearly met guidelines for sampling participation rates after replacement schools were included.
1 National defined population covers 90% to 95% of the national target population.

Students could receive a score of zero, one, or two points on this task. Students who could use the syntax correctly by dragging the "Yes" and "No" labels to the spaces above the decision point but placed the "Reduce speed 20%" and "Continue current speed" labels under the wrong decision received one score point. Similarly, students who used the syntax incorrectly by placing the "Yes" and "No" labels in the spaces below the decision point but placed the "Reduce speed 20%" and "Continue current speed" labels under the correct decision also received one score point. In summary, students who could complete the decision tree with either the correct syntax or the correct logic received one score point. This was indicative of achievement in the middle region of the CT scale. Students who could complete the decision tree with the correct syntax and the correct logic received two score points. This was indicative of achievement in the upper region of the CT scale. On average across all countries, 56 percent of students achieved a score of at least one on this task and 28 percent of students achieved a score of two. These percentages varied across countries and the benchmarking participant between 50 percent and 65 percent for a score of at least one and between 20 percent and 37 percent for a score of two.

Comparison of CT across countries

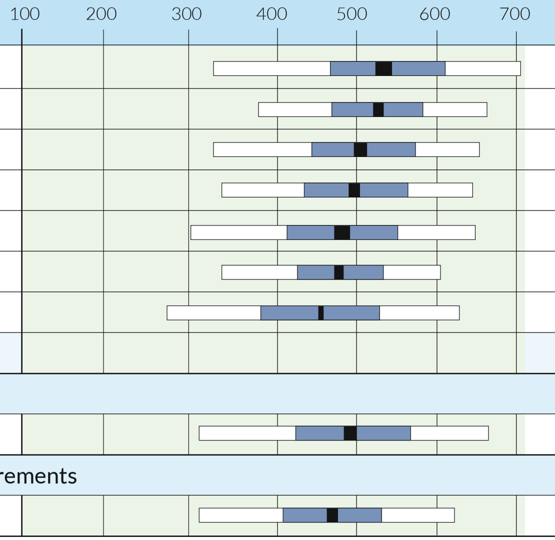
Distribution of student achievement scores

When considering the distribution of student achievement on the CT test for all participating countries and benchmarking participants, it is important to bear in mind that only a small number of countries participated in the ICILS CT international option (Table 4.1; note that the length of the bars in these tables reflect the spread of student scores within each country, highlighting differences in the within-country student score distributions).

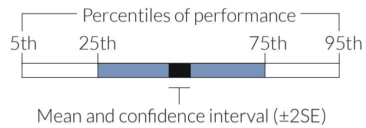
The average country scores on the CIL scale ranged between 460 and 536 scale points; the range was equivalent to approximately 0.7 international standard deviations (Table 4.1). Unlike the distribution of CIL, for CT the spread of scores within countries does not appear to be clearly associated with achievement across countries (Table 4.1). However, as was observed for student CIL, the variation in student CT scores within countries was greater than that across countries. Across countries, the distance between the lowest five percent and the highest five percent of

Table 4.1: Country averages for CT, average age, CT score, ICT development index score, and percentile graph

Country	Average age	CT achievement distribution	Average CT score		ICT development index (IDI) score (and country rank)
Korea, Republic of	14.2		536 (4.4)	▲	8.85 (2)
Denmark† 1	14.9		527 (2.3)	▲	8.71 (4)
Finland	14.8		508 (3.4)	▲	7.88 (22)
France	13.8		501 (2.4)		8.24 (15)
Germany	14.5		486 (3.6)	▼	8.39 (12)
Portugal†† 1	14.1		482 (2.5)	▼	7.13 (44)
Luxembourg	14.5		460 (0.9)	▼	8.47 (9)
ICILS 2018 average	14.4		500 (1.1)		
Not meeting sample participation requirements					
United States	14.2		498 (2.5)		8.18 (16)
Benchmarking participant meeting sample participation requirements					
North Rhine-Westphalia (Germany)	14.4		485 (3.0)	▼	8.39 (12) [2]



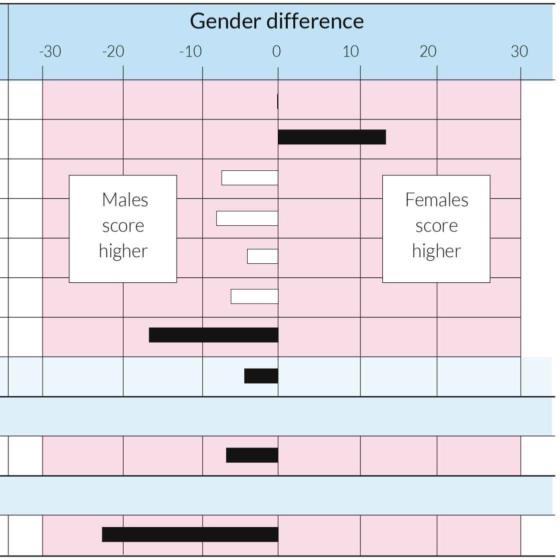
▲ Achievement significantly higher than ICILS 2018 average
▼ Achievement significantly lower than ICILS 2018 average



Notes: ICT development index (IDI) score and country rank data relate to 2017 (source: ITU 2019). Standard errors appear in parentheses.
† Met guidelines for sampling participation rates only after replacement schools were included.
†† Nearly met guidelines for sampling participation rates after replacement schools were included.
1 National defined population covers 90% to 95% of national target population.
2 Data relate to all of Germany.

Table 4.2: Gender differences in CT

Country	Mean scale score females	Mean scale score males	Difference (females – males)	Gender difference
Denmark† 1	527 (2.7)	527 (3.1)	0 (3.5)	
Finland	515 (3.7)	502 (4.3)	**13** (4.4)	
France	498 (3.1)	505 (3.0)	−7 (3.8)	
Germany	482 (3.7)	490 (4.7)	−8 (4.4)	
Korea, Republic of	534 (4.6)	538 (5.5)	−4 (4.9)	
Luxembourg	457 (2.0)	463 (1.7)	−6 (3.3)	
Portugal†† 1	473 (2.7)	490 (3.3)	**−16** (3.3)	
ICILS 2018 average	498 (1.2)	502 (1.4)	**−4** (1.5)	
Not meeting sample participation requirements				
United States	495 (2.6)	502 (3.3)	**−7** (3.1)	
Benchmarking participant meeting sample participation requirements				
North Rhine-Westphalia (Germany)	474 (3.4)	496 (4.1)	**−23** (4.8)	



■ Gender difference statistically significant at $p < 0.05$ level
□ Gender difference not statistically significant

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent. Statistically significant differences ($p < 0.05$) between subgroups are shown in **bold**.
† Met guidelines for sampling participation rates only after replacement schools were included.
†† Nearly met guidelines for sampling participation rates after replacement schools were included.
1 National defined population covers 90% to 95% of national target population.

CT scores ranges between 266 and 371 scale points (with a median of 320 scale points or 3.2 standard deviations), in comparison to a range of average scores across all countries of 76 scale points (equivalent to three quarters of an international standard deviation).

CT relative to the ICT development index

In Chapter 3 we reported that, on average, students had higher levels of CIL in countries with higher ICT development index (IDI) scores. The association between CT achievement and the IDI scores across countries was moderate, with a Pearson's correlation coefficient of 0.43. However, the association between student CIL and country IDI in the same countries (i.e., only those countries in which students completed the tests of both CIL and CT) was weak, with a Pearson's correlation coefficient of 0.23. A comparison of these correlation coefficients suggest that, across the countries taking part in the ICILS CT assessment, the association between the IDI and a broader notion of digital competence (CIL) was lower than one might have predicted. This is potentially due to this smaller set of countries being relatively more homogeneous with respect to IDI and achievement than the broader set of ICILS 2018 participating countries.

Variation in CT across countries with respect to student background characteristics

In this section we address Research Question CT 4: *What aspects of students' personal and social backgrounds (such as gender and socioeconomic background) are related to students' CT?*

Our focus is therefore on the associations between students' CT and student gender, variables associated with students' socioeconomic status, whether or not students had an immigrant background, and the language students spoke at home. Chapter 7 documents further investigation, based on regression modeling, of the relationships between student CT and student-level and school-level factors.

Gender and CT

In Chapter 3 we reported that the CIL scale scores of female students were statistically significantly higher than those of male students on average across all countries and within all countries and benchmarking participants except Chile, Uruguay, and North Rhine-Westphalia (Germany). A different pattern was evident for the relationship between gender and CT achievement to that of gender and CIL achievement. Across all countries the average CT scale score of male students was statistically significantly higher than that of female students (Table 4.2). However, this difference was not consistent at the country level. In Portugal and North Rhine-Westphalia (Germany) the average achievement of male students was statistically significantly higher than that of female students and in Finland the CT achievement of female students was statistically significantly higher than that of male students. In France, Germany, Korea, and Luxembourg the average scores of male students appeared to be higher than those of female students but the differences were not statistically significant. In Chapter 8 we discuss the implications of the differences in achievement by gender found in both CIL and CT in further detail.

Socioeconomic background

In Chapter 3, details of how background data variables were collected and derived were presented. We reported statistically significant associations between each of the three socioeconomic background variables (parental occupation, parental education, and number of books in the home) and CIL across all countries.

We found a similar pattern of association between socioeconomic background and CT achievement across all countries (Table 4.3) to that reported for CIL in Chapter 3.

Table 4.3: Average CT by category of parental occupation, parental education, and number of books in the home

Country	Average CT scores by parental occupation (SEI)		Average CT scores by parental education (ISCED)		Average CT scores by books in the home	
	SEI below 50	SEI 50 and above	Short-cycle tertiary or below	Bachelor's degree or higher	Below 26	26 and above
	-100 -80 -60 -40 -20 0 20 40 60 80 100		-100 -80 -60 -40 -20 0 20 40 60 80 100		-100 -80 -60 -40 -20 0 20 40 60 80 100	
Denmark† 1	513 (3.4)	**539** (2.7)	516 (2.7)	**546** (3.8)	494 (4.3)	**541** (2.3)
Finland	493 (3.8)	**538** (3.6)	498 (4.0)	**520** (4.1)	475 (5.6)	**524** (3.0)
France	485 (3.0)	**531** (3.2)	493 (2.5)	**529** (4.4)	465 (3.5)	**528** (2.6)
Germany	470 (4.1)	**522** (4.8)	485 (4.3)	**515** (6.4)	431 (7.5)	**508** (3.6)
Korea, Republic of	530 (4.3)	**551** (5.3)	523 (4.9)	**542** (5.1)	495 (8.4)	**542** (4.4)
Luxembourg	438 (1.8)	**505** (2.0)	443 (1.6)	**489** (2.1)	407 (2.1)	**482** (1.4)
Portugal†† 1	466 (3.1)	**504** (3.1)	470 (2.8)	**507** (4.0)	458 (4.1)	**498** (2.5)
ICILS 2018 average	485 (1.3)	**527** (1.4)	490 (1.3)	**521** (1.7)	461 (2.1)	**517** (1.1)
Not meeting sample participation requirements						
United States	483 (2.8)	**527** (3.1)	473 (2.9)	**526** (3.0)	451 (3.3)	**525** (2.7)
Benchmarking participant meeting sample participation requirements						
North Rhine-Westphalia (Germany)	472 (3.8)	**518** (4.3)	485 (3.9)	**507** (5.8)	439 (6.0)	**507** (3.2)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent. Score averages that are significantly larger ($p < 0.05$) than those in the comparison group are shown in **bold**. SEI = socioeconomic index.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

1 National defined population covers 90% to 95% of the national target population.

Difference between comparison groups statistically significant at $p < 0.05$

Difference between comparison groups not statistically significant at $p < 0.05$

For each of the three socioeconomic background variables in each country, and overall across countries and the benchmarking participant, the average CT scores of students in the "higher" groups were statistically significantly higher than those of students in the "lower" groups. However, the magnitude of the differences between groups for all three variables varied across countries.

On average across all countries, the difference between students in the higher and lower parental occupation categories was 42 CT scale points, with the minimum difference being 21 scale points in Korea and the maximum difference being 67 scale points in Luxembourg.

On average across all countries, the difference between the CT scores of students in the low (short-cycle tertiary or below) and in the high (Bachelor's degree or higher) parental education groups was 31 scale points, with the minimum difference being 19 scale points in Korea and the maximum difference being 46 scale points in Luxembourg.

Cross-nationally, the difference between the average CT scale scores of students who reported having 26 or more books at home and those students who reported fewer than 26 books at home was 57 scale points, with the minimum difference being 40 scale points in Portugal and the maximum difference being 77 scale points in Germany.

All three indicators of students' socioeconomic status contributed to a composite index of socioeconomic status. This index is used in the multilevel regression analyses presented in Chapter 7.

Immigrant status and language use

In Chapter 3 we reported that the CIL scores of students without immigrant background tended to be higher than those with an immigrant background. We found similar results with CT where, for six of the seven countries that met the ICILS technical requirements, the students from non-immigrant family backgrounds had statistically significantly higher average CT scores than students from immigrant backgrounds (Table 4.4). On average across countries, the difference between students with immigrant backgrounds and those without was 46 CT scale points, with the minimum difference being 14 scale points in Portugal and the maximum difference being 56 in Finland.

In most participating countries, the majority of students indicated speaking the test language at home. Across countries, CT scores tended to be higher among students speaking the test language at home; the average difference was 47 scale points. For five of the seven participating countries meeting the technical requirements, we recorded statistically significant differences between students speaking the test language and those speaking other languages at home. The statistically significant positive differences varied from 38 scale points in Finland to 64 in France.

Computers at home and experience using computers

In Chapter 3 we noted that students with more computers at home tended to have higher CIL scores. A similar pattern of association can be seen when considering students' CT achievement (Table 4.5). On average across countries, the CT scores of students reporting having two or more computers at home were 29 scale points higher than those who reported having fewer than two computers at home. This difference varied from 18 points in Portugal to 43 points in Germany and was statistically significant in all countries.

Table 4.4: Average CT by category of immigrant background and language spoken at home

Country	Average CT by immigrant background		Average CT by language spoken at home	
	Immigrant family	Non-immigrant family	Other	Language of test
Denmark† [1]	481 (6.3)	**534** (2.4)	481 (8.8)	**531** (2.3)
Finland	457 (15.9)	**513** (3.3)	474 (10.0)	**512** (3.3)
France	459 (6.5)	**511** (2.2)	446 (9.2)	**510** (2.3)
Germany	454 (7.4)	**503** (3.7)	439 (6.9)	**502** (3.8)
Korea, Republic of	^ ^	538 (4.3)	^ ^	537 (4.3)
Luxembourg	450 (2.1)	**470** (1.9)	452 (1.1)	**496** (3.6)
Portugal†† [1]	470 (6.7)	**484** (2.7)	489 (10.5)	482 (2.5)
ICILS 2018 average	462 (3.5)	**508** (1.1)	463 (3.4)	**510** (1.2)
Not meeting sample participation requirements				
United States	471 (8.2)	**500** (2.6)	468 (5.1)	**502** (2.4)
Benchmarking participant meeting sample participation requirements				
North Rhine-Westphalia (Germany)	454 (5.4)	**503** (3.4)	439 (5.9)	**504** (3.3)

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent. Score averages that are significantly larger ($p < 0.05$) than those in the comparison group are shown in **bold**.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

[1] National defined population covers 90% to 95% of the national target population.

^ Number of students too small to report group averages.

Difference between comparison groups statistically significant at $p < 0.05$

Difference between comparison groups not statistically significant at $p < 0.05$

Table 4.5: Percentages by category of computer availability at home and years' experience of ICT use, and comparison of average CT between categories

Country	Average CT by computer availability at home		Average CT by years' experience of ICT use	
	Fewer than two computers	Two computers or more	Less than five years	Five years or more
Denmark† 1	501 (8.6)	**531** (2.3)	510 (3.1)	**541** (2.8)
Finland	490 (3.8)	**518** (3.5)	487 (5.0)	**522** (3.1)
France	486 (3.7)	**511** (2.4)	500 (2.8)	**508** (3.0)
Germany	461 (7.5)	**504** (3.7)	487 (4.7)	**498** (4.3)
Korea, Republic of	525 (5.0)	**546** (4.8)	512 (4.6)	**562** (5.2)
Luxembourg	430 (3.4)	**471** (1.2)	458 (1.4)	**468** (1.6)
Portugal†† 1	472 (4.1)	**490** (2.5)	470 (3.5)	**489** (2.8)
ICILS 2018 average	481 (2.1)	**510** (1.2)	489 (1.4)	**512** (1.3)
Not meeting sample participation requirements				
United States	465 (3.6)	**515** (2.7)	479 (2.8)	**517** (2.9)
Benchmarking participant meeting sample participation requirements				
North Rhine-Westphalia (Germany)	463 (5.4)	**502** (3.2)	488 (3.7)	493 (4.8)

Difference between comparison groups statistically significant at $p < 0.05$
Difference between comparison groups not statistically significant at $p < 0.05$

Notes: Standard errors appear in parentheses. Because results are rounded to the nearest whole number, some totals may appear inconsistent. Score averages that are significantly larger ($p < 0.05$) than those in the comparison group are shown in **bold**.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

1 National defined population covers 90% to 95% of the national target population.

Similarly, students' years of experience using computers was positively associated with CT. On average across all countries, the CT scores of students with five or more years of experience using computers were 23 scale points higher than those with less than five years' experience. This difference varied from eight scale points in France to 50 scale points in Korea and was statistically significant in all countries but not statistically significant in the benchmarking participant North Rhine-Westphalia (Germany).

The association between CT and CIL

In this section we address Research Question CT 5: *What is the association between students' CIL and CT?*

CIL has an operational emphasis on students' abilities to use computer technologies to collect and manage information and to produce and exchange information. In the ICILS test of CIL, this is assessed through students' responses to a broad range of tasks that focus on (receptive and productive) information literacy in a digital environment. CT is a new assessment construct to ICILS. Its focus is on the planning, formulation, implementation, and evaluation of "algorithmic solutions to [real-world] problems so that the solutions could be operationalized with a computer" (Fraillon et al. 2019, p. 27). The ICILS CT assessment combines tasks involving planning for a computer-operationalized solution to a real-world problem and a suite of visual coding tasks.

CIL and CT are quite different achievement constructs. This is evident through examination of their definitions and descriptions as well as the tasks used to assess each domain. However, they do share some common features. They both are, and can only be, completed on computer. Each one therefore draws on understandings of how computers can be used to solve problems as described by CIL Aspect 1.1 (knowing about and understanding computer use) and CT Aspect 1.1 (knowing about and understanding digital systems). Furthermore, achievement in each of CIL and CT draws on students' literacy skills (in reading and responding to tasks) and critical thinking (through the evaluation of information, data, and solutions to problems). In ICILS 2018 it was expected that, while the differences between the two domains would result in differences in achievement within students, the commonalities in the foundations of achievement of the domains would lead to a positive association between CIL and CT achievement.

On average across all countries, the correlation between students' CIL and CT scale scores was 0.82 (Table 4.6). This strong correlation between CIL and CT scores was consistent across countries and varied from 0.74 in Korea to 0.89 in Finland. We report the correlations between CIL and CT scores across all countries as well as the average CT scale scores for students performing within each CIL proficiency level across countries (Table 4.3). Across all countries the average CT scores of students increase as the CIL levels of students increase. On average across all countries, the difference in student CT scale scores between students in adjacent CIL levels of achievement varied from 90 CT scale points (between students with CIL of Level 1 and Below Level 1) and 60 CT scale points (between students with CIL of Level 3 and Level 4 or above). Across countries there was a general tendency for the difference in average CT scale scores of students in adjacent CIL levels of achievement to be larger between the lower levels of achievement than between the higher levels. We further investigated the correlation between students' CIL and CT by gender and found that they varied little from the overall correlations reported across countries. The consequences of the strong correlation between CIL and CT and their relationships to student gender are further discussed in Chapter 8.

Table 4.6: Correlations between CT and CIL and average CT performance for students at each CIL proficiency level across countries

Country	Correlation CT and CIL	Average CT score for students below CIL Level 1	Average CT score for students at CIL Level 1	Average CT score for students at CIL Level 2	Average CT score for students at CIL Level 3	Average CT score for students at CIL Level 4 and above
Denmark† 1	0.81 (0.01)	360 (11.6)	437 (4.9)	510 (2.8)	583 (3.5)	645 (13.8)
Finland	0.89 (0.01)	329 (7.6)	430 (4.9)	514 (2.6)	591 (3.7)	652 (14.8)
France	0.87 (0.01)	370 (6.9)	458 (3.2)	533 (2.4)	605 (5.4)	^ ^
Germany	0.81 (0.01)	328 (7.0)	420 (3.6)	502 (3.2)	578 (4.2)	659 (17.0)
Korea, Republic of	0.74 (0.01)	370 (8.9)	463 (5.9)	535 (4.3)	597 (3.8)	658 (7.9)
Luxembourg	0.80 (0.01)	331 (3.0)	430 (2.1)	510 (1.9)	585 (3.3)	651 (15.3)
Portugal†† 1	0.78 (0.01)	353 (7.2)	431 (3.2)	497 (2.4)	553 (5.3)	600 (13.0)
ICILS 2018 average	0.82 (0.00)	349 (3.0)	439 (1.6)	515 (1.1)	585 (1.6)	644 (5.7)
Not meeting sample participation requirements						
United States	0.82 (0.01)	334 (4.5)	429 (2.6)	515 (2.0)	594 (2.6)	673 (9.8)
Benchmarking participant meeting sample participation requirements						
North Rhine-Westphalia (Germany)	0.81 (0.01)	340 (8.7)	430 (4.0)	504 (3.0)	574 (6.0)	640 (14.4)

Notes: Standard errors appear in parentheses.

^ Number of students too small to report group averages.

† Met guidelines for sampling participation rates only after replacement schools were included.

†† Nearly met guidelines for sampling participation rates after replacement schools were included.

1 National defined population covers 90% to 95% of the national target population.

References

Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). *IEA International Computer and Information Literacy Study 2018 assessment framework*. Cham, Switzerland: Springer. Retrieved from https://www.springer.com/gp/book/9783030193881.

Fraillon, J., Schulz, W., Friedman, T., & Meyer, S. (Eds.). (2020). *IEA International Computer and Information Literacy Study 2018 technical report*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Manuscript in preparation.

ITU. (2019). IDI 2017 rank [webpage]. Geneva, Switzerland: International Telecommunications Union. Retrieved from https://www.itu.int/net4/ITU-D/idi/2017/index.html.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier, & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, Volume 2* (pp. 9–36). Hamburg, Germany/Princeton, NJ: International Association for the Evaluation of Educational Achievement (IEA)/Educational Testing Service (ETS). Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_01.pdf.