

PISA 2022 Technical Report



19

Scaling procedures and construct validation of context questionnaire data

Introduction

The PISA 2022 Context Questionnaires are based on the questionnaire framework (OECD, 2023^[1]) described in Chapter 5 of this technical report. Many questionnaire items were designed to be combined in some way in order to represent latent constructs that cannot be observed directly (e.g., a student's mathematics self-efficacy; sense of belonging; or economic, social, and cultural status). To construct meaningful indices, transformations or scaling procedures were applied to these items.

In the following sections, these indices are referred to as *derived variables* (DVs). This chapter describes the DVs based on one or more items that were constructed and validated for all questionnaires across the respondent groups – students, parents, schools, and teachers – administered in PISA 2022.

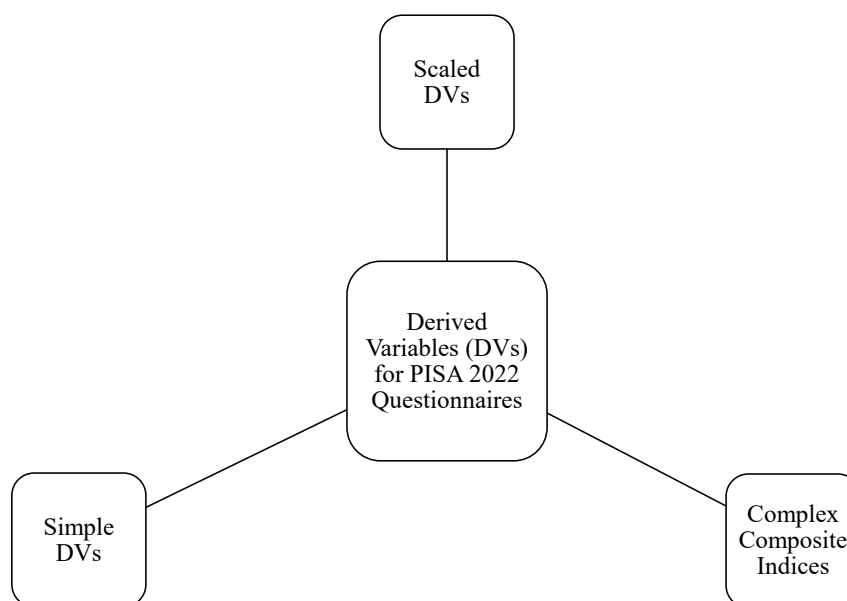
As in the previous PISA surveys, three different kinds of DVs can be distinguished (see Figure 19.1):

- *simple questionnaire indices* constructed through the arithmetical transformation or recoding of one or more items;
- *scaled indices* based on item response theory (IRT) scaling; and
- *complex composite indices* based on a combination of two or more indices.

As described in Chapter 5, the PISA 2022 Context Questionnaires included a broad scope of contextual factors assessed with different questionnaire instruments. While the student and school questionnaires were mandatory in all countries/economies, many countries/economies also administered an optional questionnaire for the parents of the participating students. In addition, countries/economies could choose to administer the optional *Financial Literacy Questionnaire*, the *Information and Communication Technology (ICT) Familiarity Questionnaire*, and the *Well-Being Questionnaire* to students. Moreover, several countries/economies also chose to administer the optional *Teacher Questionnaire*, which included questionnaires for mathematics teachers and general teachers.

This chapter describes the methodology used for the scaled DVs and also presents an overview of all the simple and scaled DVs for each questionnaire.

Figure 19.1. Types of derived variables for questionnaires in PISA 2022



Within-construct matrix sampling

Previous PISA cycles have used different strategies for collecting data on relevant contextual variables via the Student Questionnaire. For example, PISA 2012 used a three-form booklet design through which each student was administered items for some but not all of the constructs in the questionnaire. The main benefit of this design was that it allowed for the collection of data on approximately 33% more contextual items at the population level without overburdening students with a single-booklet design. However, a disadvantage of this design was the introduction of systematic missing data for students at the construct level, preventing researchers conducting secondary analyses to fully study the relationships between all possible sets of constructs, since no student was administered items for all of the constructs. PISA 2015 and PISA 2018 used a single-form booklet design through which all students were administered the same items. The main benefit of this design was that it allowed for the creation of a database without systematic missing data at the construct or item level, enabling researchers conducting secondary analyses to study the relationships between all possible sets of items and constructs. However, a disadvantage of this design was that it only allowed for the administration of a smaller set of items for each construct compared to the design used in PISA 2012, leading to a large number of relatively short 3-item scales with somewhat limited representation of the broad underlying construct.

PISA 2022 used a new within-construct matrix sampling design that combined the advantages of the multi-form and single-form booklet designs. This design was studied extensively using data from previous PISA cycles as well as the Field Trial data before it was implemented in the Main Survey (Bertling and Weeks, 2018^[1]; 2020^[2]; Bertling et al., 2020^[3]). Specifically, with this new design, every student was administered a random subset of five items for each construct. This design ensured that each item was administered to approximately the same number of students in each country/economy as well as the overall sample. It also allowed each construct to be assessed in larger breadth, kept individual students' burden comparable to previous cycles, and substantially reduced the reading load for students by displaying only five items on each screen.

This within-construct matrix design was only used for the IRT based scales in the Student Questionnaire, as the primary reporting objective for these scales was at the construct level instead of the item level. Also,

this design was not used for any scales pertaining to the economic, social, and cultural status index. In addition, it was not used for any of the optional questionnaires administered to students (i.e., Financial Literacy Questionnaire, ICT Familiarity Questionnaire, Well-Being Questionnaire) or questionnaires administered to adult respondents (i.e., Parent Questionnaire, School Questionnaire, Teacher Questionnaire) due to the smaller sample sizes for these questionnaires. Table 19.1 provides a list of the 32 scales in the Student Questionnaire that were administered using the within-construct matrix sampling design.

Scaling methodology and reporting of scores

Scaling methodology

As in previous cycles of PISA, some of the DVs were constructed using IRT. More specifically, the two-parameter logistic model (2PLM) (Birnbaum, 1968^[4]) was used to scale items with only two response categories (i.e., dichotomous items), while the generalised partial credit model (GPCM) (Muraki, 1992^[5]), was used to scale items with more than two response categories (i.e., polytomous items).¹ A detailed explanation of each model is in the following sections. The software *mdltm* (version 1.965) (Shin et al., 2017^[7]; von Davier, 2015^[8]) was used for the scaling.

In the initial scaling, item parameters were estimated using data from all individuals with available data from all participating countries/economies. Each country/economy was included in the analysis using a senate weight (SENWT). The senate weight is a linear transformation of the student full sampling weight (W_FSTUWT) such that the sum of SENWT for all cases within a country/economy add up to a constant of 5 000. Due to missing responses within each country/economy, the sum of the SENWT of the cases used in the calibration of each scale varied on a scale-by-scale basis.

For countries/economies with more than one language group, a language group was treated as an independent group in the scaling process if the group's sample size was over 150 and the sum of the weights was over 300. The groups used in the scaling are called *country-by-language groups* since they are defined by both country/economy and language group. For simplicity, the country-by-language groups are also called *groups* in the remainder of this chapter. Note that if the sample size for an entire country/economy was 150 or less, data from the country/economy were not included in the estimation of the item parameters, and the country/economy was assigned international item parameters (explained below) that had been estimated with data from the other countries/economies.

Several of the scales had items with negative valence. These are items for which a higher response category signified a lower level of the construct being measured, and vice versa. The responses to these items were reverse-coded prior to scaling. For all items, including the reverse-coded items, the responses were recoded so that the response corresponding to the lowest level of the construct was coded as 0. Any missing response data, whether it was because an item was not administered or a student did not respond to an item, were ignored and were not included in the analysis.

The two parameter logistic model (2PLM)

The 2PLM, a generalisation of the Rasch model (Rasch, 1960^[7]), assumes that the probability of a response x to be positive (coded as 1 in this case) by individual v to item i depends on the difference between the respondent v 's trait level θ and the location of the item β . In addition, the 2PLM postulates that for every item, the association between this difference and the response probability depends on an additional item discrimination parameter α . The equation for the response probability for an item under the 2PLM is presented in Formula 19.1:

Formula 19.1

$$P(x_{iv} = 1 | \theta_v, \beta_i, \alpha_i) = \frac{\exp(D\alpha_i(\theta_v - \beta_i))}{1 + \exp(D\alpha_i(\theta_v - \beta_i))}$$

The item location parameter β can be regarded as the item's general location on the latent continuum of the construct being measured. Items with a higher β parameter require a higher latent trait for a positive response to be selected.

The item discrimination parameter α , which was scaled by a constant $D = 1.7$ starting in PISA 2015 when the 2PLM was used instead of the Rasch model, characterises how quickly the probability of responding positively to an item approaches 1 with an increase in the trait level θ . In other words, α describes how well a certain item relates to the latent trait θ and, therefore, discriminates between individuals with different trait levels. To solve the indeterminacy of the IRT scale, the average of the item discrimination parameters α across all the items in the scale was constrained to 1. A special case of the 2PLM is when $\alpha = 1$ for all items, in which case the model is equivalent to the Rasch model.

The generalised partial credit model (GPCM)

The GPCM (Muraki, 1992^[5]) is a mathematical model for the probability that an individual will select a certain response category for an item with more than two response categories. Note that the GPCM is a generalisation of the 2PLM and that it reduces to the 2PLM when applied to items with only two response categories. For an item i with $m + 1$ ordered categories, the probability of an individual selecting a certain response category k ($0, 1, 2, \dots, m$) under the GPCM and adopting the same notation employed above can be written as:

Formula 19.2

$$P(x_i = k | \theta_v, \beta_i, \alpha_i, d_i) = \frac{\exp\{\sum_{r=0}^k D\alpha_i(\theta_v - \beta_i + d_{ir})\}}{\sum_{u=0}^m \exp\{\sum_{r=0}^u D\alpha_i(\theta_v - \beta_i + d_{ir})\}}$$

As with the 2PLM, the overall item location parameter β can be regarded as the item's general location on the latent continuum of the construct being measured. Items with a higher β parameter require a higher latent trait for a higher response category to be selected. d is the step parameter (of which there are m for an item with $m + 1$ categories, with the step parameters for each item summing to 0) which represents the deviation of the category intersection δ from the general location β .

The category intersection δ is the intersection between two neighbouring category characteristic curves, in other words, the point on the latent continuum θ at which a higher response category is more likely to be selected (e.g., when the individual is more likely to select "disagree" than "strongly disagree"). Note that β and d can be used to calculate the category intersection δ using Formula 19.3.

Formula 19.3

$$\delta_k = \beta - d_k$$

The discrimination parameter α , which was scaled by a constant $D = 1.7$ starting in PISA 2015, signifies the slope of the category characteristic curves. In other words, it indicates how well selecting a certain response category discriminates between individuals on the latent continuum θ . To solve the indeterminacy of the IRT scale, the average of the item discrimination parameters α across all the items in the scale was constrained to 1. A special case of the GPCM is when $\alpha = 1$ for all items, in which case the model is equivalent to the partial credit model (PCM) (Masters, 1982^[5]).

Figure 19.2 displays the category characteristic curves of a four-category item (e.g., a Likert-type item with response categories “strongly disagree”, “disagree”, “agree”, and “strongly agree”), with the three item parameters used in the GPCM (i.e., α , β and d) represented in the figure. For comparison, Figure 19.3 displays the category characteristic curves of an item for which only the α parameter has been increased while the β and d parameters were kept the same as in Figure 19.2.

Figure 19.2. Category characteristic curves for a four-category item under the generalised partial credit model (GPCM)

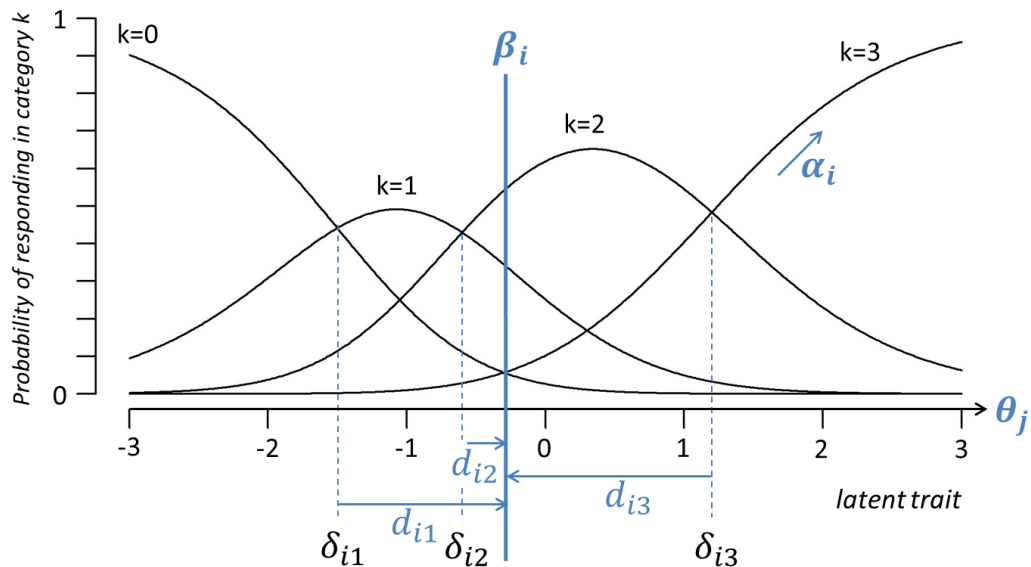
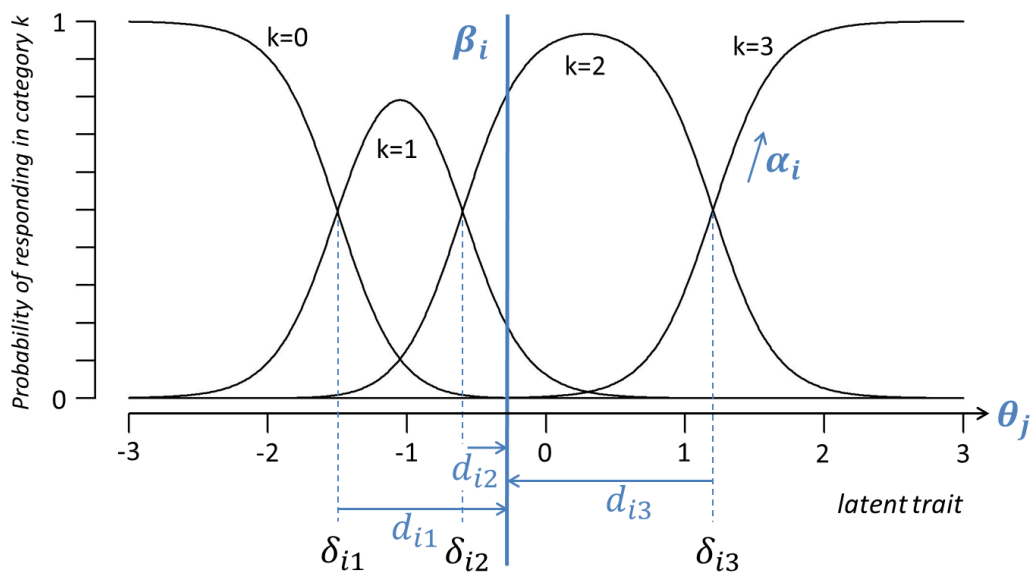


Figure 19.3. Illustration of how an increase in the slope parameter α affects the category characteristic curves of the model above



Special handling of trend scales

For the trend scales, the scaling process began by fixing the item parameters of the trend items to the parameters that had been estimated for each group in the previous cycle, a procedure called *fixed parameter linking*. Also, in line with the models that were used in the past cycles, the trend scales linked to PISA 2018 were scaled using the 2PLM and GPCM, while the trend scales linked to PISA 2012 were scaled using the Rasch model and PCM. This was done so that the scale scores from the current cycle would be comparable to the scale scores from the previous cycle. To compute trends, a scale needed to have at least three trend items, but some trend scales consisted of both trend items and new items. In this case, the item parameters for the trend items were fixed at the beginning of the scaling process, but the item parameters for the new items were estimated using the PISA 2022 data. Note that all the items in the trend scales were also evaluated for the goodness-of-fit of the trend parameters, a process described below. Please see Table 19.2 for a full list of trend scales in PISA 2022.

Releasing item parameters

PISA 2022 adopted and further refined the approach for evaluating the invariance of latent constructs across groups using multiple-group concurrent calibration with partial invariance constraints, a method which was first introduced in PISA 2015.

As explained above, in the initial scaling, item parameters were estimated using data from all individuals with available data from all countries/economies. The item parameters that were estimated in this initial scaling process are called *international parameters* since they were estimated using responses from most or all participating countries/economies. After the initial scaling, the fit of the international parameters for each item was evaluated for each group using the root mean square deviance (RMSD). The $RMSD_g$ for group g is defined as:

Formula 19.4

$$RMSD_g = \sqrt{\int [p_g^{obs}(\theta) - p_g^{exp}(\theta)]^2 f_g(\theta) d\theta}$$

quantifying the difference between the *observed item characteristic curve (ICC)* for the group based on the pseudo counts from the E-step of the Expectation-Maximisation (EM) algorithm ($P_{obs,gk}(\theta)$) with the model-based Item Characteristic Curve (ICC; $P_{exp,gk}(\theta)$) (Shin et al., 2017^[7]). RMSD values range from 0 to 1, with values close to 0 indicating good item fit, meaning that the model-based item parameters fit the data for the group well. Note that the RMSD statistic is sensitive to group-specific deviations of both the item location parameter β and the item discrimination parameter α .

When the RMSD for an item*group exceeded a pre-defined cut-off of 0.25, it was considered that the model-based item parameters did not fit the group's data well, and unique item parameters (also called *group-specific item parameters*) were estimated for the group using data only from that group. However, if more than one group had similar response patterns for an item, data from those groups were pooled together and the same unique parameters were estimated for those groups. This process is called *releasing item parameters*. Item parameters were released until all item*groups had an RMSD value under 0.25.

In PISA 2015 and PISA 2018, an RMSD value of 0.3 was used as the cut-off criterion for releasing item parameters for the context questionnaires. However, an analysis of the scaling results from the PISA 2018

Student Questionnaire suggested that this threshold may have over-emphasised international comparability of the model over group-level model-data fit, as very few item*groups received unique parameters. For PISA 2022, the RMSD threshold for releasing item parameters for the context questionnaires was lowered to 0.25, as it was found that this new threshold could improve the group-level model-data fit without weakening the comparability of the model across groups (as measured by the percent of item*groups with unique parameters, number of groups with international parameters for three or more items in a scale, and the rank order correlation of the scale scores from the models with and without unique parameters).

The final distribution of the RMSD values across groups for each scale item after the final scaling is documented in Annex E. Note that the figures do not include RMSD values for item*groups with an unweighted sample size of 150 or less, as the sample size was too small to estimate stable group-specific ICCs.

Scale scores

The scaling process described above produced weighted likelihood estimates (WLE) (Warm, 1989^[11]) for each individual. These WLE scores were subsequently standardised through the process described in the following sections. Note that if an individual had fewer than three valid responses for a scale, a WLE score was not produced for the individual and his/her scale score was replaced with “99” in the SPSS file and “.M” in the SAS file.

New scales

For the new scales, the original WLE scores were transformed into a reporting metric to have a mean of 0 and a standard deviation of 1 across the OECD countries, using senate weights for all cases with available data. The transformation was achieved by applying Formula 19.5:

Formula 19.5

$$\theta'_v = \frac{\theta_v - \bar{\theta}_{OECD}}{\sigma_{\theta(OECD)}}$$

where θ'_v is the scale score on the reporting metric, θ_v is the original WLE, $\bar{\theta}_{OECD}$ is the mean of the original WLEs across the OECD countries, and $\sigma_{\theta(OECD)}$ is the standard deviation of the original WLEs across the OECD countries. The transformation constants that were used to transform the original WLEs into the reporting scale are displayed in Table 19.3.

For the new scales, an average scale score of 0 is expected when calculated across all OECD countries using the senate weights. A negative scale score does not imply that a student responded negatively to the items in the scale. Rather, it means that the student is below the OECD average.

Trend scales

For the trend scales, to ensure the comparability of the scale scores from the current cycle to the scale scores from the previous cycle, the original WLEs of PISA 2022 were transformed using the same transformation constants of the original WLEs from the cycle to which the current cycle was linked. Table 19.4 presents the transformation constants of the original WLEs in PISA 2018 for the trend scales linked to PISA 2018, while Table 19.5 presents the transformation constants used in PISA 2012 for the trend scales linked to PISA 2012.

Criteria for suppressing scale scores

The scale scores of individuals or groups were suppressed under the following conditions.

Low internal consistency

Cronbach's alpha coefficient was used to check the internal consistency of each scale for each group. This coefficient ranges from 0 to 1, with a higher value indicating higher internal consistency. A group needed to have a Cronbach's alpha of at least 0.60 for a scale in order for the group's scale scores to be reported. Scale scores were suppressed for countries/economies in which one or more language groups had a Cronbach's alpha under 0.60 for the scale.

Few items with international parameters

For each scale, a group needed to have at least three items with international parameters in order for the scale scores of the group to be considered comparable to the scale scores of the other groups. Scale scores were suppressed for countries/economies in which one or more language groups had less than three items with international parameters for the scale. The scale scores for the individuals in these countries/economies were replaced with "97" in the SPSS file and ".N" in the SAS file.

Lack of trend items with international parameters

For the trend scales, a group needed to have at least three trend items with international parameters in order for the PISA 2022 scales scores for the group to be considered comparable to the scale scores of the previous cycle to which the current cycle was linked. Scale scores were suppressed for countries/economies in which one or more language groups had less than three trend items with international parameters for the scale. The scale scores for the individuals in these countries/economies were replaced with "97" in the SPSS file and ".N" in the SAS file.

Student Questionnaire derived variables

There were 86 variables derived from the Student Questionnaire, including 43 simple DVs, 42 IRT scaled DVs, and one complex composite index. The DVs are shown in Table 19.6 and will be described in the following sections; the first section covers all simple DVs, the second section covers those that are based on IRT scaling, and the last section covers the complex composite index. The simple and scaled DVs are organised first by framework module (please see Chapter 5 for a description of modules) and then alphabetical order within modules.

Simple questionnaire indices

Basic demographics (Module 1)

Student's age (AGE)

The age of a student (AGE) was calculated as the difference between the year and month of the testing and the year and month of a student's birth, which was obtained from school records from the student sampling data and validated by comparing to the students' responses in the questionnaire. Data on students' age were obtained from both the questionnaire (ST003) and the student tracking forms. The formula for computing AGE was:

Formula 19.6

$$\text{AGE} = (100 + T_y - S_y) + (T_m - S_m)/12$$

where T_y and S_y are the year of the test and the year of the students' birth, respectively, in two-digit format (for example "06" or "92"), and T_m and S_m are the month of the test and month of the students' birth, respectively. The result is rounded to two decimal places.

Grade compared to modal grade in country (GRADE)

The relative grade index (GRADE) was computed to capture between-country/economy variation. It indicates whether students are in the country/economy's modal grade (value of 0), or the number of grades below or above the modal grade in the country. The information about the students' grade level was obtained from school records from the student sampling data and validated by comparing the students' responses in the Student Questionnaire (ST001).

Gender (ST004D01T)

The gender of a student which was obtained from school records from the student sampling data and validated by comparing to the student's responses in the questionnaire (ST004).

Economic, social and cultural status (Module 2)

Mother's level of education (MISCED)

Student responses to questions ST005 and ST006 regarding their mothers' education were used to derive the mother's level of education (MISCED) index, where education level ranged from "1" less than ISCED level 1 to "10" ISCED level 8, as noted in Table 19.7.²

Father's level of education (FISCED)

Student responses to questions ST007 and ST008 regarding their fathers' education were used to derive the father's level of education (FISCED) index, where education level ranged from "1" less than ISCED level 1 to "10" ISCED level 8, as noted in Table 19.7 above.

Highest level of education of parents (HISCED)

Students' responses to questions ST005, ST006, ST007, and ST008 regarding their mothers' and fathers' education were used to derive the index of highest education level of parents (HISCED). The index is equal to the highest ISCED level of either parent.

Highest education of parents in years (PAREDINT)

The index of the highest education of parents in years, PAREDINT, was based on the median cumulative years of education associated with completion of the highest level of parental education (HISCED). Cumulative years of education values used in PISA 2018 were assigned to each ISCED level (see Table 19.7). Mother's occupational code (OCOD1)

Students' responses to the fill-in question ST014 about their mothers' occupation were human-coded based on the International Standard Classification of Occupations (ISCO)-08 classification system, resulting in the mother's occupational code (4-digit ISCO; ILO, 2007) index, OCOD1. These 4-digit codes range from 0000 to 9705. Codes 0000 to 9629 are occupations from the ISCO-08 classification system. Codes 9701-9705 were used to classify responses that fell outside of the ISCO-08 classification system.

Specifically, the code 9701 indicates “stay-at-home parent”, 9702 indicates “student”, and 9703 indicates “social beneficiary (e.g., unemployed, retired, sick)”. Lastly, “I don’t know” responses were coded 9704 and vague responses (e.g., a good job, a well-paid job) were coded 9705.

Father’s occupational code (OCOD2)

Students’ responses to the fill-in question ST015 about their fathers’ occupation were human-coded based on the ISCO-08 classification system, resulting in the father’s occupational code (4-digit ISCO) index, OCOD2. These 4-digit ISCO-08 codes range from 0000 to 9705. Codes 0000 to 9629 are occupations from the ISCO-08 classification. Codes 9701-9705 were used to classify responses that fell outside of the ISCO-08 classification. Specifically, the code 9701 indicates “stay-at-home parent”, 9702 indicates “student”, and 9703 indicates “social beneficiary (e.g., unemployed, retired, sick)”. Lastly, “I don’t know” responses were coded 9704 and vague responses (e.g., a good job, a well-paid job) were coded 9705.

Mother’s occupational status (BMMJ1)

The mother’s occupational status index, BMMJ1, was derived from the OCOD1 index and international socio-economic index of occupational status (ISEI) (Ganzeboom and Treiman, 2003^[5]) scores. The 4-digit ISCO-08 occupation codes in OCOD1 were mapped onto ISEI ratings.

Father’s occupational status (BFMJ2)

The father’s occupational status index, BFMJ2, was derived from the OCOD2 index and international socio-economic index of occupational status (ISEI) scores. The 4-digit ISCO-08 occupation codes in OCOD2 were mapped onto ISEI occupational status scores.

Highest parental occupational status (HISEI)

This highest parental occupational status index (HISEI) was based on the 4-digit ISCO-08 occupational codes that were human coded from students’ responses to questions ST014 and ST015 about their mother and father’s occupations, respectively. The index was equal to the higher of the mother’s (BMMJ1) and father’s (BFMJ2) ISEI scores.

Educational pathways and post-secondary aspirations (Module 3)

Duration in early childhood education and care (DURECEC)

Questions ST125 and ST126 measure the starting age in ISCED 1 and ISCED 0. The indicator DURECEC is built as the difference of ST126 and ST125 plus the value of “2” to indicate the number of years a student spent in early childhood education and care.

Study programme level and orientation (ISCEDP)

PISA collects data on study programmes available to 15-year-old students in each country/economy. This information is obtained through the student tracking form and the Student Questionnaire (ST002). In the final database, all national programmes are included in a separate DV (PROGN) where the first six digits represent the National Centre code, and the last two digits are the nationally specific programme code. All study programmes were classified using the International Standard Classification of Education (ISCED 2011).

The study programme level and orientation index (ISCEDP) is a three-digit index that describes whether students were at the lower or upper secondary level and (ISCED 2 or ISCED 3) and whether their

programmes were general or vocational and sufficient for level completion with direct access to tertiary or post-secondary non-tertiary education. ISCEDP values and labels can be found in Table 19.8.

Grade repetition (REPEAT)

Students' answers on question ST127 of whether and, if yes, how often they have ever repeated a grade at ISCED levels 1, 2, and 3 were combined into the index REPEAT. Each item included three response options ("No, never", "Yes, once", "Yes, twice or more"). REPEAT took the value of "0" if the student never repeated a grade (student did not select options 2 or 3 for any of the three items) and the value of "1" if the student repeated a grade at least once (student selected options 2 or 3 for at least one of the three items). The index was assigned a missing value if none of the three response options were selected in any levels.

Missing school (MISSSC)

Students' answers on question ST260 of whether and, if yes, how often they have ever missed school for more than three months in a row at ISCED levels 1, 2, and 3 were combined into the index MISSSC. Each item included three response options ("No, never", "Yes, once", "Yes, twice or more"). MISSSC took the value of "1" if the student selected options 2 or 3 for at least one of the three items, and the value "0" otherwise. The index was assigned a missing value if none of the three response options were selected in any levels.

Skipping classes or days of school (SKIPPING)

Students' responses to whether, in the two weeks prior to the PISA test, they had skipped classes (ST062Q02TA) or days of school (ST062Q01TA) at least once were used to derive an indicator of student truancy. Both questions have four response options ("Never", "One or two times", "Three or four times", "Five or more times"). The indicator takes a value of 0 if students reported that they had not skipped any class or day of school in the two weeks before the PISA test, and a value of 1 if students reported that they had skipped classes or days of school at least once in the same period.

Arriving late for school (TARDYSD)

Students responded to a question about whether and how frequently they had arrived late for school during the two weeks prior to the PISA test (ST062Q03TA). TARDYSD takes a value of "0" for on-time students if students reported that they had not arrived late for school, a value of "1" for occasional late arrivals if students report they arrived late for school one or two times, and "2" for frequent late arrivals if students reported they had arrived late for school three or more times.

Highest expected educational level (EXPECEDU)

Students' responses which of a list of possible educational levels they expect to complete in question ST327 were transformed into the index of "Highest Expected Educational Level". This DV has been newly created for 2022. Values on the index can range from "Less than ISCED level 2" to "ISCED level 8". Scores are assigned as shown in Table 19.9.

Expected occupation (OCOD3) and Expected occupation status (BSMJ)

Students' responses to the fill-in question ST329 about what kind of job they expect to have when they are about 30 years old were human-coded based the ISCO-08 classification system, resulting in the index "Expected Occupation (OCOD3)". These ISCO codes were then mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom and Treiman, 2003^[5]) in variable BSMJ. Higher scores on this variable indicate higher levels of a student's expected occupational status.

Clear idea about future job (SISCO)

The students who had a clear idea about their future job index (SISCO) was based on the human-coded open-ended expected occupation index, OCOD3, which was derived from question ST329. Students who had no clear idea about their future jobs were considered those who indicated “I do not know” or gave a vague answer such as “a good job”, “a quiet job”, “a well-paid job”, “an office job” in response to question ST329. In the OCOD3 index, “I don’t know” responses were coded 9704 and vague responses were coded 9705. Examples of invalid responses include students who did not answer the question or gave an answer, such as a smiley face. Specifically, a value of “0” is assigned on the index if OCOD3 values are 9704 or 9705, and a value of “1” is assigned if OCOD3 values are 0000 to 9703.

Migration and language exposure (Module 4)

Based on students’ responses to question ST019 (“In what country were you and your parents born?”), five indices are created as outlined below.

Student’s country of birth (COBN_S)

This index has the value “1” if the student selected the country of test (“Country A”) in question ST019AQ01T, and “0” otherwise.

Student mother’s country of birth (COBN_M)

This index has the value “1” if the student selected the country of test (“Country A”) in question ST019BQ01T, and “0” otherwise.

Student father’s country of birth (COBN_F)

This index has the value “1” if the student selected the country of test (“Country A”) in question ST019CQ01T, and “0” otherwise.

Index on immigrant background (IMMIG)

The index on immigrant background (IMMIG) is calculated from the three variables above (COBN_S, COBN_M, COBN_F), and has the categories as listed below. Students with missing responses for either the student or for both parents were given missing values for this variable.

1. Native students (those students who had at least one parent born in the country/economy);
2. Second-generation students (those born in the country/economy of assessment but whose parent[s] were born in another country/economy);
3. First-generation students (those students born outside the country/economy of assessment and whose parents were also born in another country/economy).

Language spoken at home (LANGN)

Students also indicated what language they usually spoke at home, and the database includes a variable (LANGN) containing country/economy-specific code for each language.

*Subject-specific beliefs, attitudes, feelings and behaviours (Module 7)***Relative motivation to do well in mathematics compared to other core subjects (MATHMOT)**

This simple index captures whether students indicate being more motivated to do well in mathematics than in Test Language and Science class. If students endorsed question ST268Q07JA (“I want to do well in my mathematics class.”) stronger than both items ST268Q08JA (“I want to do well in my <test language> class.”) and ST268Q09JA (“I want to do well in my <science> class.”), they received a “1” on this index, otherwise “0”. Please note that this index captures students’ relative motivation for math rather than their absolute motivation for mathematics. The latter is captured by the original response to the item.

Perception of mathematics as easier than other core subjects (MATHEASE)

This simple index captures whether students indicate they perceive mathematics as easier compared to the Test Language and Science. If students endorsed question ST268Q014A (“Mathematics is easy for me.”) stronger than both items ST268Q05JA (“<Test language> is easy for me.”) and ST268Q06JA (“<Science> is easy for me”), they received a “1” on this index, otherwise “0”. Please note that this index captures students’ relative easiness of math rather than their absolute easiness rating for mathematics. The latter is captured by the original response to the item.

Preference of mathematics over other core subjects (MATHPREF)

This simple index captures whether students indicate they preferred mathematics over Test Language and Science. If students endorsed question ST268Q01JA (“Mathematics is one of my favourite subjects”) stronger than both items ST268Q02JA (“<Test language> is one of my favourite subjects.”) and ST268Q03JA (“<Science> is one of my favourite subjects.”), they received a “1” on this index, otherwise “0”. Please note that this index captures students’ relative preference of math rather than their absolute preference for mathematics. The latter is captured by the original response to the item.

*Out-of-school experiences (Module 10)***Exercising or practising a sport before or after school (EXERPRAC)**

Students’ answers on how many days during a typical school week they exercised or practised a sport before going to school and/or after leaving school in questions ST294 and ST295 were scaled into the index of “Exercise or practise a sport before or after school”. Each item included six response options (“0 days”, “1 day”, “2 days”, “3 days”, “4 days”, “5 or more days”). Values on this index range from 0 (no exercise or sports) to 10 (10 or more times exercise or sport a per week).

Studying for school or homework before or after school (STUDYHMW)

Students’ answers on how many days during a typical school week they studied for school or homework before going to school and/or after leaving school in questions ST294 and ST295 were scaled into the index of “Study for school or homework before or after school”. Each item included six response options (“0 days”, “1 day”, “2 days”, “3 days”, “4 days”, “5 or more days”). Values on this index range from 0 (no studying) to 10 (10 or more times of studying per week).

Working for pay before or after school (WORKPAY)

Students’ answers on how many days during a typical school week they worked for pay before going to school and/or after leaving school in questions ST294 and ST295 were scaled into the index of “Work for pay before or after school”. Each item included six response options (“0 days”, “1 day”, “2 days”, “3 days”,

“4 days”, “5 or more days”). Values on this index range from 0 (no work for pay) to 10 (10 or more times of working for pay per week).

Working in household or taking care of family members (WORKHOME)

Students’ answers on how many days during a typical school week they worked in the household or took care of a family member before going to school and/or after leaving school in questions ST294 and ST295 were scaled into the index of “Work in household or take care of family members”. Each item included six response options (“0 days”, “1 day”, “2 days”, “3 days”, “4 days”, “5 or more days”). Values on this index range from 0 (no work in household or care of family members) to 10 (10 or more times of working in household or caring for family members per week).

Derived variables based on IRT scaling

The Student Questionnaire provided data for 42 DVs based on IRT scaling. The Cronbach’s alpha for each scale and group are presented in Table 19.10, the number of items with international parameters for each scale and group are presented in Table 19.11, the number of trend items with international parameters for each trend scale and group are presented in Table 19.12, the countries/economies for which the scale scores were suppressed for each scale are presented in Table 19.13, and the groups that did not administer each scale are presented in Table 19.14.

Economic, social and cultural status (Module 2)

Home possessions (HOMEPOS)

In the HOMEPOS scale (which included questions ST250, ST251, ST253, ST254, ST255, and ST256), students indicated whether their household possessed certain items (e.g., “A room of your own”, “Educational software or apps”) or how many of an item their household possessed (e.g., “Rooms with a <flush toilet>”, “Cars, vans, or trucks”). This scale included 31 items, including four country/economy-specific items (ST250Q06JA, ST250Q07JA, ST251Q08JA, and ST251Q09JA) that were seen as local measures of family wealth within the country/economy’s context.³ In addition, students answered how many books (ST255) and digital devices with screens (ST253) were in their home. Note that all groups received unique item parameters for the country/economy-specific items (i.e., no international parameters were estimated for these items) and that for some items, the response categories were collapsed to align with the response categories used in previous cycles. Table 19.15 shows the item wording and item parameters for the items in this scale, while Table 19.16 shows how the response categories for each item were recoded prior to scaling.

ICT resources (ICTRES)

Students reported on the availability of 11 Information and Communications Technologies (ICT) resources in their home (e.g., “A computer (laptop, desktop, or tablet) that you can use for school work”, “Internet access (e.g., Wi-fi) (excluding through smartphones)”) in questions ST250 (which had two response categories), ST253 (which had eight response categories), and ST254 (which had four substantive response categories and an additional response category “I don’t know.” which was recoded as missing prior to scaling). These items were scaled into the index of “ICT resources”. Table 19.17 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded and how the response categories were recoded prior to scaling.

*Educational pathways and post-secondary aspirations (Module 3)***Information seeking regarding future career (INFOSEEK)**

Students' ratings of whether they had undertaken a range of possible activities to find out about future study or types of work (e.g., "I did an internship.", "I researched the internet for information about careers.") in question ST330 were scaled into the index of "Information seeking regarding future career". Note that this scale used a within-construct matrix sampling design. Each of the 11 items included in this scale had three response options ("Yes, once", "Yes, two or more times", "No"). Table 19.18 shows the item wording and item parameters for the items in this scale.⁴ It also shows how the response categories were recoded prior to scaling.

*School culture and climate (Module 6)***Being bullied (BULLIED)**

Students' ratings of how often they had a range of experiences at school that are indicative of being bullied during the past 12 months (e.g., "Other students left me out of things on purpose.", "Other students made fun of me.") in question ST038 were scaled into the index of "Being bullied". Note that this scale was linked to the BEINGBULLIED scale in PISA 2018. Each of the nine items included in this scale had four response options ("Never or almost never", "A few times a year", "A few times a month", "Once a week or more"). Table 19.19 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Feeling safe (FEELSAFE)

Students' ratings of their agreement with four statements about their perceived safety (e.g., "I feel safe on my way to school.", "I feel safe in my classrooms at school.") in question ST265 were scaled into the index of "Feeling safe". Each of the four items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.20 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Mathematics teacher support (TEACHSUP)

Students' frequency ratings of how often a range of situations occurred in their mathematics lessons (e.g., "The teacher shows an interest in every student's learning.", "The teacher gives extra help when students need it.") in question ST270 were scaled into the index of "Mathematics teacher support". Note that this scale was linked to the TEACHSUP scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Each of the four items included in this scale had four response options ("Every lesson", "Most lessons", "Some lessons", "Never or almost never"). Table 19.21 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which items are trend items.

Quality of student-teacher relationships (RELATST)

Students' ratings of their agreement with the eight statements (e.g., "The teachers at my school are respectful towards me.", "When my teachers ask how I am doing, they are really interested in my answer.") in question ST267 were scaled into the index of "Quality of student-teacher relationships". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.22 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

School safety risks (SCHRISK)

Students' answers of whether a range of events indicative of safety risks at school occurred during the past four weeks (e.g., "Our school was vandalised.", "I witnessed a fight on school property in which someone got hurt.") in question ST266 were scaled into the index of "School safety risks". Each of the five items included in this scale had two response options ("Yes", "No"). Table 19.23 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Sense of belonging (BELONG)

Students' ratings of their agreement with six statements (e.g., "I feel like I belong at school.", "I feel lonely at school.") in question ST034 were scaled into the index of "Sense of belonging". Note that this scale used a within-construct matrix sampling design and that it was linked to the BELONG scale in PISA 2018. Each of the six items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.24 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which items are trend items.

Subject-specific beliefs, attitudes, feelings, and behaviours (Module 7)

Growth mindset (GROSAGR)

Students' ratings of their agreement with a range of statements indicative of their mindset (e.g., "Your intelligence is something about you that you cannot change very much.", "Some people are just not good at mathematics, no matter how hard they study.") in question ST263 were scaled into the index of "Growth mindset". Each of the four items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.25 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Mathematics anxiety (ANXMAT)

Students' ratings of their agreement with statements about a range of attitudes towards mathematics (e.g., "I often worry that it will be difficult for me in mathematics classes.", "I feel anxious about failing in mathematics.") in question ST292 were scaled into the index of "Mathematics anxiety". Note that this scale was linked to the ANXMAT scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Also, it used a within-construct matrix sampling design. Each of the six items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.26 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which items are trend items.

Mathematics self-efficacy: Formal and applied mathematics (MATHEFF)

Students' ratings of how confident they felt about having to do a range of formal and applied mathematics tasks (e.g., "Calculating how much more expensive a computer would be after adding tax", "Solving an equation like $2(x+3) = (x+3)(x-3)$ ") in question ST290 were scaled into the index of "Mathematics self-efficacy: Formal and applied mathematics". Note that this scale was linked to the MATHEFF scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Also, it used a within-construct matrix sampling design. Each of the nine items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Note that in PISA 2012, the response options were presented to the students ordered from "Very confident" to "Not at all confident" possibly eliciting different response patterns related to the format of the question, and not necessarily related to the construct. Because of this, caution should be exercised when comparing scale scores across

these two cycles. Table 19.27 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Mathematics self-efficacy: Mathematical reasoning and 21st century mathematics (MATHEF21)

Students' ratings of how confident they felt about having to do a range of mathematical reasoning and 21st century mathematics tasks (e.g., "Extracting mathematical information from diagrams, graphs, or simulations", "Using the concept of statistical variation to make a decision") in question ST291 were scaled into the index of "Mathematics self-efficacy: Mathematical reasoning and 21st century mathematics". Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Table 19.28 shows the item wording and item parameters for the items in this scale.

Proactive mathematics study behaviour (MATHPERS)

Students' frequency ratings of how often they engaged in behaviours indicative of effort and persistence in mathematics (e.g., "I actively participated in group discussions during mathematics class.", "I put effort into my assignments for mathematics class.") in question ST293 were scaled into the index of "Proactive mathematics study behaviour". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had five response options ("Never or almost never", "Less than half of the time", "About half of the time", "More than half of the time", "All or almost all of the time"). Table 19.29 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Subjective familiarity with mathematics concepts (FAMCON)

Students' ratings of how familiar they were with different mathematical concepts representative of different levels of mathematical skill or understanding (e.g., "Divisor", "Exponential function", "3-dimensional geometry") in question ST289 were scaled into the index of "Subjective familiarity with mathematics concepts". Note that this scale was linked to the FAMCON scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Also, it used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options ("Never heard of it", "Heard of it once or twice", "Heard of it a few times", "Heard of it often", "Know it well, understand the concept"). Table 19.30 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

General social and emotional characteristics (Module 8)

All of the scales in this module used a within-construct matrix sampling design and included both positively and negatively valenced items. This allowed us to check the consistency of responses since we would expect those agreeing with the items with positive valence to disagree with items with negative valence, and vice versa. To this effect, some students were identified as extreme straightliners. These were students that were administered items with positive and negative valence and responded to all five items selecting the same extreme response category, "Strongly disagree" or "Strongly agree". Students that were identified as extreme straightliners were removed from the analysis.

Assertiveness (ASSERAGR)

Students' ratings of their agreement with statements about a range of behaviours indicative of assertiveness (e.g., "I take initiative when working with my classmates.", "I find it hard to influence people.") in question ST305 were scaled into the index of "Assertiveness". Note that this scale used a within-

construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.31 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.32 shows the percent of students in each country/economy that did not receive a scale score for ASSERAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Cooperation (COOPAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of cooperation (e.g., “I work well with other people.”, “I get annoyed when I have to compromise with others.”) in question ST343 were scaled into the index of “Cooperation”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.33 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.34 shows the percent of students in each country/economy that did not receive a scale score for COOPAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Curiosity (CURIOAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of curiosity (e.g., “I like to know how things work.”, “I am more curious than most people I know.”) in question ST301 were scaled into the index of “Curiosity”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.35 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.36 shows the percent of students in each country/economy that did not receive a scale score for CURIOAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Emotional control (EMOCOAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of emotional control (e.g., “I keep my emotions under control.”, “I get mad easily.”) in question ST313 were scaled into the index of “Emotional control”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.37 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.38 shows the percent of students in each country/economy that did not receive a scale score for EMOCOAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Empathy (EMPATAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of empathy (e.g., “I predict the needs of others.”, “It is difficult for me to sense what others think.”) in question ST311 were scaled into the index of “Empathy”. Note that this scale used a within-construct matrix sampling

design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.39 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.40 shows the percent of students in each country/economy that did not receive a scale score for EMPATAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Perseverance (PERSEVAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of perseverance (e.g., “I keep working on a task until it is finished.”, “I give up after making mistakes.”) in question ST307 were scaled into the index of “Perseverance”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.41 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.42 shows the percent of students in each country/economy that did not receive a scale score for PERSEVAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Stress resistance (STRESAGR)

Students’ ratings of their agreement with statements about a range of behaviours indicative of stress resistance (e.g., “I remain calm under stress.”, “I get nervous easily.”) in question ST345 were scaled into the index of “Stress resistance”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Strongly disagree”, “Disagree”, “Neither agree nor disagree”, “Agree”, “Strongly agree”). Table 19.43 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling. Table 19.44 shows the percent of students in each country/economy that did not receive a scale score for STRESAGR due to extreme straightlining or, for comparison, for not having enough responses (i.e., less than three responses for the scale). In both cases, the scale scores were replaced with “99” in the SPSS file and “.M” in the SAS file.

Exposure to mathematics content (Module 15)

Exposure to formal and applied mathematics tasks (EXPOFA)

Students’ frequency ratings of how often they had encountered a range of formal and applied mathematics tasks during their time at school (e.g., “Calculating how much more expensive a computer would be after adding tax.”, “Solving an equation like $2(x+3) = (x+3)(x-3)$ ”) in question ST275 were scaled into the index “Exposure to formal and applied mathematics tasks”. Note that this scale used a within-construct matrix sampling design. Each of the nine items included in this scale had four response options (“Frequently”, “Sometimes”, “Rarely”, “Never”). Table 19.45 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Exposure to mathematical reasoning and 21st century mathematics tasks (EXPO21ST)

Students’ frequency ratings of how often they had encountered a range of different types of mathematics tasks related to mathematical reasoning and 21st century mathematics tasks during their time at school (e.g., “Extracting mathematical information from diagrams, graphs, or simulations”, “Using the concept of statistical variation to make a decision”) in question ST276 were scaled into the index “Exposure to

mathematical reasoning and 21st century mathematics tasks”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had four response options (“Frequently”, “Sometimes”, “Rarely”, “Never”). Table 19.46 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Mathematics teacher behaviour (Module 16)

Cognitive activation in mathematics: Foster reasoning (COGACRCO)

Students’ frequency ratings of how often their mathematics teacher showed a range of behaviours indicative of fostering mathematics reasoning during the ongoing school year (e.g., “The teacher asked us to explain our reasoning when solving a mathematics problem.”, “The teacher asked us to defend our answer to a mathematics problem.”) in question ST285 were scaled into the index of “Cognitive activation in mathematics: Foster reasoning”. Note that this scale used a within-construct matrix sampling design. Each of the nine items included in this scale had five response options (“Never or almost never”, “Less than half of the lessons”, “About half of the lessons”, “More than half of the lessons”, “Every lesson or almost every lesson”). Table 19.47 shows the item wording and item parameters for the items in this scale.

Cognitive activation in mathematics: Encourage mathematical thinking (COGACMCO)

Students’ frequency ratings of how often their mathematics teacher showed a range of behaviours indicative of encouraging mathematical thinking during the ongoing school year (e.g., “The teacher encouraged us to “think mathematically.”, “The teacher asked us how different topics are connected to a bigger mathematical idea.”) in question ST283 were scaled into the index of “Cognitive activation in mathematics: Encourage mathematical thinking”. Note that this scale used a within-construct matrix sampling design. Each of the nine items included in this scale had five response options (“Never or almost never”, “Less than half of the lessons”, “About half of the lessons”, “More than half of the lessons”, “Every lesson or almost every lesson”). Table 19.48 shows the item wording and item parameters for the items in this scale.

Disciplinary climate in mathematics (DISCLIM)

Students’ frequency ratings of how often a range of situations occurred in their mathematics lessons (e.g., “Students do not listen to what the teacher said.”, “Students get distracted by using <digital resources> (e.g., smartphones, websites, apps).”) in question ST273 were scaled into the index of “Disciplinary climate in mathematics”. Note that this scale was linked to the DISCLIMA scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Also, this scale used a within-construct matrix sampling design. Each of the seven items included in this scale had four response options (“Every lesson”, “Most lessons”, “Some lessons”, “Never or almost never”). Table 19.49 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Parental/guardian involvement and support (Module 19)

Family support (FAMSUP)

Students’ ratings of how often their parents or someone else in their family engaged in a range of behaviours indicative of family support (e.g., “Discuss how well you are doing at school”, “Spend time just talking with you”) in question ST300 were scaled into the index of “Family support”. Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had five response options (“Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”). Table 19.50 shows the item wording and item parameters for the items in this scale.

*Creative thinking (Module 20)***Creative peers and family environment (CREATFAM)**

Students' ratings of their agreement with statements about the degree to which creative thinking is fostered and supported by their peers and family environment (e.g., "My friends are open to new ideas.", "At home, I am encouraged to use my imagination.") in question ST336 were scaled into the index of "Creative peers and family environment". Note that this scale used a within-construct matrix sampling design. Each of the six items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.51 shows the item wording and item parameters for the items in this scale.

Creative school and class environment (CREATSCH)

Students' ratings of their agreement with statements about the degree to which creative thinking is fostered and supported in their school and class environment (e.g., "My teachers value students' creativity.", "At school, I am given a chance to express my ideas.") in question ST335 were scaled into the index of "Creative school and class environment". Note that this scale used a within-construct matrix sampling design. Each of the six items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.52 shows the item wording and item parameters for the items in this scale.

Creative thinking self-efficacy (CREATEFF)

Students' ratings of how confident they felt about having to do a range of tasks reflective of creative thinking skills (e.g., "Coming up with creative ideas for school projects", "Inventing new things") in question ST334 were scaled into the index of "Creative thinking self-efficacy". Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Table 19.53 shows the item wording and item parameters for the items in this scale.

Creativity and openness to intellect (CREATOP)

Students' ratings of their agreement with statements regarding their own views on their creativity and openness to intellect (e.g., "Doing something creative satisfies me.", "I like games that challenge my creativity.") in question ST340 were scaled into the index of "Creativity and openness to intellect". Note that this scale used a within-construct matrix sampling design. Each of the 10 items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.54 shows the item wording and item parameters for the items in this scale.

Imagination and adventurousness (IMAGINE)

Students' ratings of their agreement with statements regarding their own views on their imagination and adventurousness (e.g., "I have difficulty using my imagination.", "Coming up with new ideas is satisfying to me.") in question ST342 were scaled into the index of "Imagination and adventurousness". Note that this scale used a within-construct matrix sampling design. Each of the seven items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.55 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Openness to art and reflection (OPENART)

Students' ratings of their agreement with statements regarding their own views on their openness to art and reflection (e.g., "I enjoy creating art.", "I reflect on movies I watch.") in question ST341 were scaled

into the index of “Openness to art and reflection”. Each of the five items included in this scale had four response options (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”). Table 19.56 shows the item wording and item parameters for the items in this scale.

Participation in creative activities at school (CREATAS)

Students’ ratings of how often they participated in creative activities that were available in their school (e.g., “Art classes/activities (e.g., painting, drawing)”, “Debate club”) in question ST337 were scaled into the index of “Participation in creative activities at school”. Note that the activities sampled in this question are the same as the activities in the “outside of school” version of this question (CREATOOS – ST338). Each of the eight items included in this scale had five substantive response options (“Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”) and an additional response option “Not available at school” which was recoded as missing prior to scaling. Table 19.57 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Participation in creative activities outside of school (CREATOOS)

Students’ ratings of how often they participated in creative activities outside of school (e.g., “Art classes/activities (e.g., painting, drawing)”, “Debate club”) in question ST338 were scaled into the index of “Participation in creative activities outside of school”. Note that the activities sampled in this question are the same as the activities in the “at school” version of this question (CREATAS – ST337). Each of the eight items included in this scale had five substantive response options (“Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”) and an additional response option “Not available” which was recoded as missing prior to scaling. Table 19.58 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Global crises (Module 21)

Note that the questions in this module were skipped for students who reported that their school had not been closed for more than a week due to COVID-19 in question ST347.

Family support for self-directed learning (FAMSUPSL)

Students’ frequency ratings of how often someone in their family provided specific kinds of learning support (e.g., “Help me create a learning schedule”; “Help me access learning materials online”) while the school building was closed due to COVID-19 in question ST353 were scaled into the index of “Family support for self-directed learning”. Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options (“Never”, “A few times”, “About once or twice a week”, “Every day or almost every day”). Table 19.59 shows the item wording and item parameters for the items in this scale.

Feelings about learning at home (FEELLAH)

Students’ ratings of their agreement with statements about how they felt about learning at home (e.g., “I enjoyed learning by myself.”, “My teachers were well prepared to provide instruction remotely.”) while the school building was closed due to COVID-19 in question ST354 were scaled into the index of “Feelings about learning at home”. Note that this scale used a within-construct matrix sampling design. Each of the six items included in this scale had four response options (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”). Table 19.60 shows the item wording and item parameters for the items in this scale.

Problems with self-directed learning (PROBSELF)

Students' frequency ratings of how often they had various problems completing their school work (e.g., "Problems with Internet access", "Problems with understanding my school assignments") while their school building was closed due to COVID-19 in question ST352 were scaled into the index of "Problems with self-directed learning". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options ("Never", "A few times", "About once or twice a week", "Every day or almost every day"). Table 19.61 shows the item wording and item parameters for the items in this scale.

Self-directed learning self-efficacy (SDLEFF)

Students' ratings of how confident they felt about having to do a range of self-directed learning tasks (e.g., "Finding learning resources online on my own", "Completing school work independently") should their school building close again in the future in question ST355 were scaled into the index of "Self-directed learning self-efficacy". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Table 19.62 shows the item wording and item parameters for the items in this scale.

School actions to sustain learning (SCHSUST)

Students' frequency ratings of how often someone from their school completed an activity to sustain their learning (e.g., "Sent me learning materials to study on my own", "Checked in with me to ensure that I was completing my assignments") while their school building was closed due to COVID-19 in question ST348 were scaled into the index of "School actions to sustain learning". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options ("Never", "A few times", "About once or twice a week", "Every day or almost every day"). Table 19.63 shows the item wording and item parameters for the items in this scale.

Types of learning resources used while school was closed (LEARRES)

Students' frequency ratings of how often they used specific learning resources (e.g., "Paper textbooks, workbooks, or worksheets", "Recorded lessons or other digital material provided by teachers from my school") while the school building was closed due to COVID-19 in question ST351 were scaled into the index of "Types of learning resources used while school was closed". Note that this scale used a within-construct matrix sampling design. Each of the eight items included in this scale had four response options ("Never", "A few times", "About once or twice a week", "Every day or almost every day"). Table 19.64 shows the item wording and item parameters for the items in this scale.

Complex composite index – Index of economic, social and cultural status (ESCS)

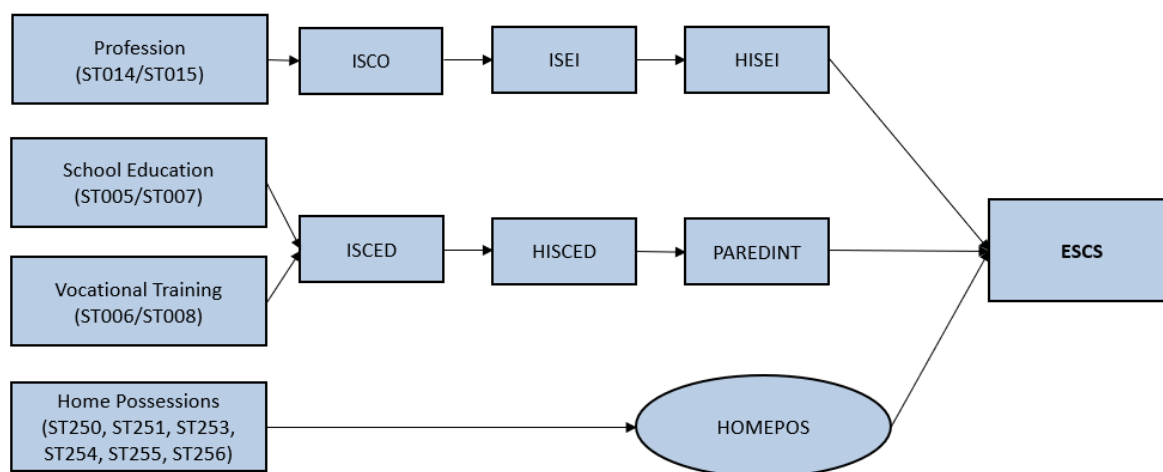
There was only one complex composite index derived from the Student Questionnaire – the index of economic, social and cultural status (ESCS).

Components of ESCS

The ESCS score was based on three indicators: highest parental occupation status (HISEI), highest education of parents in years (PAREDINT), and home possessions (HOMEPOS). The rationale for using these three components, which are consistent with the components used in previous PISA cycles, is that socio-economic status is most commonly theoretically conceptualized based on "the big 3" (occupational status, education, and income) (Cowan et al., 2012^[5]). As no direct income measure is available in the

PISA data, the existence of household items has been used as a proxy for family income. Figure 19.4 provides a schematic representation of ESCS and its components.

Figure 19.4. Computation of ESCS in PISA 2022



HISEI. For more information on HISEI, refer to the explanation on HISEI in the simple indices section above.

PAREDINT. For more information on PAREDINT, refer to the explanation on PAREDINT in the simple indices section above.

HOMEPOS. For more information on HOMEPOS, refer to the explanation on HOMEPOS in the IRT scale section above.

Computation of ESCS

The ESCS scores were computed using the same methodology used in PISA 2018 (Avvisati, 2020^[10]; OECD, 2020^[11]). For students with missing data on one out of the three components, the missing component was imputed using a regression equation which was created for each country/economy using data from students without any missing components. For each student with a missing component, this regression equation was used to predict the missing component with the two non-missing components and a random value was added to the predicted value to reflect the error of the regression model.⁵ If a student had missing data on more than one component, the ESCS score was not computed for the student, and the student's ESCS score was replaced with "99" in the SPSS file and ".M" in the SAS file.

After the imputation process, each of the three components (including the imputed values) was standardised to have a mean of 0 and a standard deviation of 1 across the OECD countries, with each OECD country weighted approximately equally using senate weights.⁶ The OECD means and standard deviations that were used to standardise each component of ESCS are displayed in Table 19.65.

Subsequently, the arithmetic mean of the three standardised components was calculated to create a preliminary ESCS score for each student. Lastly, the preliminary ESCS scores were standardised again to have a mean of 0 and a standard deviation of 1 across the OECD countries (again with each country weighted approximately equally using senate weights⁷), producing the final ESCS score for each student. The OECD mean and standard deviation that were used to transform the preliminary ESCS scores into the final ESCS scores are displayed in Table 19.65.

ESCS trend scores

In contrast to the other trend scales in the context questionnaires (for which the scale scores for the current cycle were made to be comparable to the scale scores from a previous cycle), the scores for each component of ESCS and the composite ESCS scores are not comparable to the scores from previous cycles. Instead, each of the component scores for ESCS and the composite ESCS scores for PISA 2012, PISA 2015, and PISA 2018 were recomputed to be comparable to the respective scores for PISA 2022. This was done by recoding the scores for each component of ESCS for PISA 2012, PISA 2015, and PISA 2018 using the coding scheme used in PISA 2022, then recomputing the composite ESCS score for these previous cycles using the ESCS computation methodology used in PISA 2022. More details are provided below.

HISEI. Until PISA 2009, ISCO-88 was used to code parental occupation. However, since PISA 2012, parental occupation has been coded using ISCO-08, the most recent version of ISCO. In PISA 2018, the coding scheme for ISEI was updated so that an ISEI value of 17 was attributed to ISCO codes 9701 (“stay-at-home parent”), 9702 (“student”), and 9703 (“social beneficiary”), equivalent to the ISEI value for ISCO code 9000 (“elementary occupations”). This coding scheme was also used in PISA 2022.

To make the HISEI scores for PISA 2012 and PISA 2015 comparable to the HISEI scores for PISA 2018 and PISA 2022, new HISEI scores were created for each student that participated in PISA 2012 and PISA 2015 using the coding scheme used in PISA 2018 and PISA 2022. These new HISEI scores were used in the computation of the trend ESCS scores.

PAREDINT. For some countries/economies, the mapping of ISCED levels to years of education was updated in 2009, 2015, and 2018, taking into account changes in the countries/economies’ educational systems. In PISA 2022, PAREDINT was updated again to map each ISCED level (based on ISCED-11) to the PISA 2018 cumulative years of education values, as presented in Table 19.7.

To make the PAREDINT scores for PISA 2012, PISA 2015, and PISA 2018 comparable to the PAREDINT scores for PISA 2022, new PAREDINT scores were created for each student that participated in the previous cycles the mapping presented in Table 19.7. These new PAREDINT scores were used in the computation of the trend ESCS scores.

HOMEPOS. Indicators of HOMEPOS have been dropped or added in all PISA cycles, taking into account the social, technical, and economic changes in the participating countries/economies. Moreover, the method for estimating HOMEPOS changed in PISA 2009, PISA 2012, and PISA 2015.

To make the HOMEPOS scores comparable across cycles, prior to scaling, the response categories for some items in PISA 2012, PISA 2015, and PISA 2018 were collapsed to align with the response categories used in PISA 2022 (as presented in Table 19.16). Then, the HOMEPOS WLEs for each student that participated in the past three cycles were re-estimated by fixing the item parameters to the parameters that were estimated for each group in PISA 2022 (either the international parameters or the group’s unique parameters, depending on whether the item parameters were released for the group in PISA 2022). For items that were not administered in PISA 2022, new international parameters were estimated by pooling data across all cycles and groups in which the item had been administered, then the item parameters were released until all groups in all cycles had an RMSD under 0.25. As an exception, unique parameters were estimated for all country/economy-specific items for all groups and all cycles. These newly estimated HOMEPOS WLEs were used in the computation of the trend ESCS scores.

ESCS. Prior to PISA 2018, the ESCS scores were computed using a principal component analysis (PCA), although there were differences across cycles regarding which countries/economies were included in the PCA and how the scores were standardised. In PISA 2018, the ESCS scores were computed as the arithmetic mean of the three components, with each of the component scores and the composite ESCS score standardised to have a mean of 0 and standard deviation of 1 across the OECD countries (with each

country weighted approximately equally using senate weights⁸). As noted above, this methodology was also used in PISA 2022.

To make the ESCS scores for PISA 2012, PISA 2015, and PISA 2018 comparable to the ESCS scores for PISA 2022, new ESCS scores were computed for the previous cycles using the methodology used in PISA 2022. Specifically, for students with missing data on one out of the three components, the missing component was imputed using a regression equation which was created for each country/economy in each cycle using data from students without any missing components. For each student with a missing component, this regression equation was used to predict the missing component with the two non-missing components and a random value was added to the predicted value to reflect the error of the regression model.⁹ If a student had missing data on more than one component, the ESCS score was not computed for the student, and the student's ESCS score was replaced with "99" in the SPSS file and ".M" in the SAS file.

After the imputation process, each of the three components (including the imputed values) was standardised using the OECD mean and standard deviation of the respective component in PISA 2022 (presented in Table 19.65 above). Next, the arithmetic mean of the three standardised components was calculated to create a preliminary ESCS score for each student. Lastly, the preliminary ESCS scores were standardised again using the OECD mean and standard deviation of the preliminary ESCS scores in PISA 2022 (also presented in Table 19.65). This process ensured that the trend ESCS scores produced for PISA 2012, PISA 2015, and PISA 2018 were directly comparable to the ESCS scores produced for PISA 2022.

Financial Literacy Questionnaire derived variables

The Financial Literacy Questionnaire is an international option that countries/economies could choose to implement. It was administered to students after they had completed the Student Questionnaire. It addresses familiarity of students related to financial literacy and their confidence about financial matters. There were 10 variables derived from this questionnaire, including one simple DV and nine IRT scaled DVs. An overview of all DVs in this questionnaire is shown in Table 19.66 and each are described in the following sections.

Simple questionnaire indices

Familiarity with concepts of finance (FCFMLRTY)

Students' ratings of how familiar they were with various financial topics in question FL164 were used to derive an indicator of familiarity with concepts of finance. There were three response options ("Never heard of it", "Heard of it, but I don't recall the meaning", "Learnt about it, and I know what I means"). For each item, a value of "1" was assigned to "Learnt about it, and I know what I means" responses, and all other responses were assigned a value of "0". This index was constructed as the sum of values across all 16 items. Values range from "0" to "16".

Derived variables based on IRT scaling

The Financial Literacy Questionnaire provided data for nine DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.67, the number of items with international parameters for each scale and group are presented in Table 19.68, the number of trend items with international parameters for each trend scale and group are presented in Table 19.69, and the groups that did not administer each scale are presented in Table 19.70 (in this case, the scale scores for the individuals in the group were replaced with "99" in the SPSS file and ".M" in the SAS file). Note that there were no

countries/economies for which the scale scores were suppressed for the scales in the Financial Literacy Questionnaire.

Financial education in school lessons (FLSCHOOL)

Students' frequency ratings of how often they encountered financial tasks and activities in school lessons (e.g., "Describing the purpose and uses of money", "Exploring ways of planning to pay an expense") in question FL166 were scaled into the index of "Financial education in school lessons". Note that this scale was linked to the FLSCHOOL scale in PISA 2018. Each of the six items included in this scale had three response options ("Never", "Sometimes", "Often"). Table 19.71 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Financial education in school lessons – Multiple subjects (FLMULTSB)

Students' responses to questions about where they encountered lessons about financial topics (e.g., "During your mathematics class", "During classes about economics or business") in question FL174 were scaled into the index of "Financial education in school lessons – Multiple Subjects (FLMULTSB)". Each of the seven items included in this scale had two substantive response options ("Yes", "No") and two additional response options ("I don't know.", "I don't have this class.") which were recoded as missing prior to scaling. Table 19.72 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Parental involvement in matters of financial literacy (FLFAMILY)

Students' frequency ratings of how often they discuss various financial issues with their parents (e.g., "Your spending decisions", "Shopping online") in question FL167 were scaled into the index of "Parental involvement in matters of financial literacy". Note that this scale was linked to the FLFAMILY scale in PISA 2018. Each of the seven items included in this scale had four response options ("Never or hardly ever", "Once or twice a month", "Once or twice a week", "Almost every day"). Table 19.73 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Access to money and financial projects – Sources of money (ACCESSFP)

Students' frequency ratings about how often their money came from different sources (e.g., "An allowance or pocket money for doing chores at home", "Working in a family business") in question FL170 were scaled into the index of "Access to money and financial projects – Sources of Money (ACCESSFB)". Each of the seven items included in this scale had five response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day"). Table 19.74 shows the item wording and item parameters for the items in this scale.

Confidence about financial matters (FLCONFIN)

Students' ratings of their confidence with various financial matters (e.g., "Understanding bank statements", "Keeping track of my account balance") in question FL162 were scaled into the index of "Confidence about financial matters". Note that this scale was linked to the FLCONFIN scale in PISA 2018. Each of the six items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Table 19.75 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Confidence about financial matters using digital devices (FLCONICT)

Students' ratings of their confidence in doing various financial tasks with electronic devices (e.g., "Transferring money", "Paying with a mobile device (e.g., mobile phone or tablet) instead of using cash") in question FL163 were scaled into the index of "Confidence about financial matters using digital devices". Note that this scale was linked to the FLCONICT scale in PISA 2018. Each of the five items included in this scale had four response options ("Not at all confident", "Not very confident", "Confident", "Very confident"). Table 19.76 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Access to money and financial products – Financial activities (ACCESSFA)

Students' frequency ratings of how often they completed different financial activities (e.g., "Checked how much money you have", "Saved money at home") in question FL171 were scaled into the index of "Access to money and financial products – Financial activities (ACCESSFA)". Each of the 11 items included in this scale had five response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day"). Table 19.77 shows the item wording and item parameters for the items in this scale.

Attitudes towards and confidence about financial matters (ATTCONFM)

Students' rating of their agreement with different statements about their attitudes towards and confidence about financial matters (e.g., "I enjoy talking about money matters.", "I know how to manage my money.") in question FL169 were scaled into the index of "Attitudes towards and confidence about financial matters (ATTCONFM)". Each of the seven items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.78 shows the item wording and item parameters for the items in this scale.

Friends' influence on financial matters (FRINFLFM)

Students' ratings of their agreement with various statements about their friends' influence on finance decisions (e.g., "My friends have a strong influence on my spending decisions.", "Sometimes I spend more than I would like when I am with my friends.") in question FL172 were scaled into the index of "Friends' influence on financial matters". Each of the four items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.79 shows the item wording and item parameters for the items in this scale.

ICT Familiarity Questionnaire derived variables

The ICT Familiarity Questionnaire is an international option that countries/economies could choose to implement. It was administered to students after they had completed the Student Questionnaire. There were 15 variables derived from this questionnaire, including three simple DVs and 12 IRT scaled DVs. All of the IRT scaled DVs were new for PISA 2022, as the ICT framework had been revised for this cycle. An overview of all DVs in this questionnaire is shown in Table 19.80 and each are described in the following sections.

Simple questionnaire indices

Availability and usage of ICT at school (ICTAVSCH)

The availability of ICT at school was gathered from IC170 where students' frequency ratings of how often they use various digital resources at school (e.g., "Desktop or laptop computer", "Smartphone") was used for the index of "ICT availability at school". Each of the seven items in this question included six response options ("Never or almost never", "About once or twice a month", "About once or twice a week", "Every day or almost every day", "Several times a day", "This resource is not available to me at school"). The index was calculated as the number of all seven items that were marked with a value other than "This resource is not available to me at school", thus ranging from 0-7. Items 2-4 were included in various previous versions of the ICT Questionnaire.

Availability and usage of ICT outside of school (ICTAVHOM)

The availability of ICT outside of school was gathered from IC171 where students' frequency ratings of how often they use various digital resources outside of school (e.g., "Desktop or laptop computer", "Smartphone") was used for the index of "ICT use outside of school". Each of the six items in this question included six response options ("Never or almost never", "About once or twice a month", "About once or twice a week", "Every day or almost every day", "Several times a day", "This resource is not available to me outside of school"). For each of the six items, a score of "0" was assigned when students choose the "This resource is not available to me outside of school" response options and all other responses were coded "1". The index was calculated as the sum of "0" and "1" designations across the six items that were marked with a value other than "This resource is not available to me at school", thus ranging from 0-6. Items 2-4 were included in various previous versions of the ICT Questionnaire.

Distress from online content and cyberbullying (ICTDISTR)

Students' ratings of how upset they were when various situation occurred online (e.g., "Encountering content online that was inappropriate for my age", "Receiving unkind, vulgar or offending messages, comments or videos") in question IC181 were scaled into the index of "Distress from online content and cyberbullying". Each item included five response options ("This did not happen to me", "Not at all upset", "A little upset", "Quite upset", "Very upset"). Values in this index range from 0-16, with "This did not happen to me" recoded as a missing variable, "Not at all upset" coded "1", "A little upset" coded "2", "Quite upset" coded "3", and "Very upset" coded "4". Values across all items were summed.

Derived variables based on IRT scaling

The ICT Familiarity Questionnaire provided data for 12 DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.81, the number of items with international parameters for each scale and group are presented in Table 19.82, the countries/economies for which the scale scores were suppressed for each scale are presented in Table 19.83 (in this case, the scale scores for the individuals in the country/economy were replaced with "97" in the SPSS file and ".N" in the SAS file), and the groups that did not administer each scale are presented in Table 19.84 (in this case, the scale scores for the individuals in the group were replaced with "99" in the SPSS file and ".M" in the SAS file).

ICT availability at school (ICTSCH)

Students' frequency ratings of how often they use various digital resources at school (e.g., "Desktop or laptop computer", "Smartphone (i.e., mobile phone with internet access)") in question IC170 were scaled into the index of "ICT availability at school". Each of the seven items included in this scale had six response options ("Never or almost never", "About once or twice a month", "About once or twice a week", "Every day

or almost every day”, “Several times a day”, “This resource is not available to me at school”). “This resource is not available to me at school” was recoded as 0, while the five other response options were recoded as 1 prior to scaling. Table 19.85 shows the item wording and item parameters for the items in this scale.

ICT availability outside school (ICTHOME)

Students’ frequency ratings of how often they use various digital resources outside of school (e.g., “Desktop or laptop computer”, “Smartphone (i.e., mobile phone with internet access)”) in question IC171 were scaled into the index of “ICT availability outside school”. Each of the six items included in this scale had six response options (“Never or almost never”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”, “Several times a day”, “This resource is not available to me outside of school”). “This resource is not available to me outside of school” was recoded as 0, while the five other response options were recoded as 1 prior to scaling. Table 19.86 shows the item wording and item parameters for the items in this scale.

Quality of access to ICT (ICTQUAL)

Students’ ratings of their agreement with various statements about ICT resources at their school (e.g., “There are enough digital devices with access to the Internet at my school.”, “The school’s Internet speed is sufficient.”) in question IC172 were scaled into the index of “Quality of access to ICT”. Each of the nine items included in this scale had four response options (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”). Table 19.87 shows the item wording and item parameters for the items in this scale.

Subject-related ICT use during lessons (ICTSUBJ)

Students’ frequency ratings of how often digital resources are used in various subject lessons (e.g., “Mathematics”, “Science”) in question IC173 were scaled into the index of “Subject-related ICT use during lessons”. Each of the four items included in this scale had five substantive response options (“Never or almost never”, “In less than half of the lessons”, “In about half of the lessons”, “In more than half of the lessons”, “In every or almost every lesson”) and an additional response option “I do not have this subject” which was recoded as missing prior to scaling. Table 19.88 shows the item wording and item parameters for the items in this scale.

Use of ICT in enquiry-based learning activities (ICTENQ)

Students’ frequency ratings of how often they use digital resources for various school-related activities (e.g., “Create a multi-media presentation with pictures, sound or video”, “Track the progress of your own work or projects”) in question IC174 were scaled into the index of “Use of ICT in enquiry-based learning activities”. Each of the 10 items included in this scale had five response options (“Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”). Table 19.89 shows the item wording and item parameters for the items in this scale.

Support or feedback via ICT (ICTFEED)

Students’ frequency ratings of how often they use digital resources in various activities related to support or feedback (e.g., “Read or listen to feedback sent by my teachers regarding my work and academic results”, “Read or listen to feedback sent by other students on my work”) in question IC175 were scaled into the index of “Support or feedback via ICT”. Each of the four items included in this scale had five response options (“Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, “Every day or almost every day”). Table 19.90 shows the item wording and item parameters for the items in this scale.

Use of ICT for school activities outside of the classroom (ICTOUT)

Students' frequency ratings of how often they use digital resources for various school-related activities outside of the classroom (e.g., "See my grades or results from specific assignments (e.g., homework or tests)", "Communicate with my teacher") in question IC176 were scaled into the index of "Use of ICT for school activities outside of the classroom". Each of the eight items included in this scale had five response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day"). Table 19.91 shows the item wording and item parameters for the items in this scale.

Frequency of ICT activity – Weekday (ICTWKDY)

Students' frequency ratings of how often they did various leisure activities using ICT during a typical week day (e.g., "Play video-games (using my smartphone, a gaming console or an online platform or apps)", "Look for practical information online (e.g., find a place, book a train ticket, buy a product)") in question IC177 were scaled into the index of "Frequency of ICT activity – Weekday". Each of the seven items included in this scale had six response options ("No time at all", "Less than 1 hour a day", "Between 1 and 3 hours a day", "More than 3 hours and up to 5 hours a day", "More than 5 hours and up to 7 hours a day", "More than 7 hours a day"). Table 19.92 shows the item wording and item parameters for the items in this scale.

Frequency of ICT activity – Weekend (ICTWKEND)

Students' frequency ratings of how often they did various leisure activities using ICT during a typical weekend day (e.g., "Play video-games (using my smartphone, a gaming console or an online platform or apps)", "Look for practical information online (e.g., find a place, book a train ticket, buy a product)") in question IC178 were scaled into the index of "Frequency of ICT activity – Weekend". Each of the seven items included in this scale had six response options ("No time at all", "Less than 1 hour a day", "Between 1 and 3 hours a day", "More than 3 hours and up to 5 hours a day", "More than 5 hours and up to 7 hours a day", "More than 7 hours a day"). Table 19.93 shows the item wording and item parameters for the items in this scale.

Views of regulated ICT use in school (ICTREG)

Students' ratings of their agreement with various statements about regulation of ICT use at school (e.g., "Students should not be allowed to bring mobile phones to class.", "The school should set up filters to prevent students from playing games online.") in question IC179 were scaled into the index of "Views of regulated ICT use in school". Each of the six items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.94 shows the item wording and item parameters for the items in this scale.

Students' practices regarding online information (ICTINFO)

Students' ratings of their agreement with various statements about their practices regarding online information (e.g., "When searching for information online I compare different sources.", "I discuss the accuracy of online information with friends or other students.") in question IC180 were scaled into the index of "Students' practices regarding online information". Each of the six items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.95 shows the item wording and item parameters for the items in this scale.

Self-efficacy in digital competencies (ICTEFFIC)

Students' ratings of how well they can do various tasks using digital resources (e.g., "Search for and find relevant information online", "Write or edit text for a school assignment") in question IC183 were scaled into the index of "Self-efficacy in digital competencies". Each of the 14 items included in this scale had four substantive response options ("I cannot do this", "I struggle to do this on my own", "I can do with a bit of effort", "I can easily do this") and an additional response option "I don't know what this is" which was recoded as missing prior to scaling. Table 19.96 shows the item wording and item parameters for the items in this scale.

Well-Being Questionnaire derived variables

The Well-Being Questionnaire is an international option that countries/economies could choose to implement. It was administered to students after they had completed the Student Questionnaire. It addresses the well-being of students. There were seven variables derived from this questionnaire, including one simple DV and six IRT scaled DVs. An overview of all DVs in this questionnaire is shown in Table 19.97 and each are described in the following sections.

Simple questionnaire indices

Body mass index (STUBMI)

The only simple DV from the Well-Being Questionnaire is STUBMI, indicating the student's body mass index (STUBMI). It is based on two questions, WB151 and WB152, which asked about the weight and the height of the student, respectively, in the units of measurement that are more common in the respective country/economy. The index is constructed as it was in PISA 2018. Specifically, the index was constructed as the weight (transformed to kilograms) divided by the square of the body height (transformed to metres).

Derived variables based on IRT scaling

The Well-Being Questionnaire provided data for six DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.98, the number of items with international parameters for each scale and group are presented in Table 19.99, the number of trend items with international parameters for each trend scale and group are presented in Table 19.100, and the groups that did not administer each scale are presented in Table 19.101 (in this case, the scale scores for the individuals in the group were replaced with "99" in the SPSS file and ".M" in the SAS file). Note that there were no countries/economies for which the scale scores were suppressed for the scales in the Well-Being Questionnaire.

Body image (BODYIMA)

Students' ratings of their agreement with statements about their body image (e.g., "I like my look just the way it is.", "I like my body.") in question WB153 were scaled into the index of "Body image". Note that this scale was linked to the BODYIMA scale in PISA 2018. Each of the five items included in this scale had four substantive response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree") and an additional response option "I don't have an opinion" which was recoded as missing prior to scaling. Table 19.102 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Social connection to parents (SOCONPA)

Students' ratings of how often their parents engage in various activities (e.g., "Show that they care", "Encourage me to make my own decisions") in question WB163 were scaled into the index of "Social connection to parents". Note that this scale was linked to the SOCONPA scale in PISA 2018. Each of the six items included in this scale had three response options ("Almost never", "Sometimes", "Almost always"). Table 19.103 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Students' life satisfaction across domains (LIFESAT)

Students' ratings of their satisfaction with different areas of their lives (e.g., "Your health", "The neighbourhood you live in") in question WB155 were scaled into the index of "Students' life satisfaction across domains". Each of the 10 items included in this scale had four response options ("Not at all satisfied", "Not satisfied", "Satisfied", "Totally satisfied"). Table 19.104 shows the item wording and item parameters for the items in this scale.

Psychosomatic symptoms (PSYCHSYM)

Students' ratings of how often they experienced different psychosomatic symptoms (e.g., "Headache", "Stomach pain") in question WB154 were scaled into the index of "Psychosomatic symptoms". Each of the nine items included in this scale had five response options ("Rarely or never", "About every month", "About every week", "More than once a week", "About every day"). Table 19.105 shows the item wording and item parameters for the items in this scale.

Social connections: Ease of communication about worries and concerns (SOCCON)

Students' ratings of how easy it is to communicate about their worries and concerns with different people (e.g., "Your father", "Your brother(s)") in question WB162 were scaled into the index of "Social connections: Ease of communication about worries and concerns". Each of the nine items included in this scale had four substantive response options ("Very difficult", "Difficult", "Easy", "Very Easy") and an additional response option "I don't have or see this person" which was recoded as missing prior to scaling. Table 19.106 shows the item wording and item parameters for the items in this scale.

Experienced well-being – Previous day (EXPWB)

Students' responses regarding their experienced well-being in the previous day (e.g., "Were you treated with respect all day yesterday?", "Did you smile or laugh a lot yesterday?") in question WB178 were scaled into the index of "Experienced well-being – Previous day". Each of the six items included in this scale had two response options ("Yes", "No"). Note that prior to scaling, all responses were recoded as missing if the student did not respond "yes" to WB178Q07JA ("Was yesterday a typical day?"). Table 19.107 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Parent Questionnaire derived variables

The Parent Questionnaire is an international option that countries/economies could choose to implement. It was administered to the parents of students participating in the PISA assessment. There were 12 variables derived from this questionnaire, including one simple DV and 11 IRT scaled DVs. An overview of all DVs in this questionnaire is shown in Table 19.108 and each are described in the following sections.

Simple questionnaire indices

Parents' expectations in child's future educational career (PAREXPT)

Parents' responses to a list of possible educational levels they expected their children to complete in question PA183 were transformed into the index of "Parents' expectations in child's future educational career". The categories were specified using country-specific terms that were understood by the respondents. Each qualification was mapped to the ISCED classification of educational levels [see *ISCED 2011 Operational Manual: Guidelines for Classifying National Educational Programmes and Related Qualifications* (OECD/Eurostat/UNESCO Institute for Statistics, 2015_[10])]. Values on the index ranged from "Less than ISCED level 2" to "ISCED level 8" and scores were assigned as noted in Table 19.109.

Derived variables based on IRT scaling

The Parent Questionnaire provided data for 11 DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.110, the number of items with international parameters for each scale and group are presented in Table 19.111, the number of trend items with international parameters for each trend scale and group are presented in Table 19.112, the countries/economies for which the scale scores were suppressed for each scale are presented in Table 19.113 (in this case, the scale scores for the individuals in the country/economy were replaced with "97" in the SPSS file and ".N" in the SAS file), and the groups that did not administer each scale are presented in Table 19.114 (in this case, the scale scores for the individuals in the group were replaced with "99" in the SPSS file and ".M" in the SAS file).

Current parental/guardian support (CURSUPP)

Parents' frequency ratings of how often they or someone else in their home provides education-related support (e.g., "Discuss how well my child is doing at school," "Talk to my child about any problems he/she may have at school") in question PA003 were scaled into the index of "Current parental/guardian support". Note that this scale was linked to the CURSUPP scale in PISA 2018. Each of the 14 items included in this scale had five response options ("Never or hardly ever", "Once or twice a year", "Once or twice a month", "Once or twice a week", "Every day or almost every day"). Table 19.115 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Parent attitudes toward mathematics (PQMIMP)

Parents' ratings of their agreement with statements about the importance of mathematical knowledge (e.g., "Most jobs today require some mathematics knowledge and skills.", "It is an advantage in the job market to have good mathematics knowledge and skills.") in question PA196 were scaled into the index of "Parent attitudes toward mathematics". Note that this scale was linked to the PQMIMP scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Each of the four items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.116 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items and which items were reverse-coded prior to scaling.

Mathematics career (PQMCAR)

Parents' responses to questions about mathematics-related careers (e.g., "Does anybody in your family (including you) work in a <mathematics-related career>?", "Does your child show an interest in working in a <mathematics-related career>?") in question PA197 were scaled into the index of "Mathematics career". Note that this scale was linked to the PQMCAR scale in PISA 2012 and was scaled using the PCM, in line with the model used in PISA 2012. Each of the five items included in this scale had two response options

("Yes", "No"). Table 19.117 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items and which items were reverse-coded prior to scaling.

Parental involvement (PARINVOL)

Parents' responses to questions about their involvement in their child's schooling in the past year (e.g., "Discussed my child's progress with a teacher on my own initiative", "Attended a scheduled meeting or conferences for parents") in question PA008 were scaled into the index of "Parental involvement". Each of the 10 items included in this scale had two substantive response options ("Yes", "No") and an additional response option "Not supported by school" which was recoded as missing prior to scaling. Table 19.118 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

School quality (PQSCHOOL)

Parents' ratings of their agreement with statements about school quality (e.g., "Most of my child's school teachers seem competent and dedicated.", "Standards of achievement are high in my child's school.") in question PA007 were scaled into the index of "School quality". Note that this scale was linked to the PQSCHOOL scale in PISA 2018. Each of the seven items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.119 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which items are trend items.

School policies for parental involvement (PASCHPOL)

Parents' ratings of their agreement with statements about school policies for parental involvement (e.g., "My child's school provides effective communication between the school and families.", "My child's school involves parents in the school's decision-making process.") in question PA007 were scaled into the index of "School policies for parental involvement". Note that this scale was linked to the PASCHPOL scale in PISA 2018. Each of the six items included in this scale had four response options ("Strongly agree", "Agree", "Disagree", "Strongly disagree"). Table 19.120 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which items are trend items.

Parents' attitudes towards immigrants (ATTIMMP)

Parents' ratings of their agreement with statements about immigrants (e.g., "Immigrant children should have the same opportunities for education that other children in the country have.", "Immigrants who live in a country for several years should have the opportunity to vote in elections.") in question PA167 were scaled into the index of "Parents' attitudes towards immigrants". Note that this scale was linked to the ATTIMMP scale in PISA 2018. Each of the four items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.121 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Creative home environment (CREATHME)

Parents' ratings of their agreement with statements about creativity in the home environment (e.g., "In our family, we encourage participating in extra-curricular activities that require creativity.", "At home, we try to fix things that are broken.") in question PA185 were scaled into the index of "Creative home environment". Each of the nine items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.122 shows the item wording and item parameters for the items in this scale.

Participation in creative activities outside of school (CREATACT)

Parents' ratings of how often their child participated in creative activities outside of school (e.g., "Art classes/activities (e.g., painting, drawing)", "Debate club") in question PA186 were scaled into the index of "Participation in creative activities outside of school". Each of the eight items included in this scale had five substantive response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day") and an additional response option "Not available" which was recoded as missing prior to scaling. Table 19.123 shows the item wording and item parameters for the items in this scale.

Creativity and openness to intellect (CREATOPN)

Parents' ratings of their agreement with statements regarding their views on their own creativity and openness to intellect (e.g., "I am very creative.", "I enjoy projects that require creative solutions.") in question PA188 were scaled into the index of "Creativity and openness to intellect". Each of the nine items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.124 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Openness to creativity: Other's report (CREATOR)

Parents' ratings of their agreement with statements regarding their views about their child's creativity (e.g., "My child is very creative.", "My child enjoys projects that require creative solutions.") in question PA189 were scaled into the index of "Openness to creativity: Other's report". Each of the eight items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.125 shows the item wording and item parameters for the items in this scale.

School Questionnaire derived variables

The School Questionnaire consisted mainly of questions used in previous cycles. There were 57 variables derived from this questionnaire, including 32 simple DVs and 25 IRT scaled DVs. An overview of all DVs in this questionnaire is shown in Table 19.126 and each are described in the following sections. The simple and scaled DVs are organised first by framework module (please see Chapter 5 for a description of modules) and then alphabetical order within modules.

Simple questionnaire indices

Out-of-school experiences (Module 10)

Creative extra-curricular activities (CREACTIV)

School principals were asked in SC053 to report what extra-curricular activities their schools offered to 15-year-old students. The two response categories were "Yes" and "No" for the 10 items. The index of creative extra-curricular activities at school (CREACTIV) was computed as the total number of the following 3 activities that occurred at school: i) band, orchestra or choir (SC053Q01TA); ii) school play or school musical (SC053Q02TA); and iii) art club or art activities (SC053Q09TA). The index ranges from 0 to 3. Additionally, a separate DV (SC053D11TA) combines all the customizations across countries to SC053C11TA (please see Annex D).

Mathematics extension courses offered at school (MATHEXC)

School principals were asked in SC181 to report what additional mathematics lessons are offered at their school. The two response categories were “Yes” and “No”. The index of mathematics extension course offered at school (MATHEXC) was computed from SC181 by assigning schools to one of three different categories based on the type of additional mathematics lessons offered at the school. Schools that responded “Yes” to offering additional mathematics courses without differentiation based on prior achievement (SC181Q03JA) and “No” to offering enrichment (SC181Q01JA) and remedial (SC181Q02JA) mathematics classes were assigned a ‘1’. Schools that responded “Yes” to offering either enrichment mathematics lessons or remedial mathematics lessons were assigned a ‘2’. Schools that responded “Yes” to offering both enrichment and remedial mathematics classes were assigned a ‘3’.

Mathematics-related extra-curricular activities at school (MACTIV)

School principals were asked in SC053 to report what mathematics-related extra-curricular activities their schools offered to 15-year-old students. The two response categories were “Yes” and “No”. The index of mathematics-related extra-curricular activities at school (MACTIV) was computed as follows. First the question SC181 was assigned the value of ‘1’ if “Yes” was selected for “Enrichment” (SC181Q01JA), “Remedial” (SC181Q02JA), or “Without differentiation depending on the prior achievement level of the students” (SC181Q03JA). SC181 was assigned the value of ‘2’ if “Yes” was selected for both “Enrichment” and “Remedial”. Second, each of three items about a mathematics club (SC053Q05NA), mathematics competitions (SC053Q06NA), or club with a focus on computers (SC053Q08TA) was assigned the value of ‘1’ if a school selected “Yes” to these activities. If a school did not offer one of these three activities (i.e., selected “No”), the corresponding variable received the value of ‘0’. Third, these recoded variables were summed up to result in a range of 0 to 5 for MACTIV. For example, if the purpose of additional lessons was both “Enrichment” and “Remedial” and the school offered a mathematics club, but not mathematics competitions or a club with a focus on computers, the value of MACTIV was coded as “3”.

School type and infrastructure (Module 11)

Availability of computers (RATCMP1)

School principals were asked in SC004 to report the number of digital devices available for 15-year-old students at their school. The index of availability of computers (RATCMP1) is the ratio of the number of desktop or laptop computers available for these students for educational purposes (SC004Q02TA) to the total number of students in the modal grade for 15-year-olds at their school (SC004Q01TA).

Availability of tablet devices (RATTAB)

School principals were asked in SC004 to report the number of tablet devices or e-book readers available for 15-year-old students at their school for educational purposes (SC004Q08JA). The index of availability of tablet devices (RATTAB) is the ratio of the number of tablet devices available for these students for educational purposes to the total number of students in the modal grade for 15-year-olds at their school (SC004Q01TA).

Computers connected to the Internet (RATCMP2)

School principals were asked in SC004 to report the number of desktop or laptop computers at their school that are connected to the Internet. The index of computers connected to the Internet (RATCMP2) is the ratio of the number of desktop or laptop computers available for 15-year-olds for educational purposes (SC004Q02TA) to the number of these computers that are connected to the Internet (SC004Q03TA).

Proportion of personnel for pedagogical support (PROPSUPP)

Principals were asked in SC168 to report the number of personnel for pedagogical support currently working in their school (SC168Q01JA). The proportion of personnel for pedagogical support (PROPSUPP) was calculated by dividing the number of these personnel by the total number of non-teaching staff at the school (TOTSTAFF).

Proportion of school administrative personnel (PROADMIN)

Principals were asked in SC168 to report the number of school administrative personnel currently working in their school (SC168Q02JA). The proportion of school administrative personnel (PROADMIN) was calculated by dividing the number of these personnel by the total number of non-teaching staff at the school (TOTSTAFF).

Proportion of school management personnel (PROMGMT)

Principals were asked in SC168 to report the number of school management personnel currently working in their school (SC168Q03JA). The proportion of school management personnel (PROMGMT) was calculated by dividing the number of these personnel by the total number of non-teaching staff at the school (TOTSTAFF).

Proportion of other non-teaching staff (PROOSTAF)

Principals were asked in SC168 to report the number of other non-teaching staff (i.e., not personnel for pedagogical support, school administrative personnel, or school management personnel) currently working in their school (SC168Q04JA). The proportion of other non-teaching staff (PROOSTAF) was calculated by dividing the number of these personnel by the total number of non-teaching staff at the school (TOTSTAFF).

School size (SCHSIZE)

The index of school size (SCHSIZE) contains the total enrolment at school. It is based on the enrolment data provided by the school principal in SC002, summing the number of girls (SC002Q02TA) and boys (SC002Q01TA) at a school.

School type (SCHLTYPE)

Schools are classified as either public or private according to whether a private entity or a public agency has the ultimate power for decision making concerning its affairs. As in previous PISA surveys, the index of school type (SCHLTYPE) was constructed by recoding SC013 and SC016. SC013 asks whether the school is public or private, and SC016 asks about the source and proportion of resources (government, resources from students/parents, benefactors/donations, or other). SCHLTYPE has the following three categories:

1. Private independent (if SC013Q01TA=2 and SC016Q01TA < 50), or (SC013Q01TA=2 and SUM(SC016Q02TA, SC016Q03TA, SC016Q04TA) >=50).
2. Private Government-dependent (if SC013Q01TA=2 and SC016Q01TA >=50), or
3. Public (if SC013Q01TA=1).

Since PISA 2018, PRIVATESCH was created from sampling information in order to improve the public/private indicators. If SC013 is missing, PRIVATESCH is used to create SCHLTYPE.

Similar to 2018, IRL had special treatment for this designation – based solely on the STRATUM sampling variable.

Student-teacher ratio (STRATIO)

The student-teacher ratio (STRATIO) was obtained by dividing the number of enrolled male and female students (SCHSIZE) provided by the principal in SC002 (SC002Q01TA, SC002Q02TA) by the total number of full-time and part-time teachers (TOTAT) provided by the principal in SC018 (SC018Q01TA01, SC018Q01TA02).

Student-mathematics teacher ratio (SMRATIO)

The student-mathematics teacher ratio (SMRATIO) was obtained by dividing the number of enrolled male and female students (SCHSIZE) provided by the principal in SC002 (SC002Q01TA, SC002Q02TA) by the total number of full-time and part-time mathematics teachers (TOTMATH) provided by the principal in SC182 (SC182Q01WA01, SC182Q01WA02).

Total number of mathematics teachers at school (TOTMATH)

Principals were asked in SC182 to report the number of full-time and part-time mathematics teachers at their school (SC182Q01WA01, SC182Q01WA02) and provide additional information on how many of the staff was full-time and part-time employed mathematics teachers qualified at different ISCED levels. The total number of mathematics teachers at the school (TOTMATH) was computed as the sum of full-time and part-time mathematics teachers.

Total number of non-teaching staff at school (TOTSTAFF)

Principals were asked in SC168 to report the number of non-teaching staff currently working in their school. The total number of non-teaching staff at the school (TOTSTAFF) is a sum of the numbers of personnel for pedagogical support (SC168Q01JA), school administrative personnel (SC168Q02JA), school management personnel (SC168Q03JA), and other non-teaching staff (SC168Q04JA).

Total number of all teachers at school (TOTAT)

Principals were asked in SC018 to report the total number of full-time and part-time teachers at their school (SC018Q01TA01, SC018Q01TA02) and provide additional information on how many of the staff was full-time and part-time employed teachers qualified at different ISCED levels.

Selection and enrolment (Module 12)

School selectivity (SCHSEL)

Principals were asked in SC012 about admittance policies at their school, including student academic performance and recommendation by feeder schools. The three response categories for this question were “Never”, “Sometimes”, and “Always”. An index of academic school selectivity (SCHSEL) was computed by assigning schools to one of three categories based on how often two factors, namely “Student’s record of academic performance” (SC012Q01TA) and “Recommendation of feeder schools” (SC012Q02TA), were considered when admitting students to the school as follows:

1. the two factors (student’s record of academic performance and recommendation of feeder schools) were never considered (if SC012Q01TA=1 and SC012Q02TA=1),
2. at least one of the factors was considered sometimes but neither always (if SC012Q01TA=2 or SC012Q02TA=2, and if SC012Q01TA<3 and SC012Q02TA<3), and
3. at least one of the factors was considered always (if SC012Q01TA=3 or SC012Q02TA=3).

School autonomy (Module 13)

School responsibility for curriculum (SRESPCUR)

Principals were asked in SC202 about who had the main responsibility for various decisions or activities at their school. The six response categories for this question were “Principal”, “Teachers or members of <school management team>”, “<School governing board>”, “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. An index of the relative level of responsibility of school staff in deciding issues related to curriculum and assessment (RESPCUR) was computed from the school principals’ reports regarding who had the main responsibility for 4 items in SC202. The index was calculated on the basis of the ratio of responses for “Principal”, “Teachers or members of <school management team>”, or “<School governing board>” on the one hand to responses for “<Local or municipal authority>”, “<Regional or state authority>”, or “<National or federal authority>” on the other hand. In the first step, a count for school responsibility was calculated by counting the number of “Principal”, “Teachers or members of <school management team>”, and “<School governing board>” responses. In the second step, a count for non-school responsibility was calculated by counting the number of “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. In the third step, the school responsibility count was divided by the non-school responsibility count. To avoid dividing by “0”, “1” was added to both the numerator and denominator; when the ratio of school responsibility to non-school responsibility was 4:0, an index value of 4 was assigned. Higher values indicated relatively higher levels of school responsibility in deciding issues related to curriculum and assessment.

School responsibility for resources (SRESPRES)

Principals were asked in SC202 about who had the main responsibility for various decisions or activities at their school. The six response categories for this question were “Principal”, “Teachers or members of <school management team>”, “<School governing board>”, “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. An index of the relative level of responsibility of school staff in deciding issues related to allocating resources (RESPRES) was computed from the school principals’ reports regarding who had the main responsibility for 6 items in SC202. The index was calculated on the basis of the ratio of responses for “Principal”, “Teachers or members of <school management team>”, or “<School governing board>” on the one hand to responses for “<Local or municipal authority>”, “<Regional or state authority>”, or “<National or federal authority>” on the other hand. In the first step, a count for school responsibility was calculated by counting the number of “Principal”, “Teachers or members of <school management team>”, and “<School governing board>” responses. In the second step, a count for non-school responsibility was calculated by counting the number of “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. In the third step, the school responsibility count was divided by the non-school responsibility count. To avoid dividing by “0”, “1” was added to both the numerator and denominator; when the ratio of school responsibility to non-school responsibility was 6:0, an index value of 6 was assigned. Higher values on the scale indicated relatively higher levels of school responsibility in this area.

Organisation of student learning at school (Module 14)

Ability grouping for mathematics classes (ABGMATH)

School principals were asked in SC187 to report the extent to which their mathematics classes catered to students with different abilities. The three response categories were “For all classes”, “For some classes”, and “Not for any classes”. An index of ability grouping between mathematics classes (ABGMATH) was derived from the first two items (SC187Q01WA, SC187Q02WA) by assigning schools to three categories: (1) schools with no ability grouping for any classes, (2) schools with one of these forms of ability grouping between some classes and (3) schools with one of these forms of ability grouping for all classes.

Class size (CLSIZE)

Principals were asked in SC003 about the average size of the test language classes in their school. The nine response categories were “15 students or fewer”, “16-20 students”, “21-25 students”, “26-30 students”, “31-35 students”, “36-40 students”, “41-45 students”, “46-50 students”, and “More than 50 students”. The average class size (CLSIZE) was derived from the midpoint of each response category, resulting in a value of 13 for the lowest category, and a value of 53 for the highest.

Math class size (MCLSIZE)

Principals were asked in SC176 about the average class size of mathematics classes in their school. The nine response categories were “15 students or fewer”, “16-20 students”, “21-25 students”, “26-30 students”, “31-35 students”, “36-40 students”, “41-45 students”, “46-50 students”, and “More than 50 students”. The average math class size (TBD) was derived from the midpoint of each response category, resulting in a value of 13 for the lowest category, and a value of 53 for the highest.

Teacher qualification, training, and professional development (Module 17)

Proportion of all teachers fully certified (PROATCE)

Principals were asked in SC018 to report the number of full-time and part-time teachers fully certified by the appropriate authority (SC018Q02TA01, SC018Q02TA02). The proportion of fully certified teachers (PROATCE) was computed by dividing the number of fully certified teachers by the total number of teachers (TOTAT).

Proportion of all teachers with at least ISCED level 6 bachelor qualification (PROPAT6)

Principals were asked in SC018 to report the number of full-time and part-time teachers with an ISCED level 6 (Bachelor’s or equivalent level) qualification (SC018Q08JA01, SC018Q08JA02). The proportion of teachers with *at least* an ISCED 6 Bachelor qualification (PROPAT6) was calculated by dividing the number full-time and part-time teachers with an ISCED level 6 (Bachelor’s or equivalent level) qualification (SC018Q08JA01, SC018Q08JA02), ISCED level 7 (Master’s or equivalent level) qualification (SC018Q09JA01, SC018Q09JA02), and ISCED level 8 (Doctoral or equivalent level) qualification (SC018Q10JA01, SC018Q10JA02) by the total number of teachers (TOTAT).

Proportion of all teachers with at least ISCED level 7 master qualification (PROPAT7)

Principals were asked in SC018 to report the number of full-time and part-time teachers with an ISCED level 7 (Master’s or equivalent level) qualification (SC018Q09JA01, SC018Q09JA02). The proportion of teachers with *at least* an ISCED 7 Master qualification (PROPAT7) was calculated by dividing the number of full-time and part-time teachers with an ISCED level 7 (Master’s or equivalent level) qualification (SC018Q09JA01, SC018Q09JA02) and ISCED level 8 (Doctoral or equivalent level) qualification (SC018Q10JA01, SC018Q10JA02) by the total number of teachers (TOTAT).

Proportion of all teachers with ISCED level 8 doctoral qualification (PROPAT8)

Principals were asked in SC018 to report the number of full-time and part-time teachers with an ISCED level 8 (Doctoral or equivalent level) qualification (SC018Q10JA01, SC018Q10JA02). The proportion of teachers with an ISCED 8 Doctoral qualification (PROPAT8) was calculated by dividing the number of these teachers by the total number of teachers (TOTAT).

Proportion of mathematics teachers at school (PROPMATH)

The proportion of mathematics teachers (PROPMATH) was computed as the total number of full-time and part-time mathematics teachers at their school (TOTMATH) provided by the principal in SC182 (SC182Q01WA01, SC182Q01WA02), divided by the total number of teachers at their school (TOTAT) provided by the principal in SC018 (SC018Q01TA01, SC018Q01TA02).

Global crises (Module 21)

School closure support from education authorities (SCSUPRTD)

School administrators' responses to three items in SC222 comprise the index on school closure support from education authorities. School administrators who indicated that their school was closed for one or more school days because of COVID-19 were asked to rate the extent that they felt their school was supported by educational authorities during the time that their school building was closed to students because of COVID-19. They reported this information by selecting one of four response options: "Not at all"; "Very little"; "To some extent"; "A lot". Respondents may interpret support broadly to include any kind of assistance (i.e., financial support, volunteer support, etc). If school administrators chose "A lot" to any of the three items, they received a value of "2" on the index. If school administrators chose "Not at all" to all three items, they received a value of "0" on the index. All other responses received a value of "1". The values of the index range from 0-2. This variable was skipped for respondents who reported that their schools had not been closed for COVID-19 on question SC213.

School closure support from other sources (SCSUPRT)

School administrators' responses to two items in SC222 comprise the index on school closure support from other sources. School administrators who indicated that their school was closed for one or more school days because of COVID-19 were asked to rate the extent that they felt their school was supported by students' parents or guardians and by private donors during the time that their school building was closed to students because of COVID-19. They reported this information by selecting one of four response options: "Not at all"; "Very little"; "To some extent"; "A lot". Respondents may interpret support broadly to include any kind of assistance (i.e., financial support, volunteer support, etc). If school administrators chose "A lot" to either of the two items, they received a value of "2" on the index. If school administrators chose "Not at all" to both items, they received a value of "0" on the index. All other responses received a value of "1". The values of the index will range from 0-2. This variable was skipped for respondents who reported that their schools had not been closed for COVID-19 on question SC213.

Derived variables based on IRT scaling

The School Questionnaire provided data for 25 DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.127, the number of items with international parameters for each scale and group are presented in Table 19.128, the number of trend items with international parameters for each trend scale and group are presented in Table 19.129, the countries/economies for which the scale scores were suppressed for each scale are presented in Table 19.130, and the groups that did not administer each scale are presented in Table 19.131.

School culture and climate (Module 6)

Negative school climate (NEGSCCLIM)

Principals were asked in SC172 about the extent of problem behaviours that contribute to a negative school climate in their school (e.g., "Profanity", "Vandalism"). The four response categories for the six items in the

scale were “Not at all”, “Small extent”, “Moderate extent”, and “Large extent”. Higher scale score values indicate that problem behaviours contribute to a negative school climate to a greater extent, while lower scale score values indicate that problem behaviours impact school climate to a lesser extent. Table 19.132 shows the item wording and item parameters for the items in this scale.

School diversity and multi-cultural views (DMCVIEWS)

Principals were asked in SC173 about the school staff’s efforts to promote a diversity-oriented culture and climate during the last academic year (e.g., “They encouraged students of different backgrounds to resolve disagreements by finding common ground.”, “They taught students how to respond to discrimination.”). The five response categories for the six items in the scale were “Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, and “Every day or almost every day”. Higher scale score values indicate that diversity-related views were encouraged in the school with greater frequency, while lower scale score values indicate that diversity-related views were encouraged with lesser frequency. Table 19.133 shows the item wording and item parameters for the items in this scale.

Student-related factors affecting school climate (STUBEHA)

Principals were asked in SC061 about the extent to which student learning is hindered by student behaviours (e.g., “Student truancy”, “Student use of alcohol or illegal drugs”). Note that this scale was linked to the STUBEHA scale in PISA 2018. The four response categories for the six items in the scale were “Not at all”, “Very little”, “To some extent”, and “A lot”. Higher scale score values indicate that students’ learning is hindered to a greater extent by negative student behaviours, while lower scale score values indicate that students’ learning is hindered by negative student behaviours to a lesser extent. Table 19.134 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Teacher-related factors affecting school climate (TEACHBEHA)

Principals were asked in SC061 about the extent to which student learning is hindered by teacher behaviours (e.g., “Teacher absenteeism”, “Staff resisting change”). The four response categories for the five items in the scale were “Not at all”, “Very little”, “To some extent”, and “A lot”. Higher scale score values indicate that students’ learning is hindered to a greater extent by negative teacher behaviours, while lower scale score values indicate that students’ learning is hindered by negative teacher behaviours to a lesser extent. Table 19.135 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Out-of-school experiences (Module 10)

Extra-curricular activities offered (ALLACTIV)

School principals were asked in SC053 to report what extra-curricular activities their schools offered to 15-year-old students (e.g., “School play or school musical”, “Mathematics club”). The two response categories for the 10 items in the scale were “Yes” and “No”. Higher scale score values indicate that more extra-curricular activities were offered by the school, while lower scale score values indicate that fewer extra-curricular activities were offered. Table 19.136 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

School type and infrastructure (Module 11)

Shortage of educational material (EDUSHORT)

Principals were asked in SC017 about the extent to which instruction is hindered by a shortage of educational materials in their school (e.g., “A lack of educational material (e.g., textbooks, IT equipment, library or laboratory material)”, “Inadequate or poor quality educational material (e.g. textbooks, IT equipment, library or laboratory material)”). Note that this scale was linked to the EDUSHORT scale in PISA 2018. The four response categories for the four items in the scale were “Not at all”, “Very little”, “To some extent”, and “A lot”. Higher scale score values indicate that the school is impacted by a shortage of educational materials to a greater extent, while lower scale score values indicate that the school is impacted to a lesser extent by a shortage of educational materials. Table 19.137 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Shortage of educational staff (STAFFSHORT)

Principals were asked in SC017 about the extent to which instruction is hindered by a shortage of educational staff in their school (e.g., “A lack of teaching staff”, “Inadequate or poorly qualified assisting staff”). Note that this scale was linked to the STAFFSHORT scale in PISA 2018. The four response categories for the four items in the scale were “Not at all”, “Very little”, “To some extent”, and “A lot”. Higher scale score values indicate that the school is impacted by a shortage of educational staff to a greater extent, while lower scale score values indicate that the school is impacted to a lesser extent by a shortage of educational staff. Table 19.138 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

School autonomy (Module 13)

Educational leadership (EDULEAD)

Principals were asked in SC201 about how often they or other members of their school management team engaged in activities or behaviours related to educational leadership during the past 12 months (e.g., “Collaborating with teachers to solve classroom discipline problems”, “Providing parents or guardians with information on the school and student performance”). The five response categories for the seven items in the scale were “Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, and “Every day or almost every day”. Higher scale score values indicate higher frequencies of engagement by the principal and school management team in educational leadership activities, while lower scale values indicate lower frequencies of engagement by the principal and school management team in educational leadership activities. Table 19.139 shows the item wording and item parameters for the items in this scale.

Instructional leadership (INSTLEAD)

Principals were asked in SC201 about how often they or other members of their school management team engaged in activities or behaviours related to teaching or instructional leadership during the last 12 months (e.g., “Providing feedback to teachers based on observations of instruction in the classroom”, “Taking actions to ensure that teachers feel responsible for their students' learning outcomes”). The five response categories for the five items in the scale were “Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, and “Every day or almost every day”. Higher scale score values indicate higher frequencies of engagement by the principal and school management team in instructional leadership activities, while lower scale score values indicate lower frequencies of engagement by the principal and school management team in instructional leadership activities. Table 19.140 shows the item wording and item parameters for the items in this scale.

School autonomy (SCHAUTO)

Principals were asked in SC202 about who had the main responsibility for various decisions or activities at their school (e.g., “Appointing or hiring teachers”, “Determining teachers’ salary increases”). The six response categories for the 12 items in the scale were “Principal”, “Teachers or members of <school management team>”, “<School governing board>”, “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. Higher scale score values indicate that the principal, teachers or members of the school management team, and the school governing board had a greater level of autonomy in decision-making activities at their school. Lower scale score values indicate that these groups had less autonomy. Table 19.141 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Teacher participation (TCHPART)

Principals were asked in SC202 about who had the main responsibility for various decisions or activities at their school (e.g., “Formulating the school budget”, “Choosing which learning materials are used”). The six response categories for the 12 items in the scale were “Principal”, “Teachers or members of <school management team>”, “<School governing board>”, “<Local or municipal authority>”, “<Regional or state authority>”, and “<National or federal authority>”. Higher scale score values indicate that the teachers or members of the school management team participated to a greater extent in decision-making activities at their school. Lower scale score values indicate that they participated to a lesser extent. Table 19.142 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Organisation of student learning at school (Module 14)

Digital device policies at school (DIGDVPOL)

School principals were asked in SC190 to indicate whether their school had various policies regarding digital device use (e.g., “Teachers establish rules for when students may use digital devices during lessons.”, “The school has a specific programme to prepare students for responsible internet behaviour.”). The two response categories for the nine items in the scale were “Yes” and “No”. Higher scale score values indicate that digital device policies are enforced at the school to a greater extent, while lower scale score values indicate that such policies are enforced to a lesser extent. Table 19.143 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Teacher qualification, training, and professional development (Module 17)

Mathematics teacher training (MTTRAIN)

School principals were asked in SC184 to indicate the areas in which professional development was offered to mathematics teachers in their school (e.g., “Mathematics content”, “Mathematics curriculum”). The two response categories for the seven items in the scale were “Yes” and “No”. Higher scale score values indicate that more opportunities for professional development are offered to mathematics teachers in the school, while lower scale score values indicate that fewer professional development opportunities are offered to them. Table 19.144 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Assessment, evaluation, and accountability (Module 18)

Feedback to teachers (TEAFDBK)

Principals were asked in SC193 how much impact teacher evaluations had on various matters (e.g., “A change in salary”, “A change in the likelihood of career advancement”). The four response categories for the seven items in the scale were “No impact”, “Small impact”, “Moderate impact”, and “Large impact”. Higher scale score values indicate greater impact of teacher evaluations or feedback, while lower scale score values indicate lesser impact. Table 19.145 shows the item wording and item parameters for the items in this scale.

Use of standardised tests (STDTEST)

Principals were asked in SC035 to indicate whether standardised tests were used for various purposes (e.g., “To guide students’ learning”, “To group students for instructional purposes”). The two response categories for the 11 items in the scale were “Yes” and “No”. Higher scale score values indicate that standardised tests are used for accountability purposes to a greater extent, while lower scale score values indicate that these tests are used to a lesser extent. Table 19.146 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Use of teacher-developed tests (TDTEST)

Principals were asked in SC035 to indicate whether teacher-developed tests were used for various purposes (e.g., “To guide students’ learning”, “To group students for instructional purposes”). The two response categories for the 11 items in the scale were “Yes” and “No”. Higher scale score values indicate that teacher-developed tests are used for accountability purposes to a greater extent, while lower scale score values indicate that these tests are used to a lesser extent. Table 19.147 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Parental/guardian involvement and support (Module 19)

School encouragement of parent or guardian involvement (ENCOURPG)

Principals were asked in SC192 about how often their school staff engaged parents or guardians in various aspects of students’ educational environment during the last academic year (e.g., “Invited parents or guardians to volunteer for school activities”, “Initiated communications with parents or guardians about their child’s progress”). The four response categories for the six items in the scale were “Never or almost never”, “A few times a year”, “A few times a month”, and “Once a week or more”. Higher scale score values indicate more frequent efforts by the school staff to engage parents or guardians in becoming involved at the school, while lower scale score values indicate less frequent engagement efforts. Table 19.148 shows the item wording and item parameters for the items in this scale.

Creative thinking (Module 20)

Beliefs about creativity (BCREATSC)

Principals were asked in SC204 to indicate their level of agreement with statements regarding their beliefs about creativity (e.g., “Creativity can be trained.”, “There are many different ways to be creative.”). The four response categories for the four items in the scale were “Strongly disagree”, “Disagree”, “Agree”, and “Strongly agree”. Higher scale score values indicate that principals endorse, to a greater extent, beliefs about the malleability of creativity and an expansive view of what it means to be creative. Lower scale

score values indicate that principals endorse these beliefs to a lesser extent. Table 19.149 shows the item wording and item parameters for the items in this scale.

Creative school activities offered (ACTCRESC)

Principals were asked in SC207 to indicate how often creative activities are offered in their school (e.g., “Creative writing classes/activities”, “Debate <club>”). The five substantive response categories for the eight items in the scale were “Never or almost never”, “About once or twice a year”, “About once or twice a month”, “About once or twice a week”, and “Every day or almost every day”. There was an additional response category, “Not available at our school”, which was recoded as missing prior to scaling. Higher scale score values indicate a greater frequency of creative activities being offered in school, while lower scale score values indicate creative activities are offered on a less frequent basis. Table 19.150 shows the item wording and item parameters for the items in this scale. It also indicates how the response categories were recoded prior to scaling.

Creative school environment (CREENVSC)

Principals were asked in SC205 to indicate their level of agreement with statements regarding the encouragement of creative thinking by teachers and through activities at the school (e.g., “Teachers in our school value students’ creativity.”, “Class activities in our school help students think about new ways to solve complex tasks.”). The four response categories for the six items in the scale were “Strongly disagree”, “Disagree”, “Agree”, and “Strongly agree”. Higher scale score values indicate more agreement with the overall view that students’ creativity is encouraged in the school, while lower scale score values indicate less agreement with this view. Table 19.151 shows the item wording and item parameters for the items in this scale.

Openness culture/climate (OPENCUL)

Principals were asked in SC208 to indicate the extent to which they agree or disagree with statements regarding their students’ orientation towards openness and creativity (e.g., “Most students at my school are creative.”, “Most students at my school enjoy learning new things.”). The four response categories for the nine items in the scale were “Strongly disagree”, “Disagree”, “Agree”, and “Strongly agree”. Higher scale score values indicate that students have a greater orientation towards openness and creativity, while lower scale score values indicate they have less orientation towards openness and creativity. Table 19.152 shows the item wording and item parameters for the items in this scale.

Global crises (Module 21)

Note that the questions in this module were skipped for respondents who reported that their school had not been closed for one or more school days because of COVID-19 in question SC213.

Problems with schools’ capacity to provide remote instruction (PROBSCRI)

School administrators were asked in SC216 to what extent specific challenges hindered their school’s capacity to provide remote instruction during the time when the school building was closed to students because of COVID-19 (e.g., “Lack of access to <digital devices> among students”, “Lack of learning management systems or school learning platforms (e.g., [Blackboard®], [Edmodo®], [Moodle®], [Google® Classroom™])”). The four response categories for the five items in the scale were “Not at all”, “Very little”, “To some extent”, and “A lot”. Higher scale score values indicate a greater level of problems with the schools’ capacity to provide remote instruction while the school building was closed to students during the COVID-19 pandemic, while lower scale score values indicate fewer problems with the capacity to provide remote instruction. Table 19.153 shows the item wording and item parameters for the items in this scale.

School preparation for remote instruction – Before pandemic (SCPREBP)

School administrators were asked in SC223 whether their school had taken specific actions to prepare for remote instruction (e.g., “Adapting existing curriculum plans for remote instruction (e.g., modifying course requirements, sequence of lessons, grading policies)”, “Ensuring that students have access to <digital devices> for remote instruction”). The three response categories for the 10 items in the scale were “Yes, as a standard practice before the COVID-19 pandemic”, “Yes, in response to the COVID-19 pandemic”, and “No”. Prior to scaling, the first response was coded as 1, while the two latter responses were coded as 0. Higher scale score values indicate a higher level of school preparation for remote instruction before the pandemic, while lower scale score values indicate less preparation for remote instruction before the pandemic. Table 19.154 shows the item wording and item parameters for the items in this scale.

School preparation for remote instruction – In response to pandemic (SCPREPAP)

School administrators were asked in SC223 whether their school had taken specific actions to prepare for remote instruction (e.g., “Adapting existing curriculum plans for remote instruction (e.g., modifying course requirements, sequence of lessons, grading policies)”, “Ensuring that students have access to <digital devices> for remote instruction”). The three response categories for the 10 items in the scale were “Yes, as a standard practice before the COVID-19 pandemic”, “Yes, in response to the COVID-19 pandemic”, and “No”. Prior to scaling, the first two responses were coded as 1, while the third response was coded as 0. Higher scale score values indicate a higher level of school preparation for remote instruction after the start of the pandemic, while lower scale score values indicate less preparation for remote instruction after the start of the pandemic. Table 19.155 shows the item wording and item parameters for the items in this scale.

Preparedness for digital learning (DIGPREP)

School administrators were asked in SC155 to rate their agreement with statements about their school's capacity to use digital devices to enhance learning and teaching (“Teachers have sufficient time to prepare lessons integrating digital devices”, “The school has sufficient qualified technical assistant staff”). The four response categories for the six items in the scale were “Strongly disagree”, “Disagree”, “Agree”, and “Strongly agree”. Higher scale score values indicate greater capacity for a school to use digital technology for learning and teaching, while lower scale score values indicate less capacity for a school to do so. Table 19.156 shows the item wording and item parameters for the items in this scale.

Teacher Questionnaire derived variables

The Teacher Questionnaire is an international option that countries/economies could choose to implement. Routing within the questionnaire was used to deliver specific questions to mathematics teachers and specific questions to all other, non-mathematics teachers. Other questions in the questionnaire were administered to all teachers. There were 45 variables derived from this questionnaire, including 13 simple DVs and 32 IRT scaled DVs. An overview of all DVs in this questionnaire is shown in Table 19.157 and each are described in the following sections.

Simple questionnaire indices

Originally trained teachers – Strict (OTT1) and broad (OTT2) definitions

The Teacher Questionnaire addressed two questions about teachers' initial education and professional development. The first question, TC014, asks if teacher education or training programme was completed, with response options “1, Yes, a programme of 1 year or less” “2, Yes, a programme longer than 1 year”

and “3, No”. TC015 asked about how the teacher qualification was received. Response options included “1, I attended a standard teacher education or training programme at an <educational institute which is eligible to educate or train teachers>.”, “2, I attended an in-service teacher education or training programme.”, “3, I attended a work-based teacher education or training programme.”, “4, I attended training in another pedagogical profession.”, or “5, Other”. These two questions (TC014, TC015) were used to build the DV OTT1 (Originally trained teachers, strict definition) and OTT2 (Originally trained teachers, broad definition). The strict definition implies that a teacher had intended to be trained as a teacher from the very beginning of his or her career and has finished a “standard teacher education or training programme at an <educational institute which is eligible to educate or train teachers>”. In the less strict definition, the teacher has finished any of the following three programmes: either a “standard teacher education or training programme at an <educational institute which is eligible to educate or train teachers>” (option 1 in TC015), an “in-service teacher education or training programme” (option 2) or a “work-based teacher education or training programme” (option 3 in TC015).

Trained to teach certain subjects (NTEACH1-11)

TC018 asked about the specific subjects that were included in the teacher’s education or training programme or other professional qualification and asked if the respondents taught these subjects to the national modal grade for 15-year-olds in the current school year. The DVs NTEACH1 to NTEACH 11 reflect whether the teacher was trained to teach a certain subject. A value of “1” indicate that teachers were trained to teach the subject in question, while “0” indicates they were not trained to teach the subject in question.

Subject-specific overlap between initial education and teaching the modal grade (STTMG1-11)

TC018 enquired about the specific subjects that were included in the teacher’s education or training programme or other professional qualification and asked if the respondents taught these subjects to the national modal grade for 15-year-olds in the current school year. This question is used to build the DVs STTMG1 to STTMG11, indicating the subject-specific overlap between initial education and teaching the modal grade, i.e., whether a teacher currently teaches a certain subject combined with whether it was included in the teacher’s initial training. A value of “0” indicates that teachers were neither trained nor teach the subject in question, “1” indicates they were trained to teach the subject in question but do not teach it, “2” indicates they were not trained to teach the subject in question but they do teach the subject, and “3” indicates they were trained to teach the subject in question and they teach the subject.

Country born (COBN_T)

COBN_T is based on question TC186, which asks about the country/economy a teacher is born in, coded into the following categories: (1) “Country of test” and (2) “Other country”. Each country/economy adapts the items for this question to collect relevant country/economy of birth information for teachers in their country/economy, so the index also gives detailed categories of teachers’ original countries/economies of birth within a country/economy that vary from country/economy to country/economy.

Content overlap between initial education and professional development (TC045Q01-TC045Q18)

TC045 asked about 10 content topics (e.g., “knowledge and understanding of my subject field(s)”, “knowledge of the curriculum”) that might have been included in the teachers’ initial education and training and/or in professional development activities during the last 12 months. Teachers could select both if applicable. The DVs TC045Q01 to TC045Q18 reflect the content overlap between initial education and professional development.

Higher educational level attained (TCISCED)

Teachers' responses to TC210 regarding education were classified using ISCED 2011. An index on higher educational level attained was constructed by recoding educational qualifications into the following categories: Less than ISCED Level 3.3 (upper secondary with no direct access to tertiary education), ISCED Level 3.3 (upper secondary with no direct access to tertiary education), ISCED 3.4 (upper secondary with direct access to tertiary education), ISCED 4 (post-secondary non-tertiary), ISCED 5 (short-cycle tertiary), ISCED 6 (Bachelor's or equivalent level), ISCED 7 (Master's or equivalent level), and ISCED 8 (Doctoral or equivalent level). Scores are assigned as noted in Table 19.158.

Employment status (EMPLSTAT and EMPLSTATd)

TC211 asked about employment status in terms of the contract duration with three response options ("Permanent employment", "Fixed-term contract for a period of more than 1 school year", "Fixed-term contract for a period of 1 school year or less"). The corresponding DVs reflected the duration of employment, measured via TC211, a) on the original three-point scale (EMPLSTAT) and b) dichotomous, distinguishing a permanent position from fixed-term contracts (EMPLSTATd).

Study abroad (STABROAD)

Teachers' responses to the whether they had studied abroad in question TC188 were scaled into the index of "Study abroad". There are four response options (No; Yes, for less than three months; Yes, for three to twelve months; Yes, for more than a year). This question was previously included in the PISA 2018 Teacher Questionnaire, but with no index. Values on this index are 0 (Never studied abroad) and 1 (Studied abroad). This variable was skipped for teachers who reported that they taught mathematics to students in national modal grade for 15-year-olds this school year on question TC217.

Weekly teacher workload (TCWKLOAD)

Teachers were asked to enter how many hours a week they spend on various teaching tasks (e.g., "Marking/correcting of student work", "General administrative work"). Teachers' responses to each of the eight fill-in items in question TC216 were summed to create the simple index "Weekly teacher workload".

Use of digital resources for mathematics (ICTMATTC)

Teachers' frequency ratings of how often they instruct their students to use digital resources for a range of mathematics tasks in class or for homework (e.g., "Use digital resources for simple calculations", "Use digital resources for simulations and modelling, virtual laboratories") in question TC222 were aggregated into a simple index "Use of digital resources for mathematics" as follows: A value of "1" is assigned if teachers select response options 4 ("About once or twice a week") or 5 ("Every day or almost every day") at least once across the 4 items in this question, "0" otherwise. The index captures whether teachers said that they use some digital resources for mathematics lessons at least on a weekly basis or not. This variable was skipped for teachers who reported that they did not teach mathematics to students in national modal grade for 15-year-olds this school year on question TC217.

Proportion of working years at this school (PROPWORK)

In TC007 teachers were asked to how many years of teaching experience they had at the school having them take the questionnaire and how many years of teaching experience they had in total. These responses were used to form the simple index "Proportion of working years at this school", by dividing the number of years at their current school by total number of years teaching.

Derived variables based on IRT Scaling

The Teacher Questionnaire provided data for 32 DVs based on IRT scaling. The Cronbach's alpha for each scale and group are presented in Table 19.159, the number of items with international parameters for each scale and group are presented in Table 19.160, the number of trend items with international parameters for each trend scale and group are presented in Table 19.161, the countries/economies for which the scale scores were suppressed for each scale are presented in Table 19.162, and the groups that did not administer each scale are presented in Table 19.163.

Proportion of professional development (PRPDT)

Teachers' responses about their participation in different professional development activities in the last 12 months (e.g., "Individual or collaborative research on a topic of interest to you professionally", "Course, workshop, or conference on teaching methods") in question TC020 were scaled into the index of "Proportion of professional development". Each of the 14 items included in this scale had two response options ("Yes", "No"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.164 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Exchange and co-ordination for teaching (EXCHT)

Teachers' frequency ratings of how often they participate in teaching-related co-operation (e.g., Exchange teaching materials with colleagues", "Attend team conferences") in question TC046 were scaled into the index of "Exchange and co-ordination for teaching". Note that this scale was linked to the EXCHT scale in PISA 2018. Each of the four items included in this scale had six response options ("Never", "Once a year or less", "2-4 times a year", "5-10 times a year", "1-3 times a month", "Once a week or more"). Table 19.165 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Teaching ICT awareness (ICTOTL)

Teachers' responses about whether they taught various ICT awareness activities to their students (e.g., "How to decide whether to trust information from the Internet", "How to detect phishing or spam emails") in question TC166 were scaled into the index of "Teaching ICT awareness". Each of the seven items included in this scale had two response options ("Yes", "No"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.166 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Teachers' use of specific ICT applications (TCICTUSE)

Teachers' frequency ratings of how often they used specific ICT applications while teaching (e.g., "Digital learning games", "Data logging and monitoring tools") in question TC169 were scaled into the index of "Teachers' use of specific ICT applications". Note that this scale was linked to the TCICTUSE scale in PISA 2018. Each of the 14 items included in this scale had four response options ("Never", "In some lessons", "In most lessons", "In every or almost every lesson"). Table 19.167 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Disciplinary climate in mathematics (TCDISCLIMA)

Teachers' frequency ratings of how often a range of situations occurred in their mathematics lessons (e.g., "There is noise and disorder.", "I have to wait a long time for students to quiet down.") in question TC170 were scaled into the index of "Disciplinary climate in mathematics", which measures how much discipline there is during mathematics lessons. Each of the seven items included in this scale had four response options ("Every lesson", "Most lessons", "Some lessons", "Never or almost never"). Note that this scale was skipped for teachers who reported that they did not teach mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.168 shows the item wording and item parameters for the items in this scale.

Need for professional development (DEVNEED)

Teachers' ratings of their need for professional development in various areas (e.g., "Knowledge of the curriculum", "Student behaviour and classroom management") in question TC185 were scaled into the index of "Need for professional development". Each of the 13 items included in this scale had four response options ("No need at present", "Low level of need", "Moderate level of need", "High level of need"). Table 19.169 shows the item wording and item parameters for the items in this scale.

Teachers' attitudes toward equal rights for immigrants (TCATTIMM)

Teachers' ratings of their agreement with statements about immigrants (e.g., "Immigrant children should have the same opportunities for education that other children in the country have.", "Immigrants should have the opportunity to continue their own customs and lifestyle.") in question TC196 were scaled into the index of "Teachers' attitudes toward equal rights for immigrants". This scale was linked to the TCATTIMM scale in PISA 2018. Each of the four items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.170 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Satisfaction with current job environment (SATJOB)

Teachers' ratings of their agreement with various statements indicating their satisfaction with the current job environment (e.g., "I enjoy working at this school.", "I would recommend my school as a good place to work.") in question TC198 were scaled into the index of "Satisfaction with current job environment". Note that this scale was linked to the SATJOB scale in PISA 2018. Each of the four items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.171 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Satisfaction with teaching profession (SATTEACH)

Teachers' ratings of their agreement with various statements indicating their satisfaction with the teaching profession (e.g., "The advantages of being a teacher clearly outweigh the disadvantages.", "I regret that I decided to become a teacher.") in question TC198 were scaled into the index of "Satisfaction with teaching profession". Note that this scale was linked to the SATTEACH scale in PISA 2018. Each of the five items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.172 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling and which are trend items.

Teacher's self-efficacy in classroom management (SEFFCM)

Teachers' ratings of their classroom management skills (e.g., "Control disruptive behaviour in the classroom", "Get students to follow classroom rules") in question TC199 were scaled into the index of "Teacher's self-efficacy in classroom management". This scale was linked to the SEFFCM scale in PISA 2018. Each of the four items included in this scale had four response options ("Not at all", "To some extent", "Quite a bit", "A lot"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.173 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Teacher's self-efficacy in maintaining positive relations with students (SEFFREL)

Teachers' ratings of their ability to maintain positive relations with students (e.g., "Help my students value learning", "Motivate students who show low interest in school work") in question TC199 were scaled into the index of "Teacher's self-efficacy in maintaining positive relations with students". This scale was linked to the SEFFREL scale in PISA 2018. Each of the four items included in this scale had four response options ("Not at all", "To some extent", "Quite a bit", "A lot"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.174 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Teacher's self-efficacy in instructional settings (SEFFINS)

Teachers' ratings of their self-efficacy in instructional settings (e.g., "Craft good questions for my students", "Provide an alternative explanation for example when students are confused") in question TC199 were scaled into the index of "Teacher's self-efficacy in instructional settings". This scale was linked to the SEFFINS scale in PISA 2018. Each of the four items included in this scale had four response options ("Not at all", "To some extent", "Quite a bit", "A lot"). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.175 shows the item wording and item parameters for the items in this scale. It also indicates which items are trend items.

Teacher use of ICT (TCDIGRES)

Teachers' frequency ratings of how often they use ICT for various teaching tasks (e.g., "Use <digital resources> to design tasks", "Use <digital resources> to provide feedback to students") in question TC220 were scaled into the index of "Teacher use of ICT". Each of the nine items included in this scale had five response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day"). Table 19.176 shows the item wording and item parameters for the items in this scale.

Emphasis on ICT competencies (ICTCOMP)

Teachers' ratings of how much emphasis they place on teaching various ICT competencies (e.g., "Evaluating the credibility of digital information", "Using digital tools to work collaboratively") in question TC221 were scaled into the index of "Emphasis on ICT competencies". Each of the five items included in this scale had four response options ("No emphasis", "Little emphasis", "Some emphasis", "A lot of emphasis"). Table 19.177 shows the item wording and item parameters for the items in this scale.

Teaching of mathematical reasoning and 21st century mathematics topics (EXPO21TC)

Teachers' frequency ratings of how often they had taught a range of different mathematics topics during the school year (e.g., "Extracting mathematical information from diagrams, graphs, or simulations", "Using the concept of statistical variation to make a decision") in question TC223 were scaled into the index of "Teaching of mathematical reasoning and 21st century mathematics topics". Each of the 10 items included in this scale had four response options ("Frequently", "Sometimes", "Rarely", "Never"). Note that this scale was skipped for teachers who reported that they did not teach mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.178 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Encouraging mathematical thinking (COGACMTC)

Teachers' frequency ratings of how often they showed a range of behaviours indicative of encouraging mathematical thinking during the school year (e.g., "I encouraged students to "think mathematically".", "I asked students how different topics are connected to a bigger mathematical idea.") in question TC227 were scaled into the index of "Encouraging mathematical thinking". Each of the nine items included in this scale had five response options ("Never or almost never", "Less than half of the lessons", "About half of the lessons", "More than half of the lessons", "Every lesson or almost every lesson"). Note that this scale was skipped for teachers who reported that they did not teach mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.179 shows the item wording and item parameters for the items in this scale.

Fostering reasoning (COGACRTC)

Teachers' frequency ratings of how often they showed a range of behaviours indicative of fostering mathematics reasoning during the school year (e.g., "I asked students to explain their reasoning when solving a mathematics problem.", "I asked students to defend their answer to a mathematics problem.") in question TC228 were scaled into the index of "Fostering reasoning". Each of the nine items included in this scale had five response options. ("Never or almost never", "Less than half of the lessons", "About half of the lessons", "More than half of the lessons", "Every lesson or almost every lesson"). Note that this scale was skipped for teachers who reported that they did not teach mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.180 shows the item wording and item parameters for the items in this scale.

Goals and views about teaching mathematics (TCMGOALS)

Teachers' agreement with statements about their views and goals when teaching mathematics (e.g., "Explaining why an answer is correct is just as important as getting a correct answer.", "Asking students to solve difficult problems in class helps them become good problem solvers.") in question TC230 were scaled into the index of "Goals and views about teaching mathematics". Each of the 11 items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Note that this scale was skipped for teachers who reported that they did not teach mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.181 shows the item wording and item parameters for the items in this scale.

Adaptation of instruction (ADAPTINSTR)

Teachers' frequency ratings of how often they adapt instruction for students (e.g., "I tailor my teaching to meet the needs of my students.", "I provide individual support for advanced students.") in question TC232 were scaled into the index of "Adaptation of instruction". Each of the four items included in this scale had

four response options (“Never or almost never”, “Some lessons”, “Many lessons”, “Every lesson or almost every lesson”). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.182 shows the item wording and item parameters for the items in this scale.

Feedback provided by the teachers (FEEDBINSTR)

Teachers’ frequency ratings of how often they provide feedback to students (e.g., “I give students feedback on their strengths in my course.”, “I tell students how they can improve their performance.”) in question TC232 were scaled into the index of “Feedback provided by the teachers”. Each of the five items included in this scale had four response options (“Never or almost never”, “Some lessons”, “Many lessons”, “Every lesson or almost every lesson”). Note that this scale was skipped for teachers who reported that they taught mathematics to students in the national modal grade for 15-year-olds in the current school year on question TC217. Table 19.183 shows the item wording and item parameters for the items in this scale.

Openness to creativity (OPENCTTC)

Teachers’ ratings of their agreement with statements about their openness to creative activities (e.g., “I enjoy projects that require creative solutions.”, “I express myself through art.”) in question TC234 were scaled into the index of “Openness to creativity”. Each of the eight items included in this scale had four response options (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”). Table 19.184 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Creative values (CREATVAL)

Teachers’ ratings of their agreement with statements about their values regarding creativity (e.g., “It is important that students are able to make creative works like drawing and painting.”, “It is important for students to solve science problems creatively.”) in question TC235 were scaled into the index of “Creative values”. Each of the six items included in this scale had four response options (“Strongly disagree”, “Disagree”, “Agree”, “Strongly agree”). Table 19.185 shows the item wording and item parameters for the items in this scale.

Teachers’ use of creative pedagogies (CREATPED)

Teachers’ ratings of how much importance they place on using creative pedagogies in class (e.g., “Finding ideas through brainstorming”, “Debating ideas or current issues”) in question TC236 were scaled into the index of “Teachers’ use of creative pedagogies”. Each of the seven items included in this scale had four response options (“No importance”, “Very little importance”, “Some importance”, “A lot of importance”). Table 19.186 shows the item wording and item parameters for the items in this scale.

Teachers’ capacity to concentrate at work (CAPCON)

Teachers’ frequency ratings of how often they experienced various situations during the school day (e.g., “I was distracted.”, “I felt focused.”) in question TC237 were scaled into the index of “Teachers’ capacity to concentrate at work”. Each of the six items included in this scale had four response options (“Never”, “Seldom”, “Often”, “Always”). Table 19.187 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Teachers' affect (AFFECT)

Teachers' frequency ratings of how often they felt various emotions during the school day (e.g., "I felt cheerful and in good spirits.", "I felt active and vigorous.") in question TC238 were scaled into the index of "Teachers' affect". Each of the five items included in this scale had four response options ("Never", "Seldom", "Often", "Always"). Table 19.188 shows the item wording and item parameters for the items in this scale.

Teachers' feeling of trust (TRUST)

Teachers' ratings of their agreement with statements about a climate of trust within the school (e.g., "Teachers can rely on the school's management for professional support.", "I feel that I can trust my colleagues.") in question TC241 were scaled into the index of "Teachers' feeling of trust". Each of the five items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.189 shows the item wording and item parameters for the items in this scale.

Teachers' work overload (OVERLOAD)

Teachers' ratings of their agreement with statements concerning work overload (e.g., "I am given enough time to do what is expected of me at work.", "I have too much work for one person to do.") in question TC243 were scaled into the index of "Teachers' work overload". Each of the six items included in this scale had four response options ("Strongly disagree", "Disagree", "Agree", "Strongly agree"). Table 19.190 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Teachers' work autonomy (AUTONOMY)

Teachers' ratings of how much control they have over various decisions at their school (e.g., "Determining course content", "Disciplining students") in question TC246 were scaled into the index of "Teachers' work autonomy". Each of the seven items included in this scale had four response options ("No control", "Some control", "A lot of control", "Full control"). Table 19.191 shows the item wording and item parameters for the items in this scale.

School leadership (LEADSHIP)

Teachers' frequency ratings of how often their school's principal took various actions (e.g., "My principal collaborated with teachers to solve classroom discipline problems.", "My principal observed instruction in the classroom.") in question TC253 were scaled into the index of "School leadership". Each of the seven items included in this scale had four response options ("Never or rarely", "Sometimes", "Often", "Very often"). Table 19.192 shows the item wording and item parameters for the items in this scale.

Occupational stress (OCSTRESS)

Teachers' ratings of their agreement with various statements regarding their stress at work (e.g., "I experience stress in my work.", "My job negatively impacts my mental health.") in question TC254 were scaled into the index of "Occupational stress". Each of the four items included in this scale had four response options ("Not at all", "To some extent", "Quite a bit", "A lot"). Table 19.193 shows the item wording and item parameters for the items in this scale. It also indicates which items were reverse-coded prior to scaling.

Sources of stress (STRESS)

Teachers' ratings of their agreement with various situations causing stress at work (e.g., "Having too many lessons to teach", "Being held responsible for students' achievement") in question TC255 were scaled into the index of "Sources of stress". Each of the nine items included in this scale had four response options ("Not at all", "To some extent", "Quite a bit", "A lot"). Table 19.194 shows the item wording and item parameters for the items in this scale.

Negative physical symptoms (NEGSYMPT)

Teachers' frequency ratings of how often they had various psychosomatic symptoms during the school day (e.g., "Headache", "Fatigue") in question TC256 were scaled into the index of "Negative physical symptoms". Each of the 10 items included in this scale had five response options ("Never or almost never", "About once or twice a year", "About once or twice a month", "About once or twice a week", "Every day or almost every day"). Table 19.195 shows the item wording and item parameters for the items in this scale.

References

- Avvisati, F. (2020), "The measure of socio-economic status in PISA: A review and some suggested improvements", *Large-Scale Assessments in Education*, Vol. 8/1, pp. 1-37, <https://doi.org/10.1186/s40536-020-00086-x>. [14]
- Bertling, J. and J. Weeks (2020), *Getting More Bang for Your Buck: Within-construct Questionnaire Matrix Sampling*, Paper presented to PISA Technical Advisory Group, September 2020, Princeton, NJ. [3]
- Bertling, J. and J. Weeks (2018), *Plans for Within-construct Questionnaire Matrix Sampling in PISA 2021*, Paper presented to PISA Technical Advisory Group, August 2018, Princeton, NJ. [2]
- Bertling, J. et al. (2020), "Comparison of within-construct matrix sampling with scale shortening: Impact on IRT scaling and population modelling", *Memo prepared for PISA Technical Advisory Group, TAG (2020)2*. [4]
- Birnbaum, A. (1968), "Some latent trait models and their use in inferring an examinee's ability", in Lord, F. and M. Novick (eds.), *Statistical Theories of Mental Test Scores*, Addison-Wesley. [5]
- Cowan, C. et al. (2012), *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*, National Center for Education Statistics, https://nces.ed.gov/nationsreportcard/pdf/researchcenter/socioeconomic_factors.pdf. [13]
- Ganzeboom, H. and D. Treiman (2003), "Three internationally standardised measures for comparative research on occupational status", in Hoffmeyer-Zlotnik, J. and C. Wolf (eds.), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Kluwer Academic Press. [12]
- Masters, G. (1982), "A Rasch model for partial credit scoring", *Psychometrika*, Vol. 47/2, pp. 149-174, <https://doi.org/10.1007/BF02296272>. [10]
- Muraki, E. (1992), "A generalized partial credit model: Application of an EM algorithm", *Applied* [6]

Psychological Measurement, Vol. 16/2, pp. 159-177, <https://doi.org/10.1002/j.2333-8504.1992.tb01436.x>.

- OECD (2023), *PISA 2022 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/dfe0bf9c-en>. [1]
- OECD (2020), *PISA 2018 Technical Report*, PISA, OECD Publishing, Paris, <https://www.oecd.org/pisa/data/pisa2018technicalreport/>. [15]
- OECD/Eurostat/UNESCO Institute for Statistics (2015), *ISCED 2011 Operational Manual: Guidelines for Classifying National Education Programmes and Related Qualifications*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264228368-en>. [16]
- Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Nielsen and Lydiche. [9]
- Shin, H. et al. (2017), *Multidimensional Discrete Latent Trait Models (mdltn) [Draft manual]*, Educational Testing Service. [7]
- von Davier, M. (2015), *Multidimensional discrete latent trait models (mdltn) (Version 1.965) [Computer software]*. [8]
- Warm, T. (1989), "Weighted likelihood estimation of ability in item response theory", *Psychometrika*, Vol. 54/3, pp. 427-450, <https://doi.org/10.1007/BF02294627>. [11]

Notes

1. For the seven trend scales linked to PISA 2012, the Rasch model (Rasch, 1960) was used to scale the dichotomous items, while the partial credit model (PCM) was used to scale the polytomous items, in line with the models used in PISA 2012.
2. The International Standard Classification of Education (ISCED) is used in international educational statistics to classify levels in education systems worldwide. A link to the 2011 framework, ISCED 2011, used in PISA 2022 can be found at <https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>
3. Separate simple DVs (ST250D06JA, ST250D07JA, ST251D08JA, ST251D09JA) combine all the customizations across countries/economies to ST250C06JA, ST250C07JA, ST251C08JA, and ST251C09JA, respectively. See Annex D.
4. A separate simple DV (ST330D10WA) combines all the customizations across countries/economies to ST330C10WA. See Annex D.

5. The random value was drawn from a normal distribution with a mean of 0 and a standard deviation equal to the standard deviation of the residuals of the regression model for the country/economy.
6. Due to missing data within the OECD countries, the senate weights of the individuals that were ultimately included in the calculation of the mean and standard deviation of each component did not sum to 5,000 for each OECD country. Therefore, the OECD countries were only approximately equally weighted in the calculation of the mean and standard deviation of each component.
7. Again, due to missing data within the OECD countries, the senate weights of the individuals that were ultimately included in the calculation of the mean and standard deviation of the preliminary ESCS scores did not sum to 5,000 for each OECD country. Therefore, the OECD countries were only approximately equally weighted in the calculation of the mean and standard deviation of the preliminary ESCS scores.
8. Due to missing data within the OECD countries, the senate weights of the individuals that were ultimately included in the calculation of the mean and standard deviation of each component in PISA 2018 did not sum to 5,000 for each OECD country. Therefore, the OECD countries were only approximately equally weighted in the calculation of the mean and standard deviation of each component in PISA 2018.
9. The random value was drawn from a normal distribution with a mean of 0 and a standard deviation equal to the standard deviation of the residuals of the regression model for the country/economy and cycle.

Chapter 19 tables

All tables are available online at <https://www.oecd.org/pisa/data/pisa2022technicalreport/>

Tables	Title
Table 19.1	Student Questionnaire scales using within-construct matrix sampling design
Table 19.2	List of all trend scales
Table 19.3	OECD mean and standard deviation of the original WLEs for the new scales
Table 19.4	OECD mean and standard deviation of the original WLEs in PISA 2018 for trend scales linked to PISA 2018
Table 19.5	OECD mean and standard deviation of the original WLEs in PISA 2012 for trend scales linked to PISA 2012
Table 19.6	Variables derived from the Student Questionnaire
Table 19.7	Mapping of ISCED levels to years of education
Table 19.8	ISCEDP values and labels
Table 19.9	Creation of EXPECEDU index
Table 19.10	Cronbach's alpha for the IRT scales in the Student Questionnaire
Table 19.11	Number of items with international parameters for the IRT scales in the Student Questionnaire
Table 19.12	Number of trend items with international parameters for the trend IRT scales in the Student Questionnaire
Table 19.13	Countries/economies for which the scale scores were suppressed for the IRT scales in the Student Questionnaire
Table 19.14	Groups that did not administer the IRT scales in the Student Questionnaire
Table 19.15	Items in the HOMEPOS scale
Table 19.16	Recoding of items in the HOMEPOS scale
Table 19.17	Items in the ICTRES scale
Table 19.18	Items in the INFOSEEK scale
Table 19.19	Items in the BULLIED scale
Table 19.20	Items in the FEELSAFE scale
Table 19.21	Items in the TEACHSUP scale
Table 19.22	Items in the RELATST scale
Table 19.23	Items in the SCHRISK scale
Table 19.24	Items in the BELONG scale
Table 19.25	Items in the GROSAGR scale
Table 19.26	Items in the ANXMAT scale
Table 19.27	Items in the MATHEFF scale
Table 19.28	Items in the MATHEF21 scale
Table 19.29	Items in the MATHPERS scale
Table 19.30	Items in the FAMCON scale
Table 19.31	Items in the ASSERAGR scale
Table 19.32	Percent of students that did not receive a scale score for ASSERAGR due to extreme straightlining or for not having enough responses
Table 19.33	Items in the COOPAGR scale
Table 19.34	Percent of students that did not receive a scale score for COOPAGR due to extreme straightlining or for not having enough responses
Table 19.35	Items in the CURIAGR scale
Table 19.36	Percent of students that did not receive a scale score for CURIAGR due to extreme straightlining or for not having enough responses
Table 19.37	Items in the EMOCOAGR scale
Table 19.38	Percent of students that did not receive a scale score for EMOCOAGR due to extreme straightlining or for not having enough responses
Table 19.39	Items in the EMPATAGR scale
Table 19.40	Percent of students that did not receive a scale score for EMPATAGR due to extreme straightlining or for not having enough responses
Table 19.41	Items in the PERSEVAGR scale
Table 19.42	Percent of students that did not receive a scale score for PERSEVAGR due to extreme straightlining or for not having enough responses
Table 19.43	Items in the STRESAGR scale

Tables	Title
Table 19.44	Percent of students that did not receive a scale score for STRESAGR due to extreme straightlining or for not having enough responses
Table 19.45	Items in the EXPOFA scale
Table 19.46	Items in the EXPO21ST scale
Table 19.47	Items in the COGACRSCO scale
Table 19.48	Items in the COGACMCO scale
Table 19.49	Items in the DISCLIM scale
Table 19.50	Items in the FAMSUP scale
Table 19.51	Items in the CREATFAM scale
Table 19.52	Items in the CREATSCH scale
Table 19.53	Items in the CREATEFF scale
Table 19.54	Items in the CREATOP scale
Table 19.55	Items in the IMAGINE scale
Table 19.56	Items in the OPENART scale
Table 19.57	Items in the CREATAS scale
Table 19.58	Items in the CREATOOS scale
Table 19.59	Items in the FAMSUPSL scale
Table 19.60	Items in the FEELLAH scale
Table 19.61	Items in the PROBSELF scale
Table 19.62	Items in the SDLEFF scale
Table 19.63	Items in the SCHSUST scale
Table 19.64	Items in the LEARRES scale
Table 19.65	OECD mean and standard deviation of each component of ESCS and the preliminary ESCS scores for PISA 2022
Table 19.66	Variables derived from the Financial Literacy Questionnaire
Table 19.67	Cronbach's alpha for the IRT scales in the Financial Literacy Questionnaire
Table 19.68	Number of items with international parameters for the IRT scales in the Financial Literacy Questionnaire
Table 19.69	Number of trend items with international parameters for the trend IRT scales in the Financial Literacy Questionnaire
Table 19.70	Groups that did not administer the IRT scales in the Financial Literacy Questionnaire
Table 19.71	Items in the FLSCHOOL scale
Table 19.72	Items in the FLMULTSB scale
Table 19.73	Items in the FLFAMILY scale
Table 19.74	Items in the ACCESSFP scale
Table 19.75	Items in the FLCONFIN scale
Table 19.76	Items in the FLCONICT scale
Table 19.77	Items in the ACCESSFA scale
Table 19.78	Items in the ATTCONFM scale
Table 19.79	Items in the FRINFLFM scale
Table 19.80	Variables derived from the ICT Familiarity Questionnaire
Table 19.81	Cronbach's alpha for the IRT scales in the ICT Familiarity Questionnaire
Table 19.82	Number of items with international parameters for the IRT scales in the ICT Familiarity Questionnaire
Table 19.83	Countries/economies for which the scale scores were suppressed for the IRT scales in the ICT Familiarity Questionnaire
Table 19.84	Groups that did not administer the IRT scales in the ICT Familiarity Questionnaire
Table 19.85	Items in the ICTSCH scale
Table 19.86	Items in the ICTHOME scale
Table 19.87	Items in the ICTQUAL scale
Table 19.88	Items in the ICTSUBJ scale
Table 19.89	Items in the ICTENQ scale
Table 19.90	Items in the ICTFEED scale
Table 19.91	Items in the ICTOUT scale
Table 19.92	Items in the ICTWKDY scale
Table 19.93	Items in the ICTWKEND scale
Table 19.94	Items in the ICTREG scale
Table 19.95	Items in the ICTINFO scale
Table 19.96	Items in the ICTEFFIC scale

Tables	Title
Table 19.97	Variables derived from the Well-Being Questionnaire
Table 19.98	Cronbach's alpha for the IRT scales in the Well-Being Questionnaire
Table 19.99	Number of items with international parameters for the IRT scales in the Well-Being Questionnaire
Table 19.100	Number of trend items with international parameters for the trend IRT scales in the Well-Being Questionnaire
Table 19.101	Groups that did not administer the IRT scales in the Well-Being Questionnaire
Table 19.102	Items in the BODYIMA scale
Table 19.103	Items in the SOCONPA scale
Table 19.104	Items in the LIFESAT scale
Table 19.105	Items in the PSYCHSYM scale
Table 19.106	Items in the SOCCON scale
Table 19.107	Items in the EXPWB scale
Table 19.108	Variables derived from the Parent Questionnaire
Table 19.109	Creation of the PAREXPT index
Table 19.110	Cronbach's alpha for the IRT scales in the Parent Questionnaire
Table 19.111	Number of items with international parameters for the IRT scales in the Parent Questionnaire
Table 19.112	Number of trend items with international parameters for the trend IRT scales in the Parent Questionnaire
Table 19.113	Countries/economies for which the scale scores were suppressed for the IRT scales in the Parent Questionnaire
Table 19.114	Groups that did not administer the IRT scales in the Parent Questionnaire
Table 19.115	Items in the CURSUPP scale
Table 19.116	Items in the PQMIMP scale
Table 19.117	Items in the PQMCAR scale
Table 19.118	Items in the PARINVOL scale
Table 19.119	Items in the PQSCHOOL scale
Table 19.120	Items in the PASCHPOL scale
Table 19.121	Items in the ATTIMMP scale
Table 19.122	Items in the CREATHME scale
Table 19.123	Items in the CREATACT scale
Table 19.124	Items in the CREATOPN scale
Table 19.125	Items in the CREATOR scale
Table 19.126	Variables derived from the School Questionnaire
Table 19.127	Cronbach's alpha for the IRT scales in the School Questionnaire
Table 19.128	Number of items with international parameters for the IRT scales in the School Questionnaire
Table 19.129	Number of trend items with international parameters for the trend IRT scales in the School Questionnaire
Table 19.130	Countries/economies for which the scale scores were suppressed for the IRT scales in the School Questionnaire
Table 19.131	Groups that did not administer the IRT scales in the School Questionnaire
Table 19.132	Items in the NEGSCCLIM scale
Table 19.133	Items in the DMCVIEWS scale
Table 19.134	Items in the STUBEHA scale
Table 19.135	Items in the TEACHBEHA scale
Table 19.136	Items in the ALLACTIV scale
Table 19.137	Items in the EDUSHORT scale
Table 19.138	Items in the STAFFSHORT scale
Table 19.139	Items in the EDULEAD scale
Table 19.140	Items in the INSTLEAD scale
Table 19.141	Items in the SCHAUTO scale
Table 19.142	Items in the TCHPART scale
Table 19.143	Items in the DIGDVPOL scale
Table 19.144	Items in the MTTRAIN scale
Table 19.145	Items in the TEAFDBK scale
Table 19.146	Items in the STDTEST scale
Table 19.147	Items in the TDTEST scale
Table 19.148	Items in the ENCOURPG scale
Table 19.149	Items in the BCREATSC scale
Table 19.150	Items in the ACTCRESC scale

Tables	Title
Table 19.151	Items in the CREENVSC scale
Table 19.152	Items in the OPENCUL scale
Table 19.153	Items in the PROBSCRI scale
Table 19.154	Items in the SCPREPBP scale
Table 19.155	Items in the SCPREPAP scale
Table 19.156	Items in the DIGPREP scale
Table 19.157	Variables derived from the Teacher Questionnaire
Table 19.158	Creation of the TCISCED index
Table 19.159	Cronbach's alpha for the IRT scales in the Teacher Questionnaire
Table 19.160	Number of items with international parameters for the IRT scales in the Teacher Questionnaire
Table 19.161	Number of trend items with international parameters for the trend IRT scales in the Teacher Questionnaire
Table 19.162	Countries/economies for which the scale scores were suppressed for the IRT scales in the Teacher Questionnaire
Table 19.163	Groups that did not administer the IRT scales in the Teacher Questionnaire
Table 19.164	Items in the PRPDT scale
Table 19.165	Items in the EXCHT scale
Table 19.166	Items in the ICTOTL scale
Table 19.167	Items in the TCICTUSE scale
Table 19.168	Items in the TCDISCLIMA scale
Table 19.169	Items in the DEVNEED scale
Table 19.170	Items in the TCATTIMM scale
Table 19.171	Items in the SATJOB scale
Table 19.172	Items in the SATTEACH scale
Table 19.173	Items in the SEFFCM scale
Table 19.174	Items in the SEFFREL scale
Table 19.175	Items in the SEFFINS scale
Table 19.176	Items in the TCDIGRES scale
Table 19.177	Items in the ICTCOMP scale
Table 19.178	Items in the EXPO21TC scale
Table 19.179	Items in the COGACMTC scale
Table 19.180	Items in the COGACRTC scale
Table 19.181	Items in the TCMGOALS scale
Table 19.182	Items in the ADAPTINSTR scale
Table 19.183	Items in the FEEDBINSTR scale
Table 19.184	Items in the OPENCTTC scale
Table 19.185	Items in the CREATVAL scale
Table 19.186	Items in the CREATPED scale
Table 19.187	Items in the CAPCON scale
Table 19.188	Items in the AFFECT scale
Table 19.189	Items in the TRUST scale
Table 19.190	Items in the OVERLOAD scale
Table 19.191	Items in the AUTONOMY scale
Table 19.192	Items in the LEADSHIP scale
Table 19.193	Items in the OCSTRESS scale
Table 19.194	Items in the STRESS scale
Table 19.195	Items in the NEGSYMPT scale

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Member countries of the OECD.

Note by the Republic of Türkiye

The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Türkiye recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Türkiye shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union

The Republic of Cyprus is recognised by all members of the United Nations with the exception of Türkiye. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at: <https://www.oecd.org/termsandconditions>