5 PISA 2022 Context Questionnaire Framework: Balancing Trends and Innovation

This chapter presents the framework for the background core questionnaires for students and schools and explains the goals and rationale for selecting specific questionnaire content for the eighth cycle of PISA. Like prior frameworks, it touches upon how measured constructs theoretically relate to one another and to student achievement. Additionally, the framework outlines a set of survey design principles and methodologies that are introduced to PISA 2022 with the aim of improving measurement, efficiency, and consistency of PISA in the mid to long term.

Glossary

To ensure consistent understanding of specific terms and acronyms used throughout this framework, Table 5.1 below lists key terms used throughout the framework along with brief definitions. Table 5.2 lists and clarifies acronyms used throughout the framework.

Table 5.1. Glossary of key terms

Term	Definition
Construct	A theoretically defined conceptualization (i.e., something constructed) of an aspect of human behaviour or an empirical phenomenon; a construct has empirical indicators but may not be completely observable due to deficits of existing measures. Two broad content categories of constructs are distinguished in the framework: (1) those that are specific to a PISA 2022 content domain (i.e., mathematics, reading, science, creative thinking) and (2) those that are general (i.e., not specific to a PISA 2022 content domain). Each construct is operationally measured by multiple items.
Item	The unit(s) of a question that a respondent answers. In case of a stand-alone discrete question, the item is the same as the question. In case of a matrix question, one question includes several items.
Matrix Question	A question that consists of a question stem and several items with the same response options.
Module	A grouping of two or more related constructs that mark a key topic or theme measured with the PISA 2022 questionnaires.
Question	The parts of a questionnaire designed to elicit information from a respondent. In PISA, the question can take the form of a stand-alone discrete question or a matrix question. When presented in the PISA digital platform, each question appears on a single screen.
Question Stem	The component of a question that presents a leading sentence clarifying what the respondent is being asked to consider when answering each item.
Questionnaire Matrix Sampling Design	A questionnaire design where each respondent receives only a subset of items in the entire questionnaire. In a *within-construct* matrix sampling design, a respondent answers items for all constructs but only receives a subset of items for each construct. In contrast, in a *construct-level* matrix sampling design entire constructs are rotated across questionnaire booklets.
Response Options	A typically verbally labelled set of answer choices provided to respondents for close-ended multiple-choice questions.
Scaled Index	An index or measure based on the scaling (using item response theory) of multiple items that all are indicators of the same underlying construct.

Table 5.2. Glossary of acronyms

Acronym	Term
CIPO	Context-Input-Process-Output Model
CT	Creative Thinking
ESCS	PISA index of Economic, Social, and Cultural Status
FT	PISA Field Trial
GCM	PISA Global Crises Module
ICQ	PISA Information and Communication Technology Questionnaire
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
LSA	Large-scale Assessment
MEG	PISA Mathematics Expert Group
MS	PISA Main Survey
OECD	Organisation for Economic Co-operation and Development
OTL	Opportunity to Learn
PGB	PISA Governing Board
PISA	Programme for International Student Assessment
PISA-D	PISA for Development
QEG	PISA Questionnaire Expert Group
SCQ	PISA School Questionnaire
SES	Socioeconomic Status

SSES	OECD Survey of Social and Emotional Skills
STQ	PISA Student Questionnaire
TAG	PISA Technical Advisory Group
TALIS	OECD Teaching and Learning International Survey
WBQ	PISA Well-being Questionnaire

Introduction

Aims of the PISA Questionnaires

Large-scale Assessments (LSAs) play an important role in evaluating education systems in terms of their capacity to develop human potential, advance progress and the quality of life of individuals across the globe and prepare future workforces for 21st century demands. Since its inception in the late 1990s, the Programme for International Student Assessment (PISA) has been known for its important contribution to education policy discussions within the Organisation for Economic Co-operation and Development (OECD) and partner countries and economies.

The main features of PISA are as follows:

- PISA is a *system-level assessment*, representing a commitment by governments to monitor the outcomes of education systems.
- PISA is *policy-oriented*, linking data on students' learning outcomes with data on key factors that shape learning in and out of school.
- PISA is *carried out regularly*, enabling countries and economies to monitor their progress in meeting key learning objectives.
- PISA *assesses both subject matter knowledge*, on the one hand, and *the capacity of individuals to apply that knowledge creatively*, including in unfamiliar contexts, on the other.
- PISA focuses on *knowledge and skills towards the end of compulsory schooling*. In most countries and economies, the end of compulsory education is around the age of 15, where students are supposed to have mastered the basic skills and knowledge to continue to higher education or the workforce.
- PISA is designed to *provide comparable data across a wide range of countries and economies.* Considerable efforts are devoted to achieving cultural and linguistic breadth and balance in assessment materials.
- PISA is a *collaborative effort* involving multiple parties including the OECD, PISA Governing Board (PGB), OECD member countries, and partner countries and economies.

PISA continues to yield indicators of the effectiveness, efficiency, and equity of educational systems, setting benchmarks for international comparison and monitoring trends over time. PISA also builds a sustainable database that allows researchers worldwide to study basic as well as policy-oriented questions on education, including those related to society and economy. The OECD and the PGB continue to look for ways to increase the scientific quality and policy relevance of the PISA context questionnaires to meet these needs.

Since the first cycle of PISA in 2000, the student and school context questionnaires have performed two interrelated purposes in service of the broader goal of evaluating educational systems:

- first, the questionnaires provide a context for interpreting the PISA results both within and between education systems;
- second, the questionnaires aim to provide reliable and valid measurement of *additional educational constructs*, which can further inform policy and research.

Over the seven cycles of PISA to date, education policy discussions have shifted from a heavy focus on the first objective to an increased focus on the second aim as well. This development corresponds to a shift in policymakers' views of the core goals for education systems in the 21st century, away from primarily teaching clearly defined subject knowledge and skills, to fostering broader skills (such as creativity, communication, collaboration, or learning to learn) that help individuals face the demands of a technology-rich and truly global society (UN, 2015[1]). There is now a growing recognition that other factors and

competencies aside from subject-specific knowledge play a vital role in fostering students' success in school and beyond. In order to understand and guide policy decisions regarding student development, the PISA 2022 context questionnaires will strengthen the measurement of the contexts that promote learning in these areas, as well as an array of general constructs of policy relevance.

The COVID-19 pandemic that emerged globally in early 2020 will likely have short- and long-term impacts on schooling and students' learning in these areas. The PISA 2022 context questionnaires will therefore also collect information on COVID-19-related disruptions to students' learning and well-being in participating education systems. This information can provide context for understanding PISA 2022 results, as well as serve to advance policy discussions about fostering the resiliency of students, schools, and education systems in responding to educational disruptions arising from ongoing and future global crises.

Outline of the PISA 2022 Context Questionnaire Framework

The PISA 2022 context questionnaire framework explains the goals and rationale for selecting specific questionnaire content for the eighth cycle of PISA. Like prior frameworks, the present framework touches upon how measured constructs theoretically relate to one another and to student achievement. Additionally, the framework outlines a set of survey design principles and methodologies that are introduced to PISA 2022 with the aim of improving measurement, efficiency, and consistency of PISA in the mid to long term. To achieve these goals, the framework is structured as follows:

- *Section 2.* describes a set of general considerations that led to the development of this framework and that guided instrument development for PISA 2022. These considerations included priorities for re-administration of questions from previous PISA cycles, changes to the mathematics framework since PISA 2012 that needed to be considered when prioritizing questionnaire constructs, country-specific needs across the range of participating countries and economies, directions taken with the PISA 2022 innovative domain of creative thinking, and plans for optional questionnaires.

- *Section 3.* presents the PISA 2022 two-dimensional framework taxonomy. The first dimension classifies proposed constructs into the two overarching categories distinguished by the PGB (domain-specific constructs and general constructs, with the latter including Economic, Social, and Cultural Status [ESCS]). The second dimension classifies proposed constructs into five categories based on key areas of educational policy setting at different levels of aggregation (Student Background; Student Beliefs, Attitudes, Feelings, and Behaviours; Teaching Practices and Learning Opportunities; School Practices, Policies, and Infrastructure; and Governance, System-Level Policies and Practices). Linkages between the 2022 approach and the overarching cross-cycle structure developed across the PISA 2000 – 2018 questionnaire frameworks are highlighted, with a focus specifically on the past three PISA cycles, i.e., 2012, 2015, and 2018 (OECD, 2013[2]; Klieme, 2014[3]; OECD, 2013[4]).

- *Section 4.* gives a detailed overview of the questionnaire modules and constructs measured in the MS which were selected for inclusion based on analysis of FT data and discussion of priorities among experts and policy makers (including the PGB).

- *Section 5.* summarizes the survey design principles that guided the PISA 2022 questionnaire development process, subsequent FT administration, and post-FT analyses and item selections for the MS.

Balancing re-administration of questions from previous cycles with new development

For PISA 2022, the PGB recommended re-balancing questionnaire content in the direction of a larger focus on general constructs and a slightly reduced focus on domain-specific constructs. Specifically, the PGB suggested that 40% of the content be devoted to domain-specific constructs. The remaining 60% of content focused on general constructs would be split between 20% devoted to measuring ESCS and 40% focused on other general constructs, including additional outcomes (PISA Governing Board, 2017[5]). By contrast, in 2018 the balance of questionnaire content across domain-specific constructs, ESCS, and general constructs was 50%, 17%, and 33%, respectively.

It was suggested that percentages be allocated based on estimated questionnaire administration time. For the PISA 2022 MS, of the allocated testing time for the student questionnaire (STQ) is 35 minutes. That is, approximately seven minutes of the STQ is devoted to ESCS and 14 minutes each are devoted to domain-specific and general constructs. Within the boundaries of these overall strategic priorities, two key areas of consideration guided the development of the PISA 2022 context questionnaire framework: (1) re-administration of questions from previous PISA cycles and (2) new development.

Guidelines for re-administration of questions from previous years

A key force driving the PISA design in general is the cyclical change of focus in the cognitive assessment. Mathematics was the major domain of cognitive assessment in PISA 2003 and 2012 and is the major domain again in 2022. Reading was the major domain of assessment in PISA 2000, 2009, and 2018. Science was the focus of PISA 2006 and 2015. The major domain serves as the primary focus of domain-specific content in the associated PISA context questionnaires (e.g., various mathematics-related constructs marked the focus of the 2003 and 2012 questionnaires).

In order to describe educational constructs of interest over time at the country or economy level, it is desirable to maintain a stable set of questionnaire measures that can be used as major reporting variables across PISA cycles. Given the cyclical nature of PISA, measurement stability can be considered at two levels:

- first, there is the issue of stability of measures across cycles of three years (i.e., administration of items for constructs that may appear in every cycle, e.g. ESCS);
- second, stability is desirable in measuring domain-specific constructs across cycles of nine years (i.e., mathematics-specific constructs assessed in the 2012 and/or 2003 cycles).

A priority of PISA 2022 has been to retain a reasonable number of questions that have been administered in previous PISA questionnaires. Table 5.3. summarizes guidelines used for decisions about retention or deletion of previously administered PISA items.

Table 5.3. Guidelines for retention or deletion of PISA questions from previous cycles

Guidelines for retention or deletion of PISA questions from previous cycles
1. Retain questions that best explain variations in academic achievement within and across countries;
2. Retain questions that are of highest policy relevance and/or necessary to establish or extend trend lines, which can inform policy and research;
3. Where possible and sensible, carry forward constructs intact, or with only minor changes that improve measurement precision;
4. Delete or revise questions that are outdated (e.g., questions that reference resources or technologies that are no longer in use);
5. Delete or revise questions that do not meet PISA 2022 psychometric criteria established by the Technical Advisory Group (TAG); and
6. Delete or shorten questions that provide information that is redundant with other questions or items within a matrix question.

Guidelines for new development

PISA has been making efforts to innovate in educational measurement. Over its past cycles, the program has, for instance, introduced new technologies (e.g., computer-based assessment [CBA]); expanded into measuring new innovative domains (e.g., collaborative problem solving in 2015, global competency in 2018, and creative thinking in 2022); updated its view on the measurement objectives for its major domains based on new frameworks; and has reacted to the emergence of new policy priorities (e.g., measuring student health and well-being as well as other social and emotional characteristics; measuring the impact of COVID-19-related disruptions on student learning and well-being).

For PISA 2022, the scope of the mathematics framework has been expanded to evaluate students' mathematical reasoning grounded in six core concepts or "big mathematical ideas" that undergird the specific content, skills, and algorithms of school mathematics (PISA Governing Board, 2017[5]):

1. Quantity, number systems and their algebraic properties;
2. Mathematics as a system based on abstraction and symbolic representation;
3. Mathematical structure and its irregularities;
4. Functional relationships between quantities;
5. Mathematical modelling as a lens onto the real world (e.g., those arising in the physical, biological, social, economic, and behavioural sciences); and
6. Variation as the heart of statistics.

Students will also be assessed in their familiarity with, or prior classroom exposure to, four emerging areas of mathematics content in which reasoning skills need to be applied: computer simulations, growth phenomena, conditional decision making, and geometric approximation. The questionnaire framework has been updated accordingly to better understand students' opportunities to learn these concepts, as well as the extent to which 21st century skills are emphasized in mathematics instruction.

Additionally, creative thinking will be assessed as the innovative domain in PISA 2022. A distinct module of the PISA 2022 context questionnaires is devoted to constructs that contribute to the understanding of students' performance in this innovative domain.

Several new educational systems will participate in PISA beginning in 2022, many of which belong to lower- and middle-income countries. In order to maximize the value of PISA to these participants, the context questionnaires include constructs related to student background and learning contexts that have previously been described in the PISA for Development (PISA-D) framework (OECD, 2018[6])

New development makes use of informed practices in survey methodology (e.g., principles regarding item types, response options, balancing of scales, length of matrix questions) and technological capabilities (e.g., routing, matrix sampling) to the extent that they enhance measurement. Section 5 of this framework elaborates on the survey design principles that guided PISA 2022 questionnaire development.

While this framework focuses on the conceptual underpinnings of the PISA questionnaires for students and schools, additional frameworks that are not part of this document provide in-depth theoretical foundation for additional questionnaires included in PISA 2022 as part of international options (i.e., frameworks for Financial Literacy, Information and Communication Technology [ICT] Literacy, Student Well-being, Teacher Well-being).

Table 5.4. summarizes guidelines used for considering the addition of new items for existing constructs as well as entirely new constructs in PISA 2022.

Table 5.4. Guidelines for new development

Guidelines for addition of new items for existing constructs as well as entirely new constructs
1. Develop questions for new constructs that are central to the educational research literature and the PGB priorities;
2. Develop new questions that are relevant to changes to the PISA mathematics framework (e.g., addition of mathematical reasoning);
3. Develop new questions for constructs that are related to the innovative domain assessed in PISA 2022 (i.e., creative thinking);
4. Develop new questions to replace previously used questions that do not comply with PISA 2022 psychometric criteria, substantially violate PISA 2022 survey design principles, and/or require updates to more accurately describe students' living and learning realities;
5. Develop new questions to replace previously used questions that do not offer sufficient flexibility to meet country- or region-specific needs of all participating education systems; and
6. Develop new questions to help shed light on the impact of the COVID-19 pandemic on student learning and well-being, and the degree of interruptions or changes to education across participating education systems.

PISA 2022 Context Questionnaire Framework taxonomy

Beginning with the questionnaire framework used for the PISA 2009 assessment, questionnaire content was explicitly linked to different levels of the education system: the student level, level of instruction in the classroom, school level, and system level (Jude, 2016). The questionnaire framework used for PISA 2012, and subsequently refined for PISA 2015 and 2018, further underscored the importance of collecting information on learning contexts for comparative system monitoring. These frameworks outlined an overarching two-dimensional structure of high-level questionnaire content areas to be measured and kept comparable across assessment cycles (OECD, 2013[4])

The theoretical foundation of the 2012 overarching framework is based on Purves' (1987[7]) *Context-Input-Process-Outcome* (CIPO) model. In the CIPO model, contextual variables for understanding education systems are conceptualized as a series of inputs (i.e., student background), processes (i.e., teaching and learning, school policies, governance), and outcomes (i.e., performance and non-cognitive outcomes) shaped at the student, classroom, school, and country levels. Starting with PISA 2015 and 2018, an additional dimension further classified questions more explicitly into domain-specific and domain-general modules. Domain-specific modules represent the set of constructs with strong expected relationships to student experiences, outcomes, and teaching and learning factors tied to a specific content area (e.g., reading, mathematics, or science). Domain-general modules represent the set of constructs that are important for understanding differences in achievement that are not tied to a specific subject-area. Figure 5.1 illustrates the high-level structures of the context questionnaire frameworks from 2012, 2015, and 2018.

Figure 5.1. Framework structures of PISA 2012, 2015, and 2018
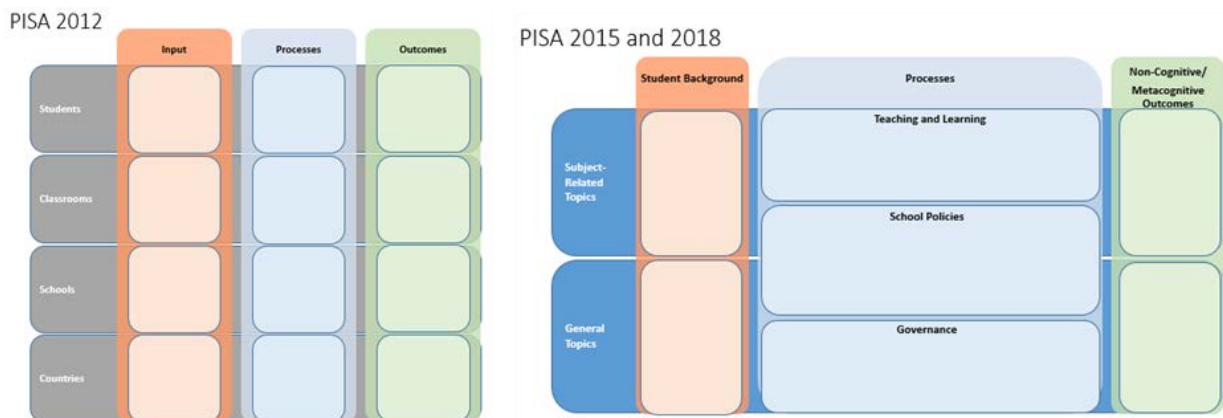
In keeping with the long-term goal of balancing continuity with innovation, the PISA 2022 context questionnaire framework retains key framework elements from previous cycles as a foundation, and introduces refinements that facilitate the strategic development of new constructs and move toward improved measurement. This updated framework structure is illustrated in Figure 5.2 below. Please note, while performance and contextual variables have been classified as "outcomes" in previous PISA frameworks per the CIPO model, both types of variables also constitute possible inputs (OECD, 2013[4]). For instance, a student's prior achievement and his/her curiosity, perseverance, achievement motivation, or confidence will likely impact the student's future achievement, as well as his/her future development of social and emotional characteristics. Due to the cross-sectional nature of PISA, variables collected through the questionnaires cannot be clearly assigned a single "role". While the CIPO model remains useful to describe an actionable policy perspective and serve as a helpful theoretical perspective for researchers on the variables measured with the PISA questionnaires, it seems less useful as a guide to classify and prioritize variables for instrument development. Due to the ambiguity in classifying variables, constructs are *not* classified as inputs, processes, or outcomes in the PISA 2022 framework taxonomy. Instead, we allude to the possible roles each variable might play in the detailed descriptions of each module. Further description of the framework dimensions and the modules is provided in subsequent sections of this framework.

Across the two overarching (vertical) framework content dimensions and of the five (horizontal) policy focus areas as shown in Figure 5.2, a total of 21 modules are specified (see Section 4. of this document). The small boxes in the taxonomy below indicate the relative distribution of constructs in the PISA 2022 MS across all modules described in this framework.

Figure 5.2. PISA 2022 two-dimensional Context Questionnaire Framework taxonomy
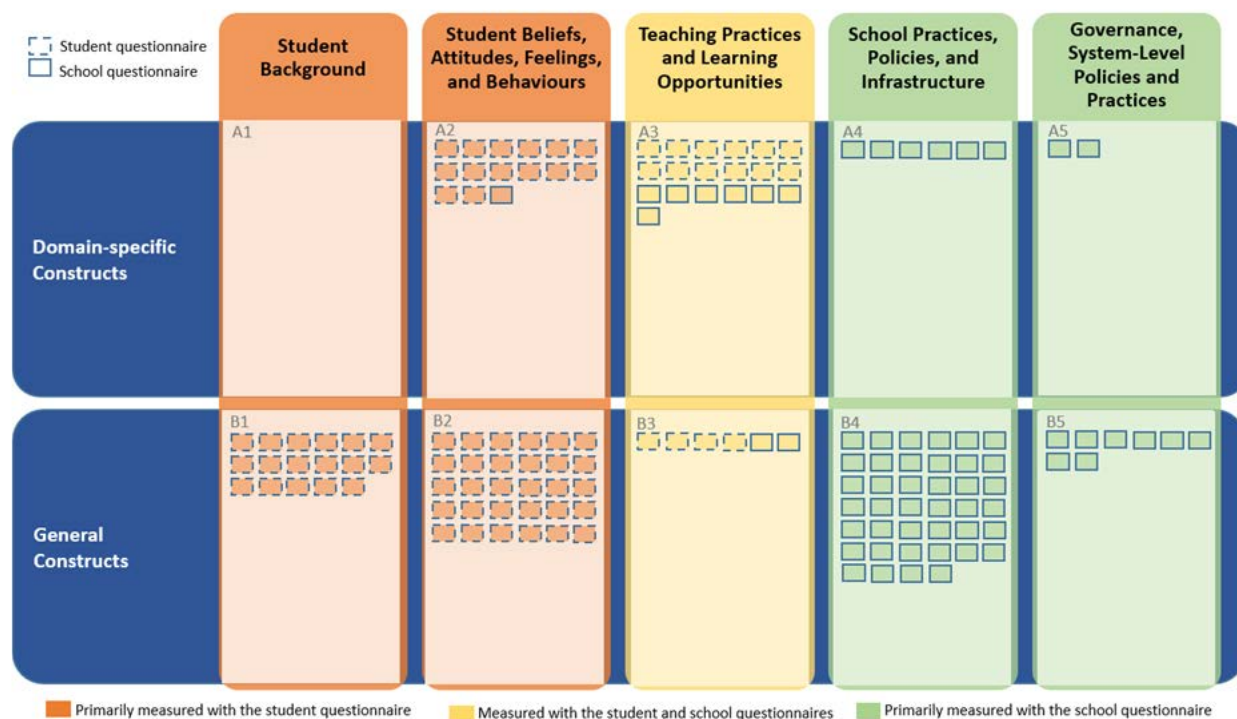
Classification based on relationships to PISA content domains

As outlined above, the PISA 2022 student and school questionnaires serve two interrelated purposes (i.e., to provide contextual information and provide additional measures) in service of the broader goal of evaluating the effectiveness of all educational systems participating in the 2022 MS.

The two categories along the vertical dimension of the taxonomy in Figure 5.2 represent the primary types of content in the student and school questionnaires:

1. *Domain-specific* Constructs; and
2. *General* Constructs (including ESCS).

Both categories of constructs represent questions that are included in PISA primarily to report their relationships with academic achievement and provide a context for interpreting the PISA results within and between education systems, as well as constructs that are included in PISA primarily to report additional variables that describe educational systems beyond academic achievement to inform policy and research.

Domain-specific constructs

Domain-specific constructs include constructs that demonstrate a relationship to students' academic achievement in the major domain of the current cycle (i.e., mathematics for PISA 2022) or hold power to explain broader outcomes in the major domain, such as students' educational career and post-secondary aspirations (e.g., course enrolment, outlook on future educational career). Examples of indicators include mathematics-related school curricula or students' interest and motivation to learn mathematics topics. Constructs that are included primarily to better understand differences in achievement in the PISA 2022 mathematics achievement scores were evaluated empirically after the FT according to their relationship with mathematics achievement to determine their inclusion in the PISA 2022 MS. The mathematics-specific constructs included in the PISA 2022 MS are summarized in Table 5.5 bellow.

In addition to constructs related to the major domain (i.e., mathematics), a smaller number of contextual variables specific to all three domains (including the two minor domains of this assessment cycle, Reading and Science) are included in the PISA 2022 MS questionnaires to provide relevant contextual information for student achievement. Lastly, the category of domain-specific constructs includes several creative thinking-related constructs that aim to contextualize achievement results in the PISA 2022 innovative domain.

Table 5.5. Mathematics-specific constructs in PISA 2022 Student and School Questionnaires

Student Questionnaire Mathematics Constructs
Subjective familiarity with mathematics concepts
Exposure to formal and applied mathematics tasks
Exposure to mathematics reasoning and 21st century mathematics topics
Perceived quality of mathematics instruction
Disciplinary climate in mathematics
Cognitive activation in mathematics: Encourage mathematical thinking
Cognitive activation in mathematics: Foster reasoning
Class periods per week in mathematics
Time spent on homework
Participation in additional mathematics instruction (types)
Mathematics teacher support
Growth mindset
Favourite subjects and self-concept in mathematics, test language, and science
Mathematics self-efficacy: Formal and applied mathematics

Mathematics self-efficacy: Reasoning and 21st century mathematics
Mathematics anxiety
Effort and persistence in mathematics

School Questionnaire Mathematics Constructs

Use of mathematics achievement data in accountability systems
Standardised mathematics curriculum
Use of mathematics assessments
Average time in class period
Average number of students in mathematics classes
Selection of courses
Percent of students who received marks below, at/above the pass mark in mathematics class their last school report
Student ability grouping in mathematics
School offering additional mathematics lessons
Mathematics teacher qualifications
Mathematics teacher training

General constructs

General constructs include constructs that demonstrate relationships to students' academic achievement across multiple domains, such as students' feelings towards school (e.g., student-teacher relationships, bullying experiences), school infrastructure (e.g., availability of digital technology for learning), or constructs that complement traditional indicators of educational effectiveness (e.g., subjective well-being, social and emotional characteristics). General constructs also include ESCS to assess students' socioeconomic status (SES) and the equity of educational opportunities within and across educational systems.

Classification based on Educational Policy Areas

The horizontal dimension of the taxonomy distinguishes five categories of educational policy focus that correspond to different aggregate levels for the collected survey responses, from individual-level variables to highly aggregated system-level indicators:

1. Student background;
2. Student beliefs, attitudes, feelings, and behaviours;
3. Teaching practices and learning opportunities;
4. School practices, policies, and infrastructure; and
5. Governance, system-level policies and practices.

Student background

The first educational policy area of interest relates to *Student Background.* In order to understand students' education pathways and to study equity within and across educational systems, basic demographic variables (e.g., gender, age, or grade), constructs related to ESCS, migration and language background, as well as information about students' early years must be considered. The distribution of educational opportunities and outcomes correlated with these background constructs may provide data about whether countries succeed in providing equity in educational opportunities.

Student beliefs, attitudes, feelings, and behaviours

The second educational policy area of interest focuses on *Student Beliefs, Attitudes, Feelings, and Behaviours.* In addition to measuring 15-year-olds' academic achievement in reading, mathematics, science, and creative thinking, measures of students' subjective attitudes and feelings, as well as their

behavioural choices may provide important indicators for an education system's success in fostering productive members of society.

Beliefs include constructs such as beliefs about learning or student's mindsets. Attitudes include constructs such as students' attitudes towards mathematics, or attitudinal aspects of social and emotional characteristics. Feelings concern feelings about their school or about specific subject-areas, and emotional aspects of social and emotional characteristics. Behaviours include participation in activities outside of school or behavioural aspects of social and emotional characteristics. Constructs such as respecting and understanding others, being motivated to learn and collaborate, or being able to regulate one's own behaviour may play a role as prerequisites of acquiring subject-area knowledge and skills. In addition, such characteristics may also be judged as goals of education in their own right (Almlund et al., 2011[8]; Bertling, Marksteiner and Kyllonen, 2016[9]; Heckman, Stixrud and Urzua, 2006[10]; Rychen and Salganik, 2003[11]).

Each of the past seven PISA cycles have included a significant number of questions tapping into students' beliefs, attitudes, feelings, and behaviours related to the major domain. In addition, recent PISA cycles have increased their focus on general constructs (e.g., "Noncognitive outcomes" modules in PISA 2015 and PISA 2018). PISA 2022 carries these developments forward and includes several modules addressing a range of constructs such as students' effort on the PISA test and questionnaires (Module 5), students' general school-related attitudes and feelings associated with school climate (Module 6), attitudes towards specific PISA content domains (Module 7), and students' general social and emotional characteristics (Module 8). A broad range of student behaviours are further assessed via a module focused on out-of-school experiences (Module 10). In addition, students' subjective views on their socioeconomic standing, as well as their future aspirations and well-being, are captured in modules 2, 3, and 9, respectively.

Teaching practices and learning opportunities

The third educational policy area of interest pertains to *Teaching Practices and Learning Opportunities*. Classroom-based instruction is the immediate and core setting of formal, systematic education. Therefore, policy makers need information on the organisation of classrooms and the teaching and learning experiences that occur within them. The knowledge base of educational effectiveness research (e.g. (Scheerens and Bosker, 1997[12]; Creemers and Kyriakides, 2007[13]) allows for the identification of core variables with an expected bearing on mathematics and student achievement in general, for example, teachers' qualifications, teaching practices and classroom climate, learning time, and learning opportunities provided in and outside of school. As such, this policy area closely links to the idea of *opportunity to learn* (OTL), which was first introduced by Carroll (Carroll, 1963[14])to indicate whether students have had sufficient time and received adequate instruction to learn (Abedi, 2006[15]). Though the meaning of OTL has since broadened, it has been an important concept in international student assessments (e.g., (Schmidt, 2001[16]) and shown to be strongly related to student performance in cross-country comparisons (Schmidt, 2009[17]).

Researchers have suggested defining OTL not only based on subject-specific teacher instruction (Callahan, 2005[18]; McDonnell, 1995[19]); and have stressed the importance of evaluating the quality of instruction in addition to mere quantity (Duncan and Murnane, 2011[20]; Little and Bell, 2009[21]; Minor et al., 2015[22]). Researchers have also pointed out the importance of informal learning opportunities and experiences in the home (Lareau and Weininger, 2003[23])and highlighted the need to evaluate OTL in country-specific contexts (Cogan and Schmidt, 2014[24]). Accounting for these broader directions, OTL could be defined as all contextual factors that capture the cumulative learning opportunities a student has been exposed to at the time of the assessment (Bertling, Marksteiner and Kyllonen, 2016[9]). These contextual factors may comprise both learning opportunities at school and informal and formal learning opportunities outside of school. In this framework, several aspects of OTL are captured across different modules, including modules capturing opportunities provided through the ways in which student learning is organised (Module 14), opportunities defined based on the mathematics content students are exposed

to (Module 15), and opportunities created based on the behaviours teachers exhibit in the classroom (Module 16).

School practices, policies, and infrastructure

The fourth educational policy area of interest examines *School Practices, Policies, and Infrastructure.* As policymakers have limited direct impact on teaching and learning processes, information on school-level factors (e.g., practices, policies, and infrastructure) that help to improve schools, and thus indirectly improve student learning, are a priority. In addition to individual student demographics and structural factors (such as school location, school type, and school size), the social, ethnic, and academic composition of the school influences students' learning processes and outcomes. Therefore, PISA uses aggregated student data to characterize demographic and other contextual factors at the level of the school community.

Similar to the *Teaching Practices and Learning Opportunities* modules and constructs, school effectiveness research has reported that "essential supports" are associated with school effectiveness (Bryk et al., 2009[25]). These essential supports include leadership and school management; well-organised curriculum, instructional, and enrolment policies; tangible resources; positive school climate; and parent or guardian involvement. Educational psychologists also emphasise teachers' collective efficacy, principals' leadership, parent or guardian involvement, and peer support as crucial for creating a positive school climate conducive to learning (LEE and SHUTE, 2010[26]). Many of these factors have been previously addressed in the PISA questionnaires as domain-general processes on the school level. Also covered is school-level support for teaching the major domain, such as the provision of learning resources and space, information and communication technology (ICT), and a school curriculum for mathematics education.

Governance, system-level policies and practices

Finally, the fifth educational policy area of interest focuses on *Governance, System Level Policies and Practices.* To meet policy requests directly, PISA also needs to address issues related to governance at the system level (Hanushek and Woessmann, 2011[27]). For instance, assessment and evaluation are basic processes that policy makers and/or school administrators use to control school quality, and to monitor and foster school improvement. These issues have been previously examined in the PISA questionnaires as domain-general context variables on the system level; domain-specific system-level context variables are also included in PISA 2022. While some information is collected through the PISA school questionnaire (SCQ), additional information can potentially be acquired by researchers and policymakers from other sources (e.g. system-level data, administrative records).

Detailed overview of PISA 2022 modules

Basic demographics

PISA questionnaires have routinely included questions on students' gender and age, as well as their grade. These questions are included again in the STQ for the PISA 2022 MS.

The PISA 2022 FT explored updates to basic demographic questions on home composition to better reflect modern living realities in traditional as well as non-traditional homes and to establish a foundation for potential routing throughout the questionnaire based on, for instance, the students' number of parents or guardians. In order to maximize the strength of trendlines to data from previous cycles in light of the disruptions to education caused by the COVID-19 pandemic, these updated questions will not yet be included in the 2022 MS. This FT exploration marked, however, an important milestone towards a more modern description of students' homes in PISA in the mid to long term.

Figure 5.3. illustrates how all constructs in this module map onto the taxonomy.

Figure 5.3. Constructs in basic demographics module
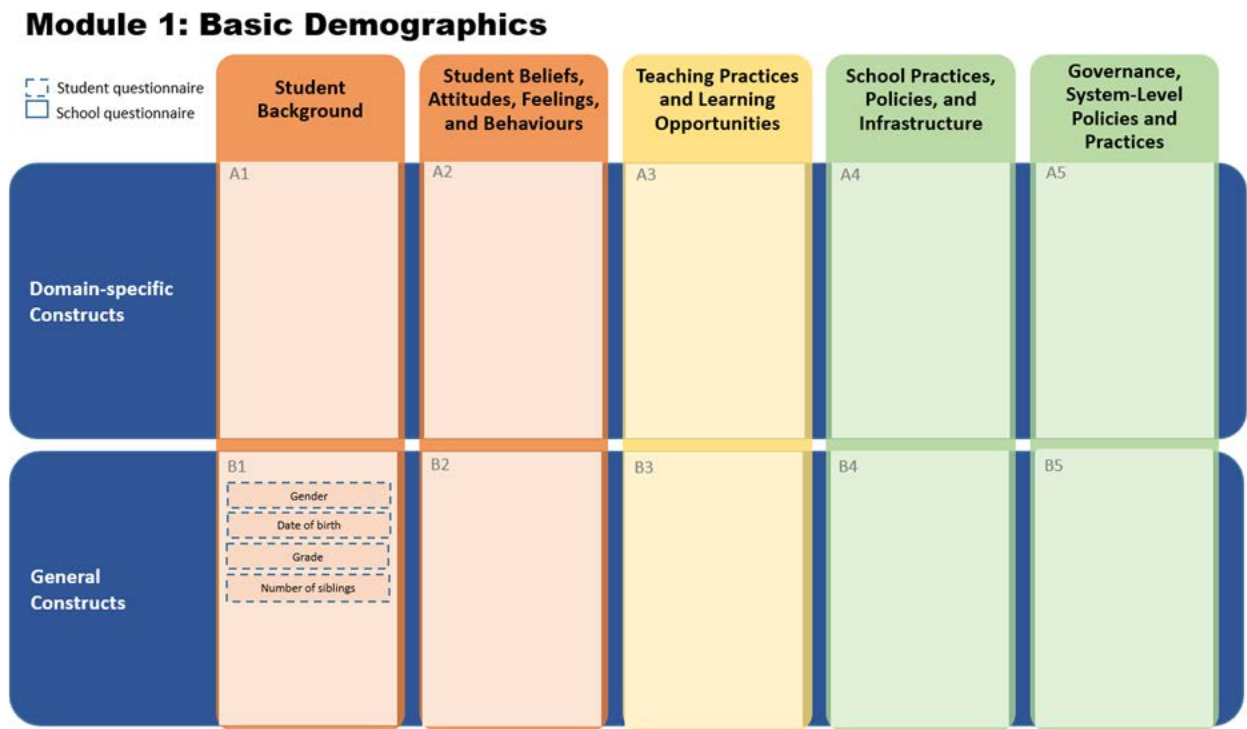
Economic, Social, and Cultural Status (ESCS)

Over the past seven PISA cycles, significant efforts have gone into the definition and operationalization of individual student background indicators, leading to the establishment of an integrated indicator for students' ESCS (Willms, 2006[28]; Lee, Zhang and Stankov, 2019[29]). Figure 5.4. displays how ESCS was created in the two most recent PISA cycles.

Figure 5.4. Computation of ESCS Index in PISA 2015 and 2018 (From PISA 2015 Technical Report)
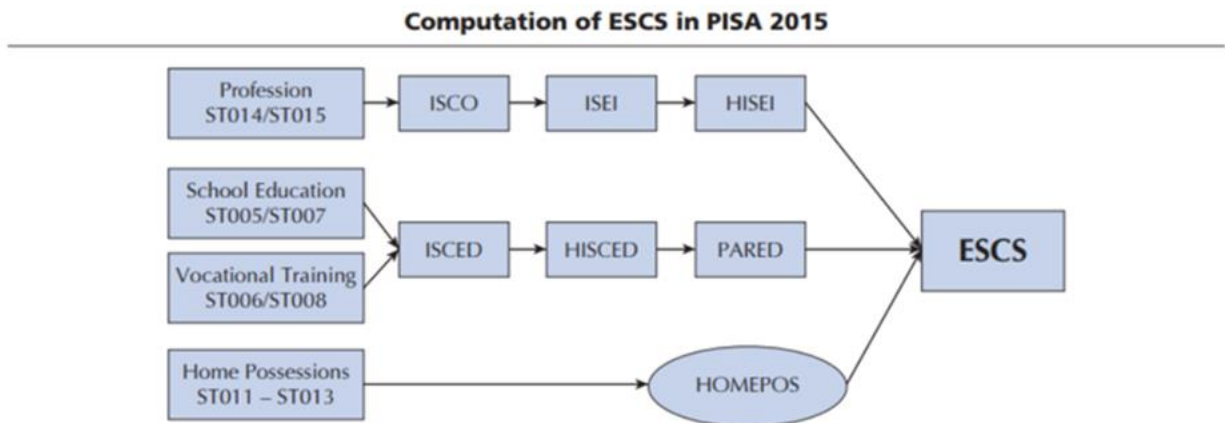
Note: ISCO: International Standard Classification of Occupations; ISEI: occupational status of mother and father; HISEI: highest parental occupational status; ISCED: International Standard Classification of Education; HISCED: Highest education of parents (ISCED); PARED: Index for highest parental education in years of schooling; HOMEPOS: Index of home possessions (WLE); ESCS: Index of economic, social and cultural status.

The PISA ESCS index is considered internationally as a gold standard measure of socioeconomic status (SES) in LSAs (e.g. (Cowan, 2012[30])). To examine trends over time and comparisons with previous PISA data on the ESCS index, it is crucial to establish minimal stability in assessing the three components. While well established, the ESCS index has also been criticized in recent years (e.g. (Rutkowski and Rutkowski, 2013[31])), calling for revisions and extensions of the index.

Few changes have been made over the years to the measurement of ESCS in PISA, resulting in current approaches only partly accounting for students' living realities within and across the much more diverse PISA population. This issue becomes more pressing with the number of participating countries more than doubling over the past cycles. For instance, the current PISA ESCS questions continue to assume a traditional nuclear family with a mother and father and give little to no room for students to provide information about their families' income and education levels if they live in non-traditional constellations (e.g. multiple households, same-sex parents, multi-generational households, etc.).

While used for several cycles, issues remain with the International Standard Classification of Occupations (ISCO) and International Standard Classification of Education (ISCED) coding of parental educational levels and occupations (Kaplan and Kuger, 2016[32]) that pose challenges when making international comparisons on the respective questions. Recent findings from other studies further suggest that student reports on their parents' occupation tend to be very inaccurate, produce larger proportions of missing values, and that these questions take substantially more time to answer than other survey questions (e.g. (Tang, 2017[33])). Note that in previous PISA cycles, information about education levels among parents has been based on ISCED 1997 classifications; beginning with PISA 2022, the more recent ISCED 2011 classifications will be used. Table 5.6. summarizes how the updated ISCED 2011 levels correspond to the ISCED 1997 levels. More detailed information about the correspondence or concordance between levels in the ISCED 2011 classification and the earlier ISCED 1997 framework can be found in *the ISCED 2011 Operational Manual: Guidelines for Classifying National Educational Programmes and Related Qualifications* (UNESCO Institute for Statistics/OECD/Eurostat, 2015[34]).

Table 5.6. Correspondence between ISCED 2011 and ISCED 1997 Levels

	ISCED 2011		ISCED 1997
01	Early childhood educational development		--
02	Pre-primary education	0	Pre-primary education
1	Primary education	1	Primary education or first stage of basic education
2	Lower secondary education	2	Lower secondary education or second stage of basic education
3	Upper secondary education	3	(Upper) secondary education
4	Post-secondary non-tertiary education	4	Post-secondary non-tertiary education
5	Short-cycle tertiary education	5	First stage of tertiary education (not leading directly to an advanced research qualification) (5A, 5B)
6	Bachelor's or equivalent level		
7	Master's or equivalent level		
8	Doctoral or equivalent level	6	Second stage of tertiary education (leading to an advanced research qualification)

The PGB has expressed a desire to increase the benefits of participation in PISA for lower- and middle-income countries. The group has further expressed a need to incorporate questionnaire items that fully reflect the context found in those countries. The broadening of the PISA population to new countries and the widened socioeconomic divides in some countries call for a better approach of assessing the entire range from low to high socioeconomic circumstances. Having common questions between the PISA-D student and out-of-school youth questionnaires and the PISA STQ could be one way of achieving that linkage. The MS questionnaire will therefore include a broader set of home possession items than previous cycles as well as additional poverty indicators (e.g., food insecurity).

In addition to these updates, a range of more fundamental potential changes to the Index of ESCS were explored in the PISA 2022 FT, specifically replacing parent-focused with guardian-focused questions and replacing fill-in with multiple choice occupation questions. While these explorations resulted in findings that will help shape the mid- to long-term enhancement of the PISA questionnaires, the nature of the three main components of the Index of ESCS (Parental Occupational Prestige, Parental Education, Home Possessions) will remain unchanged in PISA 2022. Minimizing bigger changes to the index, while not ideal from an inclusiveness perspective, will allow keeping trend lines on ESCS with past cycles as strong as possible in efforts to contextualize student learning and disruptions thereof due to the COVID-19 pandemic.

Complementing the ESCS index, which will allow for strong trend comparisons, PISA 2022 will also measure food insecurity and subjective socio-economic status to gain a fuller perspective of students' backgrounds and potential obstacles to educational success they may be facing. Research demonstrates that the types of family SES variables necessary for student achievement differ depending on the country's overall developmental status; traditional measures of parental/guardian educational and occupational levels were more relevant to student achievement in richer countries than in poor countries (Lee and Borgonovi, 2022[35]).

Research on subjective SES suggests that student's subjective beliefs about their own and their family's status can be as important as objective SES measures in predicting important outcomes, ranging from achievement and overall future aspirations, to obesity and other health outcomes (e.g. (Citro, 1995[36]; Demakakos et al., 2008[37]; Goodman et al., 2001[38]; Lemeshow et al., 2008[39]; Quon and McGrath, 2014[40])).The most common approach for measuring subjective SES is Cantril's Self-Anchoring Ladder ((Cantril, 1965[41]); see (Levin and Currie, 2014[42]), for an adaptation for adolescents). It has been used in several variations, including extensions to subjective social status within the school community (Goodman et al., 2001[38]). A subjective SES measure will complement rather than replace the established ESCS indicator in PISA.

Figure 5.5. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.5. Constructs in ESCS module
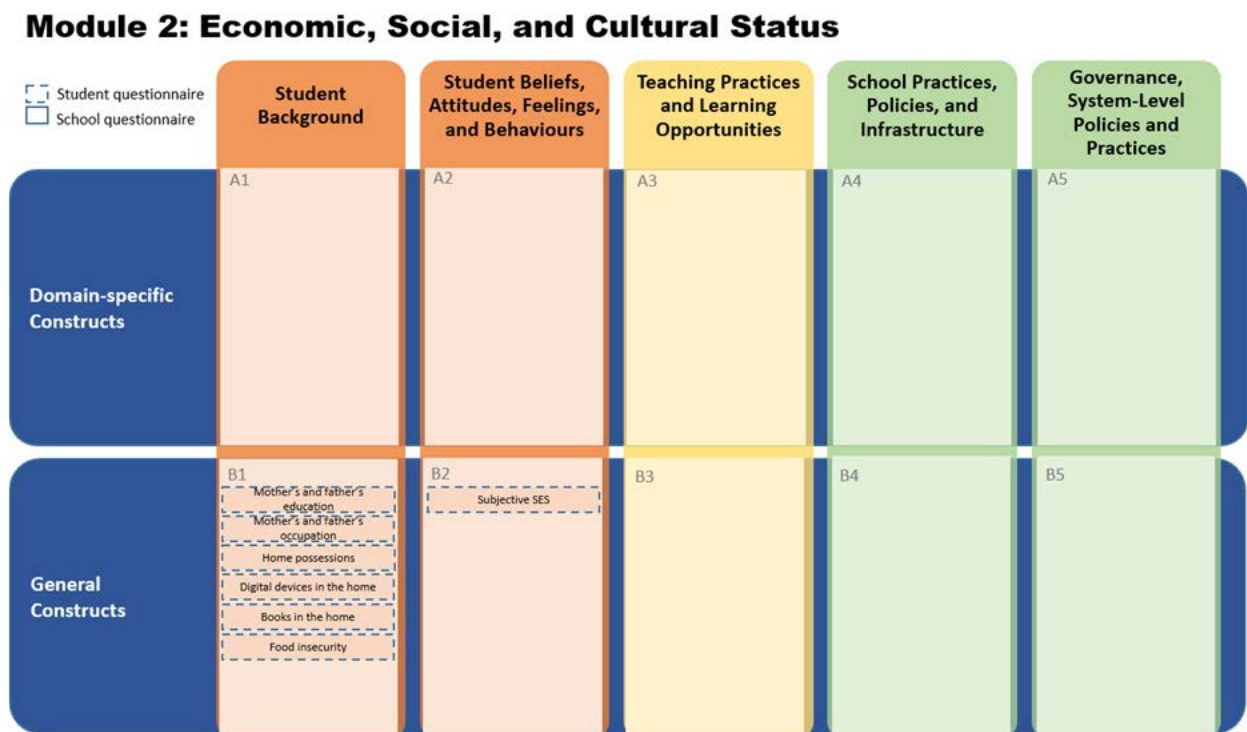
### *Educational pathways and post-secondary aspirations*

PISA gathers retrospective information about students' early years, educational pathways, and careers. Researchers and public debates in many countries have stressed the importance of early childhood education ((Blau and Currie, 2006[43]; Cunha et al., 2006[44]). PISA 2022 continues this tradition to capture essential information on primary and pre-primary education (bearing in mind that, for the most part, this would be solicited from 15-year-olds or their parents, which may pose validity challenges). Aspects of school attendance, such as truancy and grade repetition, are also captured as they have been found to significantly impact students' educational pathways. For example, school attendance problems have been linked to academic deficiencies including reduced educational performance, fewer literacy skills, and school dropout (Kearney et al., 2019[45]).

In addition to collecting data on students' early educational careers, previous PISA cycles have gathered prospective information about students' future educational pathways and preparation, and their occupational aspirations. While research in the United States has found that interpersonal relationships (e.g., peers, parents or guardians, teachers and staff who provide career guidance) play a significant role in shaping students' educational aspirations, cross-cultural research suggests that these influences may largely depend on the structural features of the educational systems in which they operate. For instance, peers and parents or guardians tend to influence educational aspirations in countries with undifferentiated secondary schooling, but this influence appears to be weaker in countries with more differentiated secondary education (Buchmann and Dalton, 2002[46]). It is possible that in differentiated systems, these effects may be indirect and mediated by early school-related decisions, such as track enrolment. An important factor to consider in understanding students' educational and work aspirations is the role that the school has in shaping these goals—for instance, through students' participation in the curriculum and activities offered by the school, and the provision of additional resources to explore educational and occupational pathways (e.g. (Beal and Crockett, 2010[47])).

Constructs measured in the STQ (e.g. attendance of ISCED 0-2; current study programme; history of students repeating a grade; missing, skipping, or arriving late to school; students' exposure to information about future studies or work; students' education and career expectations) and SCQ (e.g., school's support in providing information to students about future work and career paths) under this module are considered primarily as general constructs.

Figure 5.6. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.6. Constructs in educational pathways and post-secondary aspirations module


Module 3: Educational Pathways and Post-Secondary Aspirations
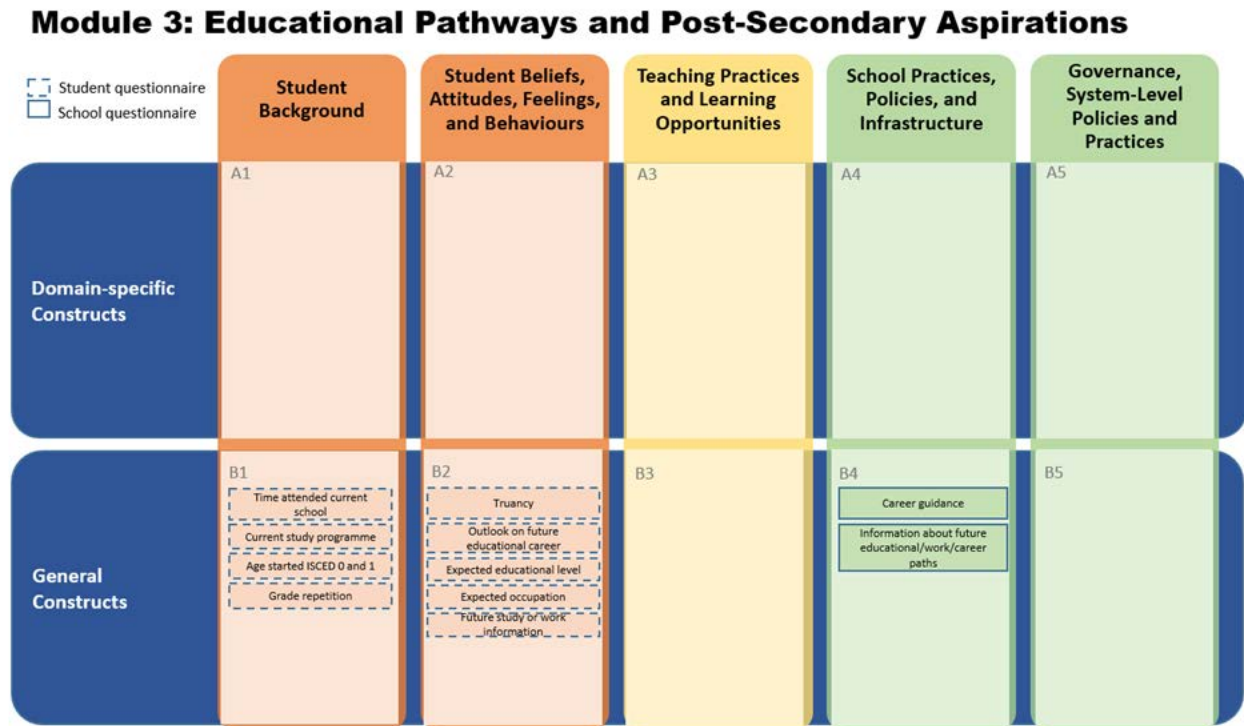
Migration and language exposure

Selected aspects of students' migration background and language exposure have been captured in previous PISA STQs as well as optional questionnaires (e.g., acculturation in the 2012 Educational Career Questionnaire). Immigration is currently a critical topic in many countries, particularly those with traditionally larger immigrant populations (e.g., the United States, Canada) as well as countries facing new challenges due to new populations of refugees (e.g., most central European countries) (Bansak, Hainmueller and Hangartner, 2016[48]; Wike, 2016[49]). Issues regarding the student's experience of a school climate that is accepting of diversity and multiculturalism are relevant to this module and overlap with content examined in the module on *School Culture and Climate* (Module 6)*.

STQ constructs in this module focus on assessing students' migration backgrounds (e.g., country of origin, age of arrival in country), and language backgrounds (e.g., primary language spoken at home). General constructs in the SCQ include the proportion of students with a migration background (e.g., immigrant or refugee status) and the number of languages taught at the school.

Figure 5.7. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.7. Constructs in migration and language exposure module

Module 4: Migration and Language Exposure
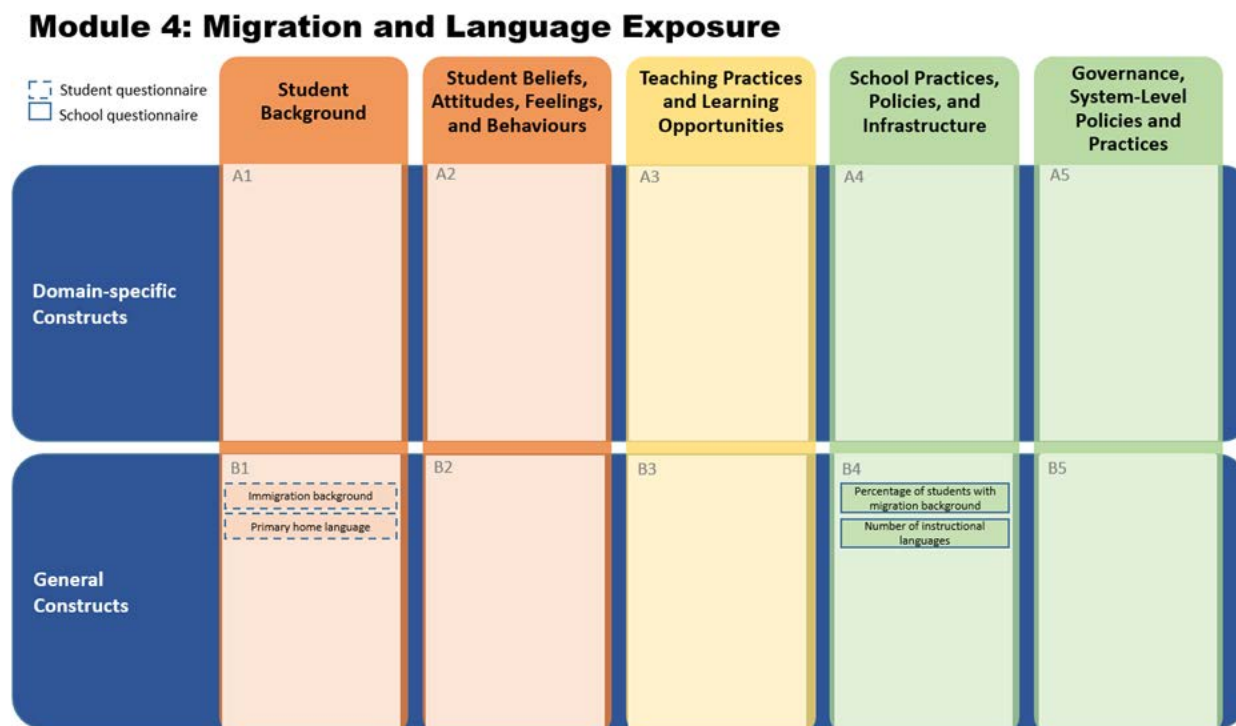
PISA preparation and effort

Several researchers have investigated questions of whether test-taker effort on low-stakes LSAs may impact achievement results or whether differential effort may play a role in explaining score differences between student groups or educational systems (e.g., (Debeer et al., 2014[50]; Eklöf, Pavešič and Grønmo, 2014[51]; Hopfenbeck and Kjærnsli, 2016[52]; Jerrim, 2015[53]; Penk, 2015[54]).

To inform educational policy regarding test-taker effort in PISA, this module covers students' subjective perceptions of how much effort they applied when answering the PISA test questions in mathematics, reading, or science, as well as filling out the STQ. Questions draw on the idea of the "effort thermometer" introduced in PISA 2003 (Butler and Adams, 2007[55]). To complement questions examining students' perceptions of effort, a new school question examines administrators' communication with teachers and parents or guardians about PISA and their encouragement of students to do their best during the PISA test. Furthermore, a project on developing and validating measures of engagement is currently under way as a part of the PISA Research, Development and Innovation (RDI) programme. The project explores, validates, and compares different approaches to developing measures of engagement, including experimentation with both innovative methods (e.g. using evidence on engagement defined in the process of item design) and more 'traditional' methods (e.g. situated self-reports, ex-post questionnaire items, indicators of performance decline and engaged time). The results of the project will become available in 2023.

Figure 5.8. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.8. Constructs in PISA preparation and effort module

School culture and climate

School climate, safety, and student well-being are important antecedents of academic achievement (Kutsyuruba, Klinger and Hussain, 2015[56]). School climate encompasses shared norms and values, the quality of relationships, and the general atmosphere of a school (Loukas, 2007[57])and is often described as the quality and character of school life that sets the tone for all the learning and teaching done in the school environment. An academic focus—that is, a consensus about the mission of the school and the value of education, shared by school leaders, staff, and parents or guardians—affects the norms in student peer groups and facilitates learning (LEE and SHUTE, 2010[26]; Opdenakker and Van Damme, 2000[58]; Rumberger and Palardy, 2005[59]). Research shows that positive school climate contributes to immediate student achievement and endures for years (Hoy, Hannum and Tschannen-Moran, 1998[60]). A positive school climate is associated with student's motivation to learn (Eccles et al., 1993[61]) and has been shown to moderate the impact of socioeconomic context on academic success (Astor, Benbenishty and Estrada, 2009[62]). Lastly, the relationships that a student encounters at all levels in school (including students' views of the quality of teacher-student support and student-student support) also impact student achievement (e.g., (Jia et al., 2009[63]; Lee, 2021[64]).

Closely related to school climate is the safety of the learning environment. An orderly, safe, and supportive learning atmosphere maximizes attendance and the use of learning time. By contrast, a learning environment characterized by disrespect, unruliness, bullying, victimisation, crime, or violence can act as a barrier to students' learning and distract from the school's overall mission and educational goals. In the area of safety, schools without supportive norms, structures, and relationships are more likely to experience violence and victimization, which is often associated with reduced academic achievement (Astor, Guerra and Van Acker, 2010[65]).

Learning in 21st century schools in many countries differs from traditional settings in terms of the diversity of the student population—for instance, diversity in racial/ethnic and cultural backgrounds, as well as

diversity in individual student characteristics and diversity of thought. Experiences with diversity in the classroom may take the form of interpersonal interactions on campus, larger classroom discussions, or diversity-related coursework or workshops. In the United States context, researchers have found that several types of diversity experiences are associated with improvements in students' academic outcomes and cognitive development (e.g., development of critical thinking and problem-solving skills). Positive diversity experiences also play an important role in fostering students' social and emotional characteristics, such as tolerance, empathy, and curiosity (e.g. (Bowman, 2010[66]; Gurin et al., 2004[67]; Gurin et al., 2002[68]; Milem, Chang and Antonio, 2005[69]; Pettigrew and Tropp, 2006[70]).

General constructs in the PISA 2022 MS STQ include students' subjective perceptions as well as their values and beliefs about their in-school experiences. Measures are drawn from previously included constructs (e.g., sense of belonging, bullying experiences, school safety, and teacher support) as well as new constructs (e.g. quality of student-teacher relationships). Constructs in the SCQ include the school's efforts to promote school diversity/multi-cultural views, school climate-related factors hindering instruction, and disorder and delinquent behaviour at school. Questions in this module show some conceptual overlap with domain-specific questions in other modules (e.g., disciplinary climate in Module 16).

Figure 5.9. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.9. Constructs in school culture and climate module
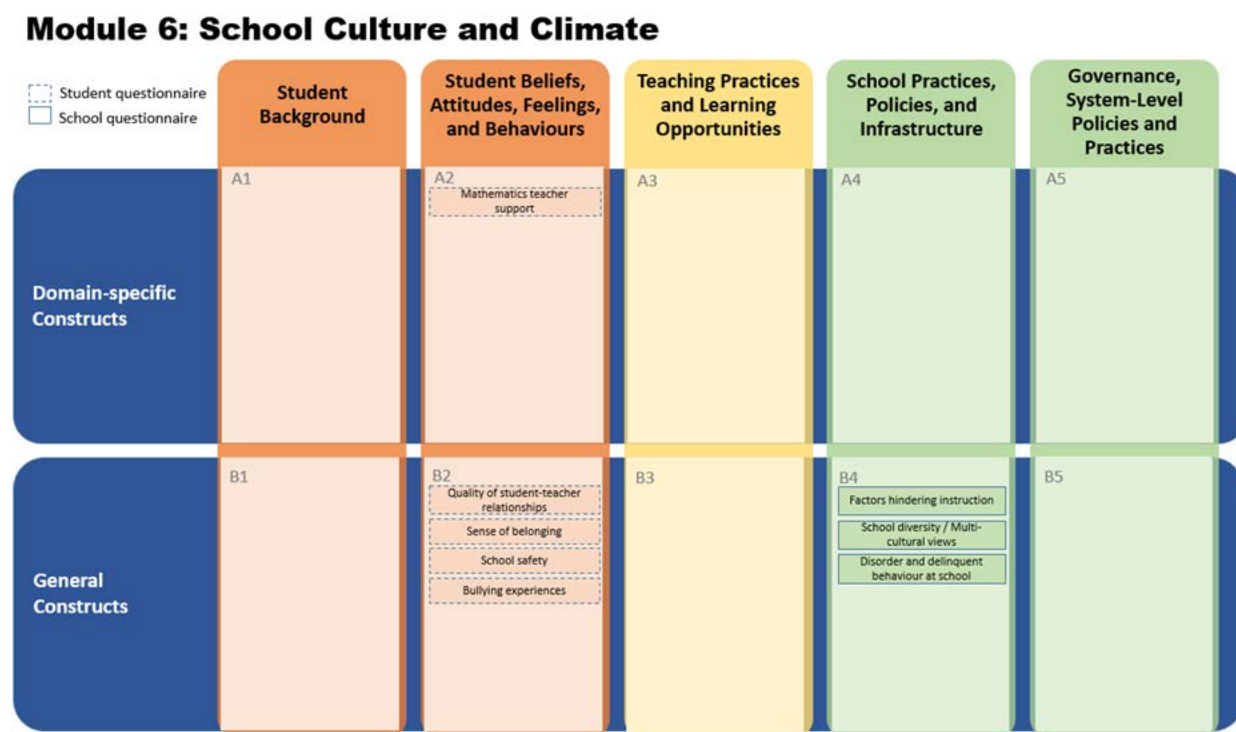
**Subject-specific beliefs, attitudes, feelings, and behaviours**

This module covers students' subjective perceptions as well as their values and beliefs, feelings and behaviours that are specific to mathematics, reading, and science. While a small set of key questions for each content-domain are included in the PISA 2022 MS, the focus of this module is on mathematics-related questions. Questions related to creative thinking are described in a separate module in this framework.

Questions related to all three domains include students' favourite subjects; whether students are motivated to achieve highly in mathematics, reading, and science; whether they think mathematics, reading, and science are easy for them; and the extent to which students think of skills in some subjects, as well as their general intelligence and creativity, as something malleable or largely robust to change (growth versus fixed mindset).

In addition, a combination of new mathematics-specific questions and questions retained from previous PISA cycles are recommended for this module. PISA 2012, for instance, assessed a number of mathematics-specific beliefs, attitudes, feelings, and behaviours. Four PISA 2012 scales (mathematics self-efficacy, mathematics anxiety, confidence in knowledge of mathematics concepts, and mathematics self-concept) were among the five constructs with consistently strongest correlational relationship with academic achievement in PISA 2012 (Lee and Stankov, 2018[71]). Based on these findings, measures for these constructs are also included in PISA 2022. Not all constructs, however, should be re-administrated without revisions and adjustments. On a trait level, mathematics self-efficacy, confidence, and self-concept are largely redundant (e.g. (Marsh et al., 2019[72])), a finding confirmed by PISA 2012 data when looking at joint relationships with achievement of these constructs. Mathematics self-efficacy and self-concept along with mathematics anxiety in PISA 2003 data formed a single second-order factor in the higher-order model to predict mathematics achievement (Lee and Stankov, 2013[73]). In both PISA 2012 and PISA 2003 data, self-efficacy showed better predictive validity for mathematics achievement than self-concept did (Lee, 2009[74]; Lee and Stankov, 2018[71]). For the PISA 2022 FT, the PISA 2012 self-efficacy scale was be retained and expanded by adding additional mathematics-reasoning related skills to the list of knowledge and skills. Self-efficacy was prioritized due to the concrete nature of the items that allow for clearer, more objective reporting than the agree/disagree type self-concept items used in PISA 2012. This difference in cross-cultural comparability of the two measures is reflected also in the finding that PISA 2012 self-efficacy showed consistently positive relationships with achievement both within and across countries, whereas relationships for self-concept were affected by the so-called "attitude-achievement-paradox" (Figure 5.25 in Section 5. of this framework). Rather than creating a second largely redundant scale focusing entirely on mathematics self-concept, this construct is operationalized for all three core PISA domains (mathematics, reading, and science) to allow for new insights based on potentially examining data as a profile across the three domains. Lastly, a new scale targeting students' invested effort and persistence in mathematics work (including homework) will provide actionable data for educators and policymakers that goes beyond the more subjective scales tapping into motivation in previous cycles.

Figure 5.10. below illustrates how all constructs in this module map onto the taxonomy.

Module 7: Subject-specific Beliefs, Attitudes, Feelings, & Behaviours
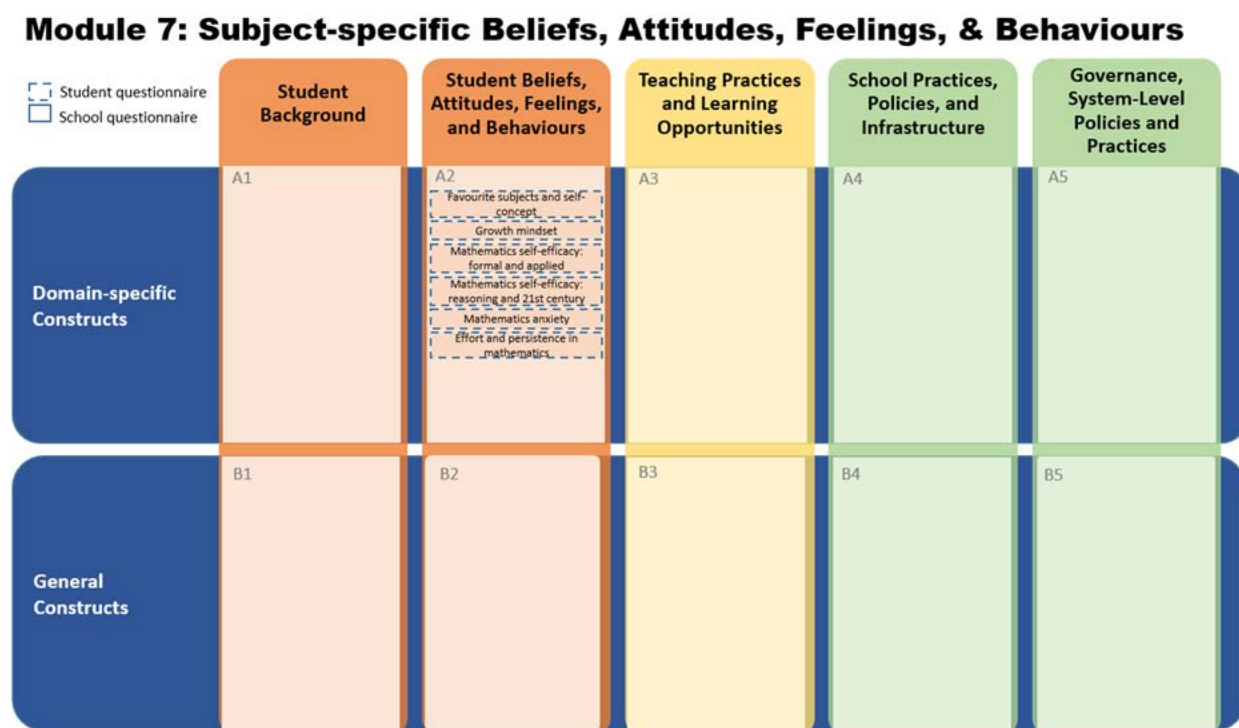
General social and emotional characteristics

Unlike the constructs listed above, constructs in this module are not primarily learning-related, but can be understood more broadly as characteristics indicative of student preparedness and social and emotional characteristics relevant to students' achievement in high school and throughout their lifetime. Two main framework approaches tend to be used to conceptualize social and emotional characteristics: one anchored to the personality psychology literature, which commonly refers to a "Big Five" taxonomy of personality traits (Abrahams et al., 2019[75]; Primi et al., 2021[76]); the other anchored to the social psychology literature, which focuses on cognitive constructs like motivations, beliefs, goals, interests, and values. PISA 2022 expands on these efforts by integrating the PISA framework with OECD's *Survey of Social and Emotional Skills* (SSES, (OECD, 2017[77])) to help policymakers and educators better link PISA data with other established frameworks and data sources. Based on the SSES framework, social and emotional characteristics can be defined as individual capacities that (a) are manifested in consistent patterns of thoughts, feelings, and behaviours, (b) can be developed through formal and informal learning experiences, and (c) influence important socioeconomic outcomes throughout individual's life. All general social and emotional characteristics measured in the PISA 2022 FT can be mapped onto the OECD SSES taxonomy (OECD, 2017[77]).

Task performance describes different aspects of students' conscientiousness and their striving for task performance, including setting high standards for themselves and working hard to meet them, fulfilling commitments and being reliable, being able to avoid distractions and focus attention on tasks, and persevering in the face of difficulty to complete tasks.

Emotional regulation covers different aspects of students' experienced range of emotions and their emotional regulation, including their ability to handle stress well, and regulate their temper, anger, and irritation in the face of frustrations.

Collaboration covers different aspects of students' approaches to collaboration, specifically their levels of agreeableness, including being kind and caring for others and valuing and investing in close relationships, building trust with others, as well as students' desire to value interconnections among people in general.

Open-mindedness covers different aspects of students' open-mindedness and openness to new experiences, including their desire to learn and approach situations with an inquisitive mindset, openness to different points of view and diversity, as well as enjoyment of generating novel ideas or visions.

Engaging with others covers different aspects of students' extraversion and their engagement with others, including their enjoyment of initiating and maintaining social connections, assertiveness in voicing their own views and exert social influence, as well as their tendency to approach daily life with energy, excitement, and spontaneity.

For the PISA 2022 MS, the following constructs are included that represent all five clusters of general social and emotional characteristics described above. Table 5.7. provides definitions for each construct that has partial item overlap between PISA 2022 and SSES.

### Table 5.7. Definitions for constructs with partial item overlap between PISA 2022 and SSES

Big Five Domain	Skill	Description	Behavioural Examples
Task Performance (Conscientiousness)	Self Control	Able to avoid distractions and focus attention on the current task in order to achieve personal goals	Doesnt rush into things, is cautious and risk averse. Opposite: is prone to impulsive shopping or binge drinking
	Persistence (Perseverance)	Persevering in tasks and activities until they get done	Finishes homework projects or work once started Opposite: Gives up easily when confronted with obstacles/distractions
Emotional Regulation (Emotional Stability)	Stress Resistance	Effectiveness in modulating anxiety and able to calmly solve problems (is relaxed, handles stress well)	Is relaxed most of the time, performs well in high-pressure situations Opposite: worries about things, difficulties sleeping
	Emotional Control	Effective strategies for regulating temper, anger and irritation in the face of frustrations	Controls emotions in situations of conflict. Opposite: gets upset easily; is moody.
Collaboration (Agreeableness)	Empathy	Kindness and caring for others and their well-being that leads to valuing and investing in close relationships	Consoles a friend who is upset, sympathises with the homeless. Opposites: Tends to disregard other persons feelings.
	Trust	Assuming that others generally have good intentions and forgiving those who have done wrong.	Lends things to people, avoids being harsh or judgmental. Opposite: is suspicious of peoples intentions.
	Cooperation	Living in harmony with others and valuing interconnectedness among all people.	Finds it easy to get along with people, respects decisions made by a group. Opposite: Has a sharp tongue, is not prone to compromises.
Open-mindedness (Openness to Experience)	Curiosity	Interest in ideas and love of learning, understanding and intellectual exploration; and inquisitive mindset	Likes to read books, to travel to new destinations. Opposite: dislikes change, is not interested in exploring new products
Engagement with Others (Extraversion)	Assertiveness	Able to confidently voice opinions, needs, and feelings, and exert social influence.	Takes charge in a class or team. Opposite: waits for others to lead the way, keeps quite when disagree with others.

Note: Descriptions and behavioural examples for each skill based on OECD Survey of Social and Emotional Skills (OECD, 2017b). This table only shows the constructs for which partial item overlap between PISA and SSES was achieved.

Each construct will be measured with a set of items that partly stem directly from the SSES and partly are unique to PISA. Perseverance and Self-control (both representing the *Task performance* cluster), Stress resistance and Emotional control (representing the *Emotional regulation* cluster), Curiosity and Perspective taking (representing the openness cluster), Cooperation, Empathy, and Trust (representing the *Collaboration* cluster), and Assertiveness (representing the *Engaging with others* cluster). Please note, in addition to these constructs, the student well-being questionnaire (SWBQ, not described in this framework) includes a range of constructs related to each of the Big Five factors, and the Creative Thinking-focused module included in the core STQ captures additional facets of openness.

Figure 5.11. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.11. Constructs in general social and emotional characteristics module

*Select items from this construct also appear in the OECD Social and Emotional Skills Study
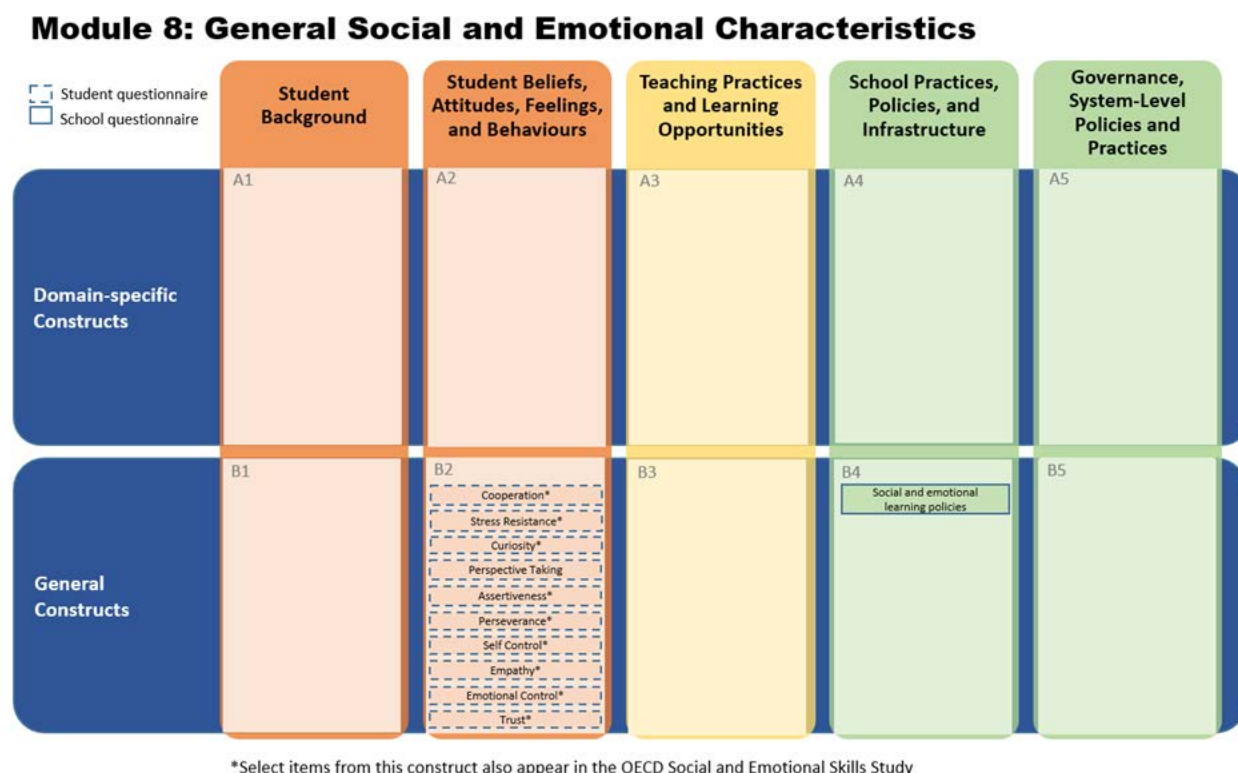
Health and well-being

PISA 2015 and 2018 started to include questions about health and well-being in the core STQ, and PISA 2018 offered an additional optional student well-being questionnaire (SWBQ) that gathered in-depth data on student well-being in participating countries. PISA 2022 carries these developments forward and includes, in addition to the optional SWBQ, a small module of health- and well-being related questions in the core STQ. Constructs for this module were selected to avoid any redundancies with the SWBQ and further prioritize well-being related questions that are important to capture student attitudes, feelings, and behaviour in all participating countries. These include students' overall life satisfaction, online activities, and potentially problematic online behaviours (e.g., extensive time spent on social networks and/or video games). The latter two constructs aim to understand the impact of online activities on students' health and well-being in light of the rapid growth of digital technology use across many aspects of daily life (e.g., socializing, communicating, and learning). Emerging research on adolescents and young adults—who are among the most active users of social media—suggests that digital technology use generally tends to have

small, negative effects on well-being, and effects differ depending on the type and frequency of activities (Dienlin and Johannes, 2020[78]; Keles, McCrae and Grealish, 2019[79]; Schønning et al., 2020[80]). Questions included in other modules (e.g., school culture and climate, general social and emotional characteristics, out-of-school experiences, physical exercise) will yield additional data that informs constructs that may be conceptualized also as part of health and well-being (e.g., activities before and after school, sense of belonging, bullying, and student-teacher relationships).

Figure 5.12. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.12. Constructs in health and well-being module

Module 9: Health and Well-Being
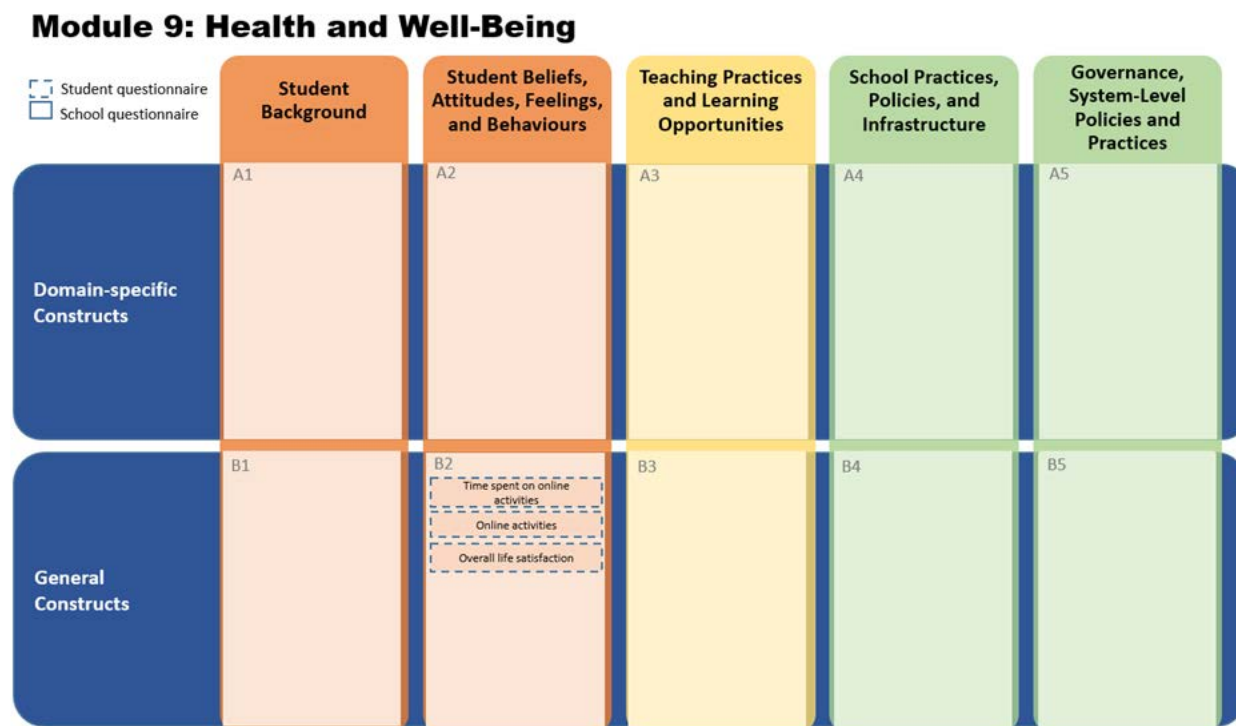
Out-of-school experiences

While classrooms serve as important settings for students' engagement in opportunities to learn, student engagement and learning also occur through formal and informal opportunities to learn outside of school. In the 2015 and 2018 questionnaire frameworks, students' out-of-school experiences focused on domain-specific indicators. The PISA 2022 framework takes a broader view on out-of-school experiences including both academic and non-academic experiences that may fall into several of the defined educational policy areas, including student attitudes, feelings, and behaviours and school practices, policies, and infrastructure.

How students spend their time outside of school, and the extent to which they engage in learning-related activities outside of school (e.g., tutoring, extracurricular activities, homework, mathematics-related activities), are important for understanding student achievement. Studies have shown that students' time use outside of school relates to mathematics achievement across several countries (Fuligni and Stevenson, 1995[81]), and engagement in extracurricular activities is associated with lower dropout rates for at-risk students, improved grade point averages, and higher educational aspirations (Broh, 2002[82]; Mahoney and Cairns, 1997[83]). Out-of-school activities can also provide important opportunities to learn, whereby students can apply subject-related content and skills that have been emphasized in class to novel

situations. This may be especially true for populations that have less exposure to formal education, as well as countries where structured out-of-school learning activities are prevalent (e.g. after-school tutoring to supplement and enhance in-school learning).

Domain-specific constructs in the STQ include students' participation in additional mathematics lessons outside of school and tutoring, and time spent on mathematics homework. Domain-specific constructs in the SCQ include administrators' reports of the school offering additional lessons in mathematics. General constructs in the STQ include students' activities before and after school (including physical activities or working for pay); general constructs in the SCQ include extracurricular activities offered by the school.

Figure 5.13. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.13. Constructs in Out-of-school experiences module

Module 10: Out-of-School Experiences
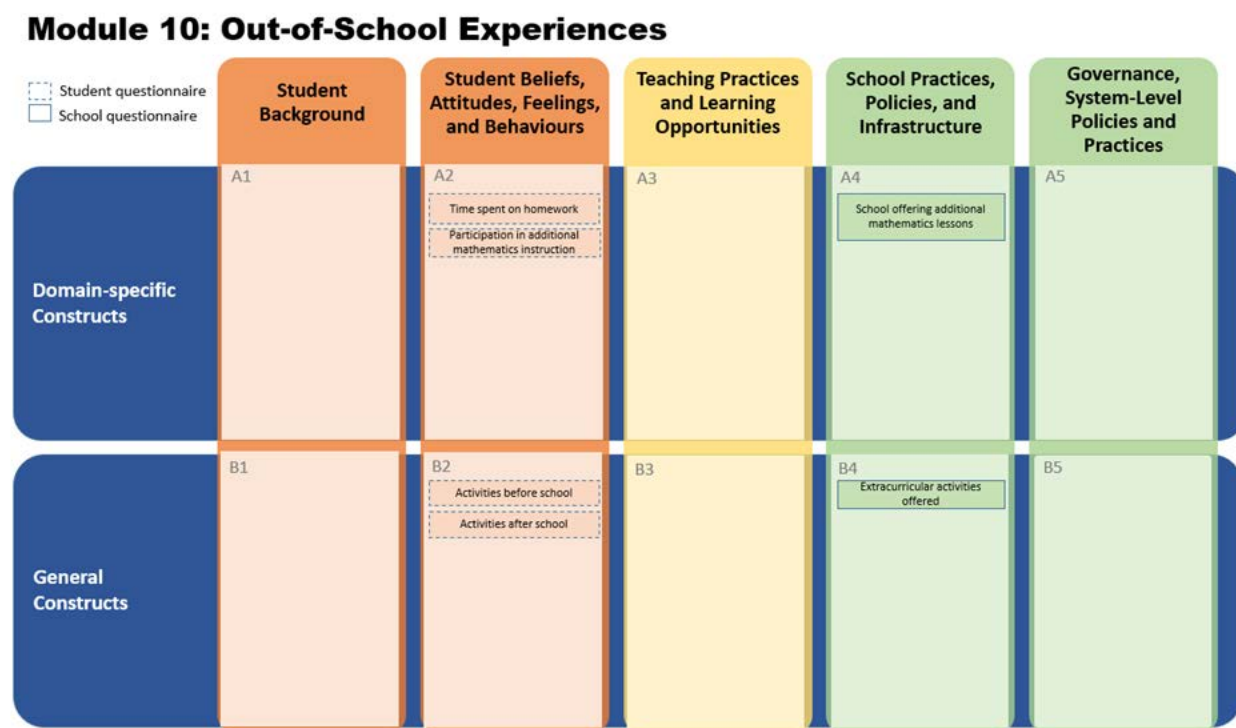
School type and infrastructure

This module examines aggregate school-level characteristics of the students' learning environment (e.g., location, type, and size of the school) and school risk factors that may hinder student learning and achievement as they relate to the physical set-up of the school, such as deficiencies in school resources and infrastructure. The quality of a school's infrastructure, and the quality and accessibility of digital educational resources (e.g., computers and other digital technology, internet access) may facilitate or hinder the learning environment's positive impact, and in turn, influence achievement.

Conceptually, this module overlaps with other modules measuring the overall characteristics of the school and school population, including those capturing school culture and climate (Module 6), organisation of student learning at school (Module 14); assessment, evaluation, and accountability (Module 18); and school autonomy (Module 13). General constructs in the SCQ include school size (teachers, students, and non-teaching staff), school type and the type of organisation running the school, school location, availability of digital technology, and lack of physical and digital infrastructures.

Figure 5.14. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.14. Constructs in school type and infrastructure module

Module 11: School Type and Infrastructure
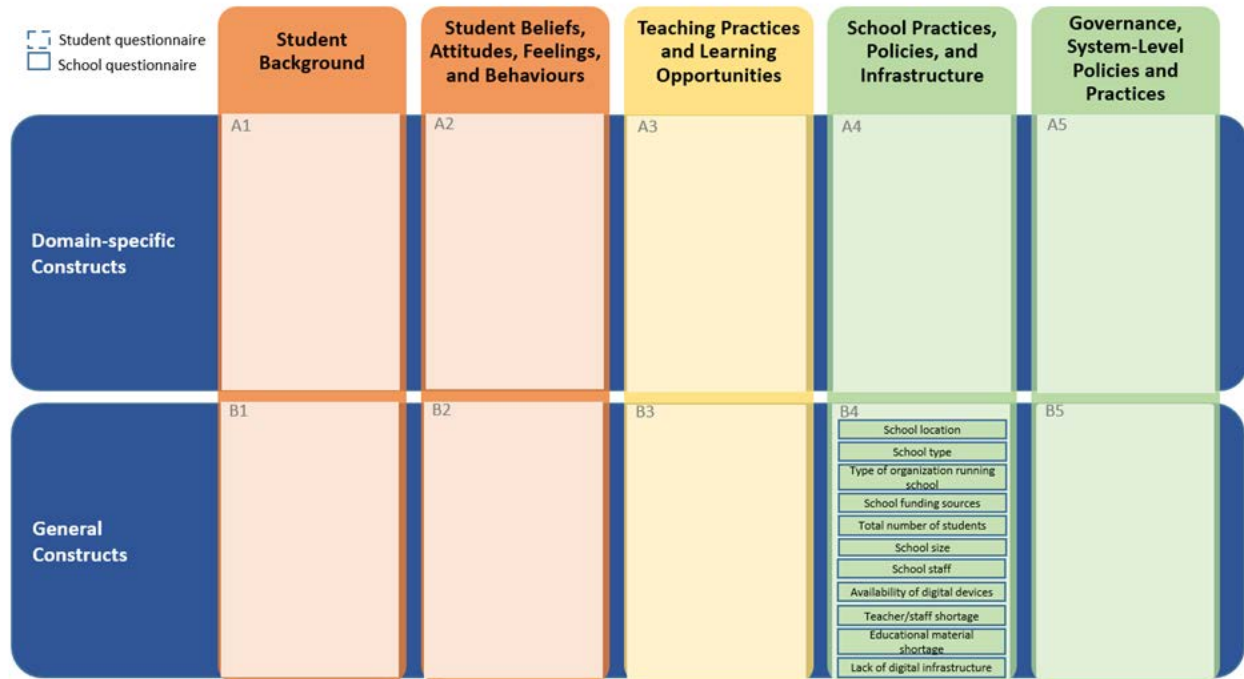
Selection and enrolment

School principals and administrators play a key role in school management and policy, as they are often seen as the primary agents of change to improve student achievement in their schools. They can shape teachers' professional development, define the school's overall mission and educational goals, ensure that instructional practices and policies within and across subjects are directed towards achieving these goals, suggest modifications to improve teaching practices, and help solve problems that may arise within the classroom or among teachers.

The way in which students are channelled into educational pathways, schools, tracks, or courses (also known as stratification, streaming, or tracking) is a core issue of educational governance and is an important aspect of school organisation and policy. For instance, highly selective schools provide a learning environment that may differ from the environment offered by schools that are more comprehensive. Some longitudinal studies have demonstrated grade retention harms individual careers and outcomes (e.g., (Griffith et al., 2010[84]; Ou and Reynolds, 2010[85]), as well as student behaviour and well-being (e.g., (Crothers et al., 2010[86])), while other research finds positive effects (Marsh et al., 2017[87]). Greene and Winters (2009[88]) showed that once a test-based retention policy has been installed, those who were exempted from the policy did worse. Additionally, Babcock and Bedard (2011[89]) showed that a large number of students being retained could have a positive effect on the cohort (i.e., all students, including those who are promoted). Kloosterman and De Graaf (2010[90]) argued that in highly tracked systems, such as in some European countries, grade repetition might serve as a preferred alternative to moving into a lower track. The authors found evidence that this strategy is preferred for students with higher SES. Thus, changing grade repetition policies might be a viable option regarding low-cost interventions (Binder, 2009[91]).

General constructs in the SCQ include the school's selection competition, academic selectivity, and student transfer policies.

Figure 5.15. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.15. Constructs in selection and enrolment module


Module 12: Selection and Enrolment
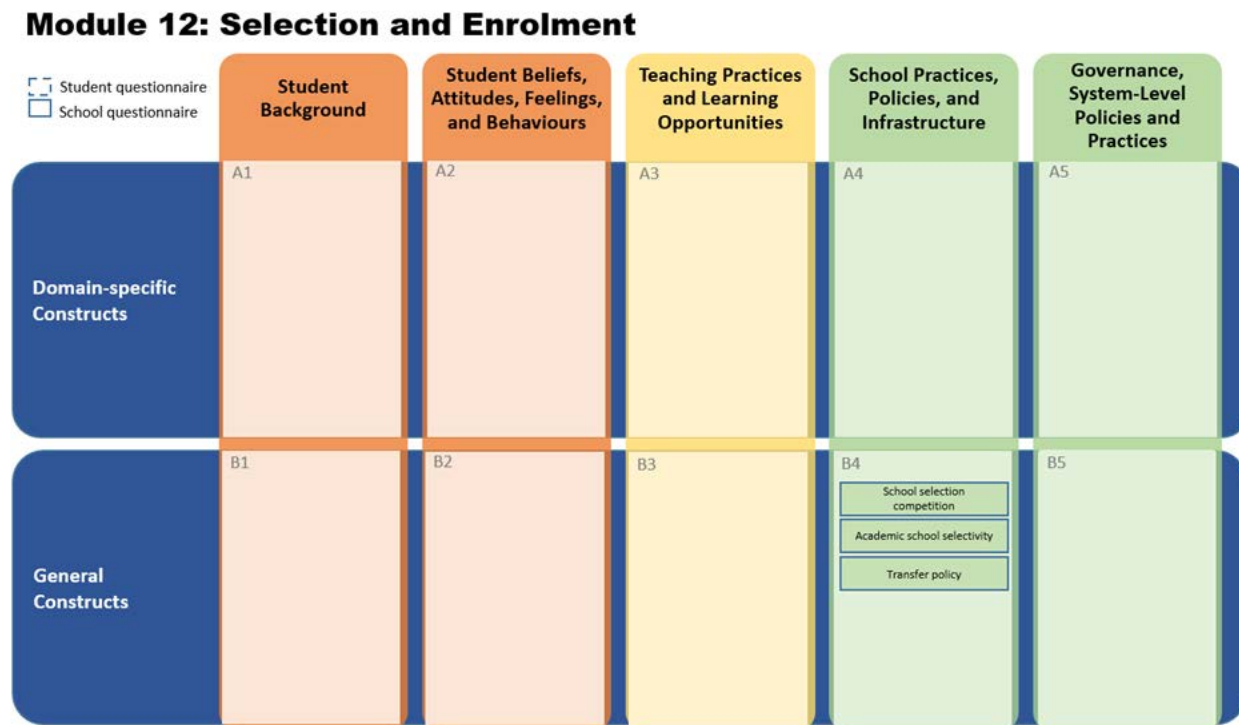
School autonomy

Education systems have been classified by the amount of control or local autonomy that is given to schools (i.e., the school board, staff, and school leaders) versus governing bodies at the local, regional, or national level when decisions on admission, curriculum, allocation of resources, and personnel must be made. These indicators have been previously included in the PISA 2012 SCQ and are revisited in 2022. General constructs in the SCQ include administrators' reports of the primary responsibility for school decision making and the role of the school management team in providing instructional leadership to teachers.

Figure 5.16. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.16. Constructs in school autonomy module
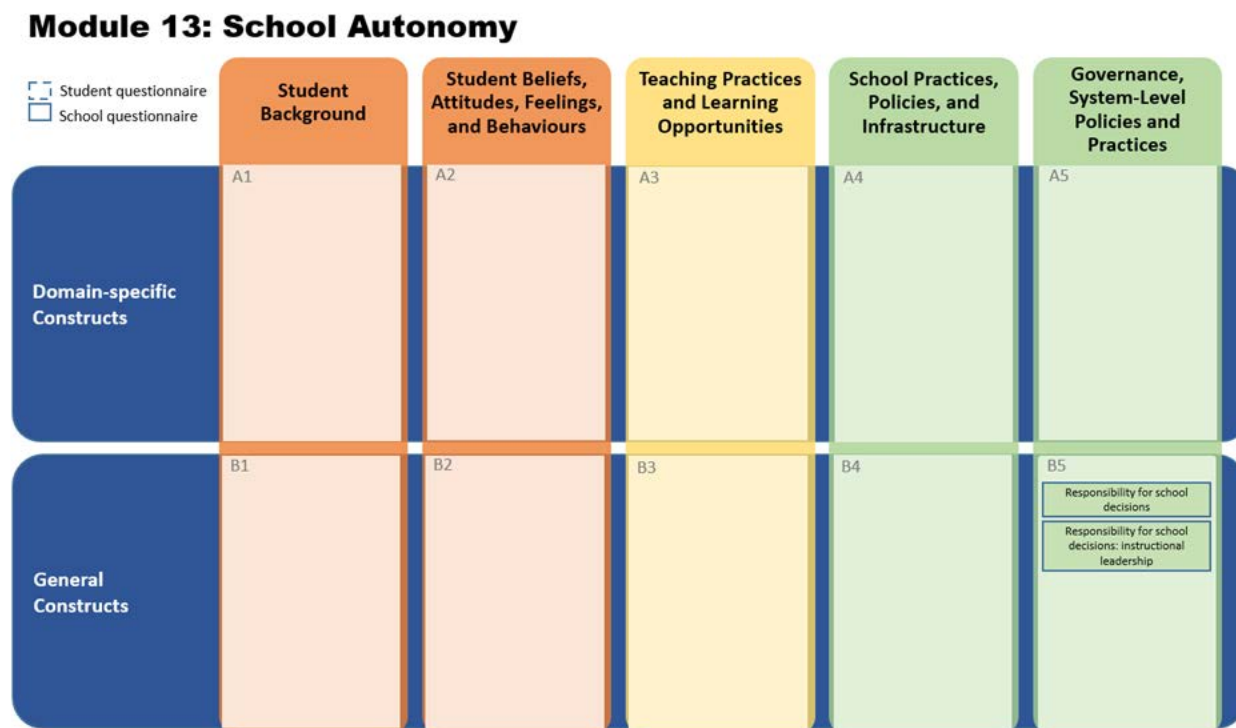
Organisation of student learning at school

Large portions of students' educational experiences tend to occur at school in the classroom environment. During time spent in the classroom, students are exposed to subject content, curriculum materials, instructional strategies, skills, and a diversity of backgrounds and perspectives contributing to overall climate. Learning time and the intended curriculum content in school have been found to be closely related to student outcomes (e.g., (Abedi, 2006[15]; Cogan, Schmidt and Guo, 2018[92]; Scherff and Piazza, 2008[93]; Schmidt, 2009[17])). Overall students' learning time and achievement are correlated as the time allowed for learning constrains students' opportunities to learn, though there are large differences within countries, across countries, and among different groups of students and schools (Ghuman and Lloyd, 2010[94]; OECD, 2011[95]). A generally positive relationship has been replicated in international comparative research (e.g. (OECD, 2011[95]; Martin, Mullis and Foy, 2008[96]; Schmidt, 2001[16]; Schmidt and Burroughs, 2016[97]; Schmidt et al., 2015[98]).

Related to learning time is the way intended learning content is designed, structured, and communicated during that time in school. Understanding how a school curriculum functions requires a consideration of how it is organised and how students gain access to it. For example, a school's curriculum can be understood by examining what coursework is required and optional; whether students are tracked or grouped by achievement; and what standards are used to develop subject content. Curriculum may vary largely across tracks, grades, schools, and countries (Schmidt, 2001[16]; Martin, Mullis and Foy, 2008[96]). Overall, there may be variations between the curriculum designed at the system level, the curriculum communicated by the teacher or in the textbook, and the curriculum as understood by students and their parents.

A domain-specific construct included in the STQ captures students' mathematics class periods per week. Domain-specific constructs in the SCQ capture administrators' reports of the average time in a class period, the average number of students in these classes, percentages of students below/above the pass mark,

student ability grouping in mathematics, the school offering study help, tracking policies, digital device policies, and selection of courses. This module complements Modules 15 (*Exposure to Mathematics Content*) and Module 16 (*Mathematics Teacher Behaviours*) in mapping out a broad view of students' OTL at school.

Figure 5.17. below illustrates how all proposed constructs in this module map on the taxonomy.

Figure 5.17. Constructs in organisation of student learning at school module

Module 14: Organisation of Student Learning at School
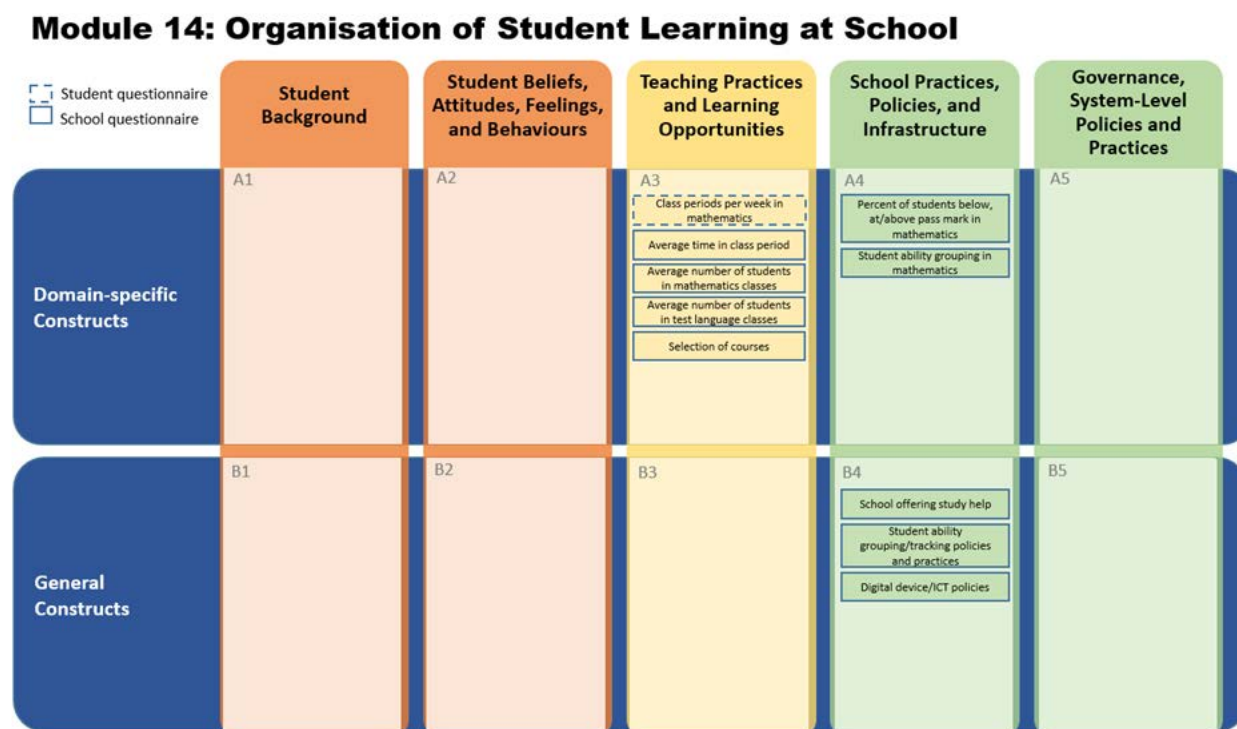
Exposure to mathematics content

This module focuses on one key aspect of the broader OTL constructs, specifically students' exposure to relevant mathematics content. In conjunction with the modules on Organisation of Student Learning at School (Module 14) and Mathematics Teacher Behaviours (Module 16), this module focuses on the first three types of OTL-related variables described by Stevens (Stevens, 1993[99]):

- *Content coverage variables* that measure whether students learn the content covered in the curriculum for a particular grade level or subject;
- *Content exposure variables* that consider the time allowed for and devoted to instruction and the depth of teaching provided;
- *Content emphasis variables* that consider which topics within the curriculum are selected for emphasis and which students are selected to receive instruction emphasizing either lower-order skills (i.e., rote memorization) or higher-order skills (i.e., critical problem solving); and
- *Quality of instructional delivery variables* that measure how classroom teaching practices (i.e., presentation of lessons) affect students' academic performance.

PISA 2012 aimed to capture domain-specific (mathematics) OTL profiles in the STQ through the presentation of tasks reflecting mathematical abilities and content categories outlined in the PISA

mathematics framework. Students were asked to judge whether and how often they had seen similar tasks in their mathematics lessons; thus, OTL measures in PISA 2012 (experience with pure and applied math tasks, experience with problem types in mathematics, and familiarity with mathematics concepts) were mainly concerned with aspects of content coverage and exposure.

One specific area for new development in PISA 2022 was around students' OTL with regard to mathematics reasoning skills. The goal of PISA 2022 is to measure in-school OTL (i.e., content coverage and exposure) at the school and country level in a way that allows for a clearer differentiation between types of mathematics problems and mathematics content—for instance, country-level differences in opportunities to learn formal mathematical modelling or applied mathematics problems. Domain-specific constructs in the STQ include students' exposure to different types of mathematics content (formal and applied mathematics tasks), as well as their exposure to mathematics reasoning and 21st century skills related to mathematics and their subjective familiarity with mathematics concepts. A domain-specific construct pertaining to the standardisation of the school's mathematics curriculum is also included in the SCQ.

Figure 5.18. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.18. Constructs in exposure to mathematics content module
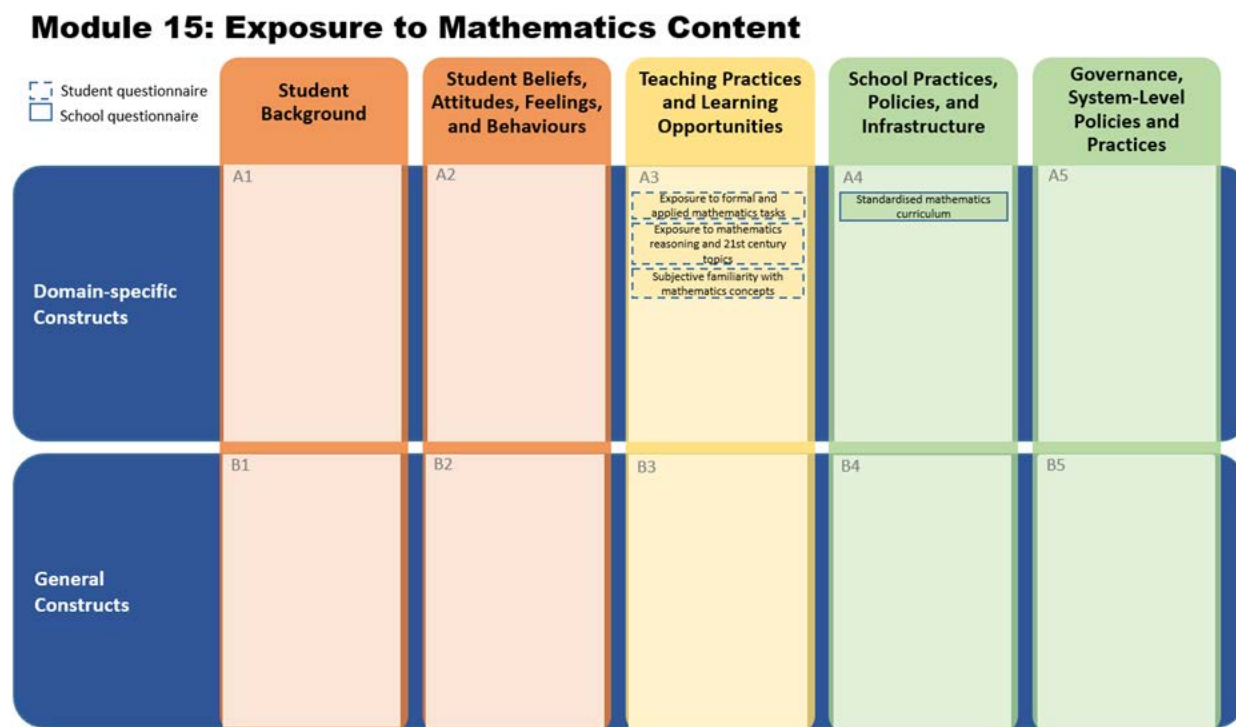
Mathematics teacher behaviours

How student learning is organised (Module 14) and what content is being taught (Module 15) are conceptually distinct from constructs that capture teaching practices and behaviours (instructional quality), in that teaching practices and behaviours can serve as vehicles through which different levels of content coverage and exposure may occur. What teachers do has the strongest direct school-based influence on student learning outcomes (Hattie, 2009[100]). Effective instruction is rooted in part in the repertoire of practices through which teachers facilitate students' thinking and understanding of subject content and concepts. Previous research has shown that proximal variables, such as classroom characteristics and

teaching and learning practices, are more closely associated with student achievement than distal variables measured at the school- and system-level (e.g. (Hattie, 2009[100]; Slavin and Lake, 2008[101]; Wang, Haertel and Walberg, 1993[102]).

 Though understood differently across the field, there is general agreement that teachers' instructional practices, or instructional quality, is a multidimensional concept (e.g., (Fauth et al., 2014[103]; Kane and Cantrell, 2010[104])). The 2018 OECD Teaching and Learning International Survey (TALIS) framework identifies the following dimensions of teaching practices as having an influence on student achievement:

- *Classroom management,* or the actions taken by teachers to ensure order and effective use of time during lessons (van Tartwijk and Hammerness, 2011[105]);

- *Teacher support,* such as providing extra help when needed, listening to and respecting students' ideas and questions, caring about and encouraging students, and providing emotional support to them ((Klieme, Pauli and Reusser, 2009[106]; Lee, 2021[64]);

- *Clarity of instruction,* that is, teachers' clear and comprehensive instruction and learning goals, connection of old and new topics, and summarization of lessons ((Hospel and Galand, 2016[107]; Kane and Cantrell, 2010[104]; Seidel, Rimmele and Prenzel, 2005[108]);

- *Cognitive activation*, or the use of instructional activities involving evaluation, integration, and knowledge application in the context of problem solving, through which students engage in knowledge construction and higher order thinking (Lipowsky et al., 2009[109]); and

- *Instructional assessment and feedback*, more specifically, the provision of constructive feedback through formative and summative assessment (Hattie and Timperley, 2007[110]; Kyriakides and Creemers, 2008[111]; Scheerens, 2016[112]) or homework (Cooper, Robinson and Patall, 2006[113]).

Previous TALIS main study results from 2008 found that in 23 countries, participation in professional development and teaching high-ability classes raised the frequency of teachers implementing practices to improve clarity of instruction, teacher support, and cognitive activation (via enhanced activities). It is important to note that while effective pedagogical practices overlap across subjects and student populations, some practices may vary by specific subjects and populations. For instance, TALIS data indicate that mathematics and science teachers reported less student-oriented instructional support and less frequent use of enhanced activities compared to teachers who taught other subjects (OECD, 2009[114]).

While TALIS has focused on measurement of general teaching practices, PISA 2022 complements these efforts by measuring closely aligned constructs that are domain-specific (i.e., mathematics focused), as has been done in previous cycles.

- *Disciplinary climate in mathematics* examines disciplinary issues that hinder mathematics learning in the classroom, complementing the TALIS dimension of *classroom management*;

- *Mathematics teacher support* covered in Module 6 (School Culture and Climate) is conceptually aligned with the dimension of *teacher support*;

- *Cognitive activation in mathematics* is conceptually similar to the dimension of *cognitive activation*, however, PISA is focused specifically on the extent to which teachers encourage mathematical thinking and reasoning skills as highlighted in the PISA 2022 mathematics framework; and

- *Use of mathematics assessments* is conceptually aligned to the dimension of *instructional assessment and feedback*, providing additional information about instructional assessment and feedback in mathematics.

Aspects of classroom disciplinary climate, teacher support, cognitive activation, and teacher behaviour (student-oriented) were measured in PISA 2012. Previous research indicates that several of the dimensions defined above correlate with students' mathematics outcomes (e.g., (Lee, 2021[64])). For instance, the international PISA 2003 report found that disciplinary climate in the mathematics classroom

was strongly associated with mathematical literacy, while other variables (e.g., class size, mathematical activities offered at the school level, avoidance of ability grouping) had no substantial relationship once socioeconomic status was accounted for (OECD, 2004[115]). The PISA 2012 data also showed strong predictive validity of disciplinary climate for students' mathematics achievement (Lee and Stankov, 2018[71]). Additionally, teacher support has been found to be positively linked to students' interest in mathematics after accounting for socioeconomic status (Vieluf et al., 2012[116]). Finally, cognitive activation in the form of providing learners opportunities to develop and practice mathematical competencies have been broadly discussed in mathematics education (e.g. (Blum and Leiss, 2007[117])).

Addressing teacher and teaching-related factors in PISA is a challenge, because sampling is by age rather than by grade or class. Nevertheless, aggregated student data and the optional teacher questionnaire can be utilized to describe several aspects of teacher background and practices, and the learning environment offered in classrooms.

Figure 5.19. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.19. Constructs in mathematics teacher behaviours module
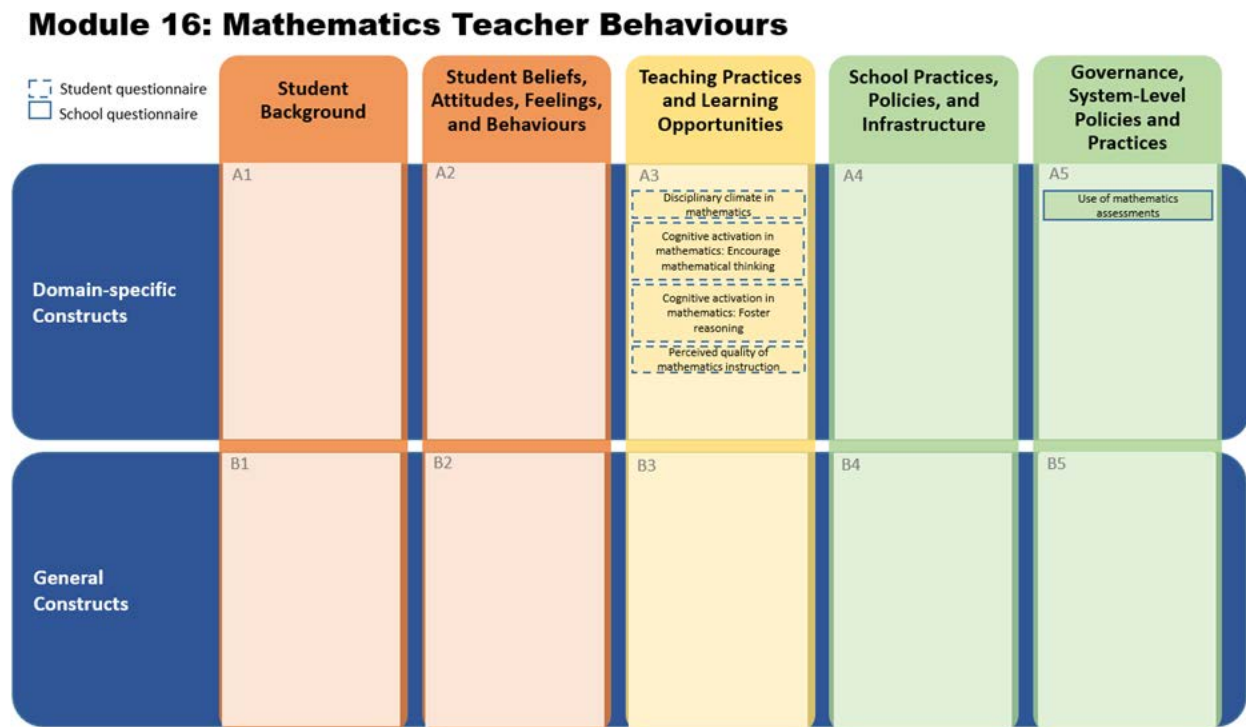
Teacher qualification, training, and professional development

OECD's annual International Summit on the Teaching Profession (ISTP; (Schleicher, 2014[118])) has exemplified the continuously growing focus on teacher-related policies for improving the quality of teachers, teaching, and learning. In addition to teacher's professional behaviour (e.g., interactions with students in the classroom and with their parents or guardians), the composition of the teaching force in terms of age and educational level, their initial education and qualifications, their individual beliefs and competencies, as well as professional practices on the school level (e.g., professional development, interactions with parents) have been topics of educational policy discussions.

A number of studies have demonstrated a clear influence of teacher-related factors on student learning and outcomes (e.g., (Schmidt et al., 2016[119])). Several studies and reviews show positive relationships between teachers' initial education and their teaching effectiveness (e.g., (Boyd et al., 2009[120]; Darling-Hammond et al., 2005[121])). Research has shown that when teachers have opportunities to expand and develop their teaching practices and their knowledge of instructional approaches, they are more likely to provide a broader range of learning opportunities for students and be more effective in improving students' learning outcomes (Harris, 2002[122]; Rankin-Erickson and Pressley, 2000[123]).

General constructs in the SCQ include administrators' reports of teacher qualifications, and in-house professional development opportunities. Domain-specific constructs in the SCQ include mathematics teacher qualifications and mathematics in-house professional development opportunities.

Figure 5.20. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.20. Constructs in teacher qualification, training, and professional development module
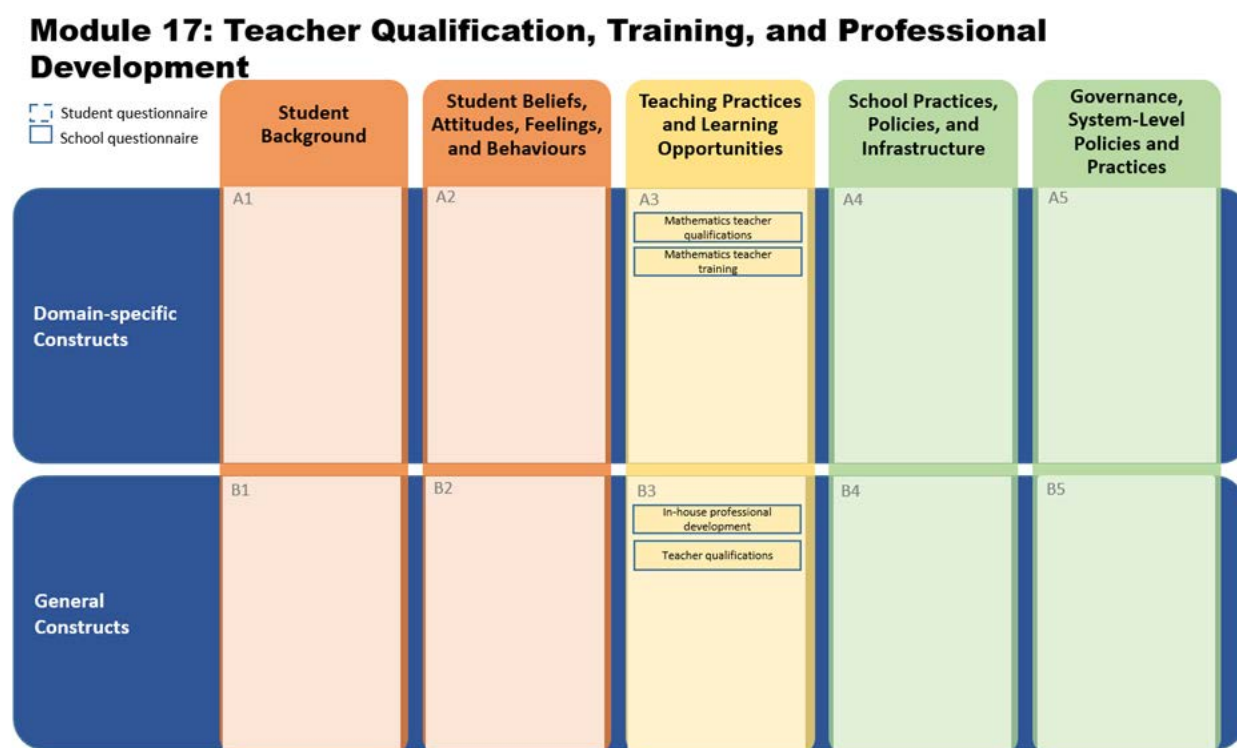
Assessment, evaluation and accountability

Assessing students and evaluating schools are common practices in most countries (Ozga, 2012[124]). Since the 1980s policy instruments, such as performance standards, standard-based assessment, annual reports on student progress, and school inspectorates, have been promoted and implemented across continents. Reporting and sharing data from assessments and evaluations with different stakeholders provides multiple opportunities for monitoring, feedback, and improvement. In recent years, there has been a growing interest in the use of assessment and evaluation results through feedback to students, parents or guardians, teachers, and schools as one of the most powerful tools for quality management and improvement (OECD, 2010, p. 76[125]). In addition, formative assessment, also known as assessment for learning, has been one of the dominant movements (Baird et al., 2014[126]; Black, 2015[127]; Hattie, 2009[100]). Accountability systems based on these instruments are increasingly common in OECD countries (Rosenkvist, 2010[128]; Scheerens, 2002, p. 36[129]).

Prior PISA cycles have covered aspects of assessment, evaluation, and accountability in the SCQ by identifying a variety of purposes for the assessment of students. School administrators have been asked whether they use test results to make comparisons with other schools at the district or national level, as well as to improve teacher instruction (e.g., by asking students for written feedback on lessons, teachers, or resources). However, extant research indicates that there are very few low-income countries that have a national assessment system in place that can track learning in a standardized manner to provide feedback into education policies and programs (Birdsall, Bruns and Madan, 2016[130])

The evaluation of schools is used as a means of assuring transparency; making judgments and decisions about systems, programs, educational resources and processes; and guiding overall school development (Faubert, 2009[131]), and evaluation criteria may be defined and applied from the viewpoints of different stakeholders (Sanders and Davidson, 2003[132]). Evaluation can either be external (i.e., the process is controlled and headed by an external body and the school does not define the areas that are judged) or internal (i.e., the process is controlled by the school itself and the school defines the areas that are judged) (Berkenmeyer and Müller, 2010[133]). The evaluation may be conducted by members of the school, or by persons/institutions commissioned by the school. Different evaluation practices generally coexist and benefit from each other (Ryan, Chandler and Samuels, 2007[134]). For instance, external evaluation can expand the scope of internal evaluation and also validate results and implement standard or goals. Additionally, internal evaluation can improve the interpretation and increase the utilization of external evaluation results. However, improvement of schools seems to be more likely when an internal evaluation is applied, compared to external evaluation. Thus, processes and outcomes of evaluation may differ between internal and external evaluation. Moreover, country and school-specific context factors may influence the implementation of evaluations as well as the conclusions and impact for schools. In many countries, individual evaluation of teachers and principals, separate from school-wide evaluation, is also common (Faubert, 2009[131]; Santiago and Benavides, 2009[135]). One study looked at 12 different school management programs in low- and middle-income countries and found that interventions from these management systems did not improve factors such as completion rates and did not have any significant effect on learning outcomes. However, in instances where the program included creating school improvement plans, decentralizing financial-decision making, and generating annual report cards on school performance, there was an improvement in learning outcomes (Snilstveit, 2016[136])

In the past several years, a number of countries have implemented national standards to assess students' learning outcomes. Together with formative assessment practices, summative assessment systems influence the way teachers teach and students learn. In particular, formative assessment practices can enhance students' achievement (Black and Wiliam, 1998[137]). However, there is a large variation in the implementation of formative assessment practices, which has also been reported in recent studies in the United States, Canada, Sweden, Scotland, Singapore, and Norway, among others (DeLuca et al., 2015[138]; Hopfenbeck, Flórez Petour and Tolo, 2015[139]; Jonsson, Lundahl and Holmgren, 2015[140]; Hayward, 2015[141]; Ratnam-Lim and Tan, 2015[142]; Wylie and Lyon, 2015[143]).

Domain-specific constructs in the SCQ include administrators' reports of the use of mathematics achievement data in accountability systems. General constructs in the SCQ include administrators' reports of monitoring teacher practices, feedback to teachers, assessment use in the school, and school evaluation.

Figure 5.21. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.21. Constructs in assessment, evaluation, and accountability module
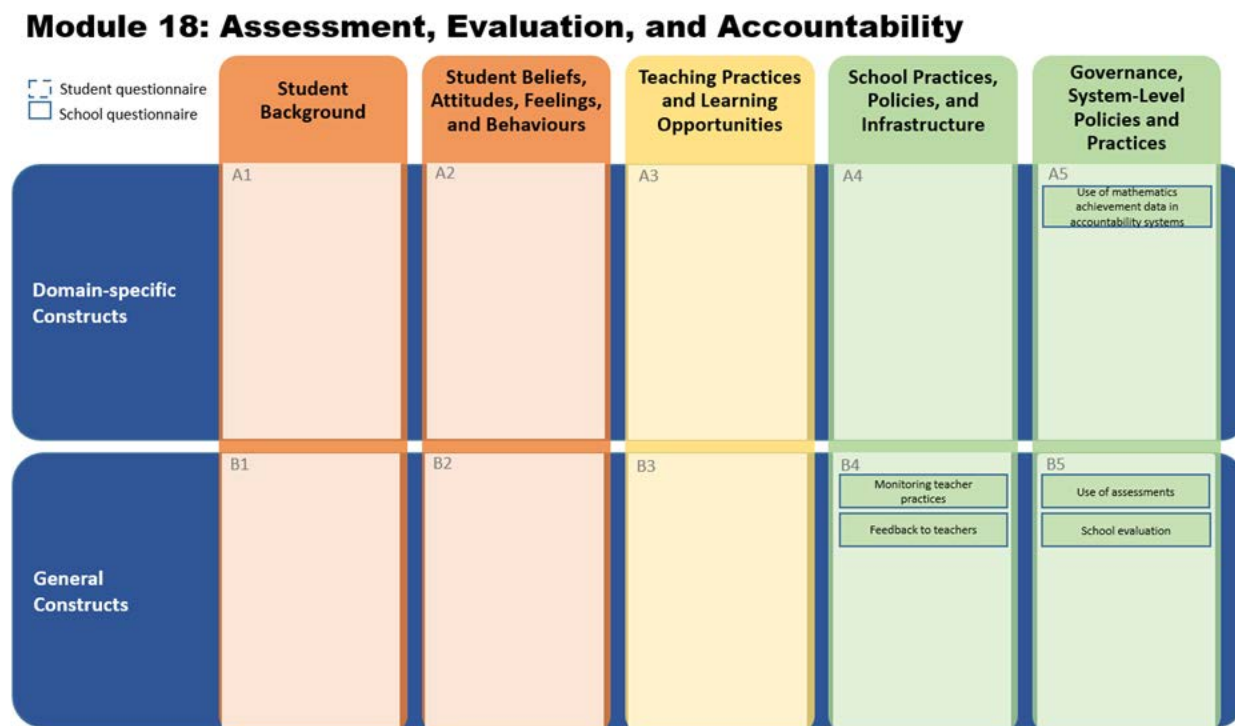
**Parental/guardian involvement and support**

Parents and guardians are an important audience as well as powerful stakeholders in education, and open communication and collaboration between school leadership and students' parents or guardians are essential to student success. Parental/guardian involvement in education has been conceptualized as parents' or guardians' interactions with schools and their children to encourage academic success (Hill and Tyson, 2009[144]). This involvement is multidimensional and includes school-based involvement (e.g., attending parent-teacher meetings, volunteering at school, or participating in school governance), home-based involvement (e.g., assisting with homework; participating in intellectual enrichment activities not directly related to school but that help develop children's cognitive and metacognitive processes), and academic socialization (i.e., parents' or guardians' educational goals and expectations for their children in general and in specific subjects, and the ways in which these goals and expectations are communicated) (Epstein, 2001[145]; Hill and Tyson, 2009[144]; Kim and Hill, 2015[146]; Murayama et al., 2016[147]). Parental/guardian involvement may also vary by whether the participation is initiated by parents or guardians, students, teachers, or schools. For example, analyses of PISA 2012 data from seven countries have found that school principals' reports of parent-initiated involvement related positively to between-school differences in student achievement, while within schools, parent reports of teacher-initiated involvement related negatively to student achievement (Sebastian, Moon and Cunningham, 2017[148]).

In addition to parents' or guardians' involvement in school activities, the support provided in the family plays an important role in fostering student learning and helping children and adolescents develop confidence, stress resistance, and other social and emotional characteristics important for academic and non-academic success. Several meta-analyses show a positive relationship between parental involvement in education and student achievement (Fan and Chen, 2001[149]; Hill and Tyson, 2009[144]; Jeynes, 2007[150]; Kim and Hill, 2015[146]), and parents' academic socialization of their children was found to have a strong positive relationship with achievement (Fan and Chen, 2001[149]; Hill and Tyson, 2009[144]; Kim and Hill,

2015[146]). This correlation generally held across race and ethnicity and when accounting for socioeconomic differences within the United States (Jeynes, 2007[150]; Kim and Hill, 2015[146]).

Figure 5.22. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.22. Constructs in parental/guardian involvement and support module
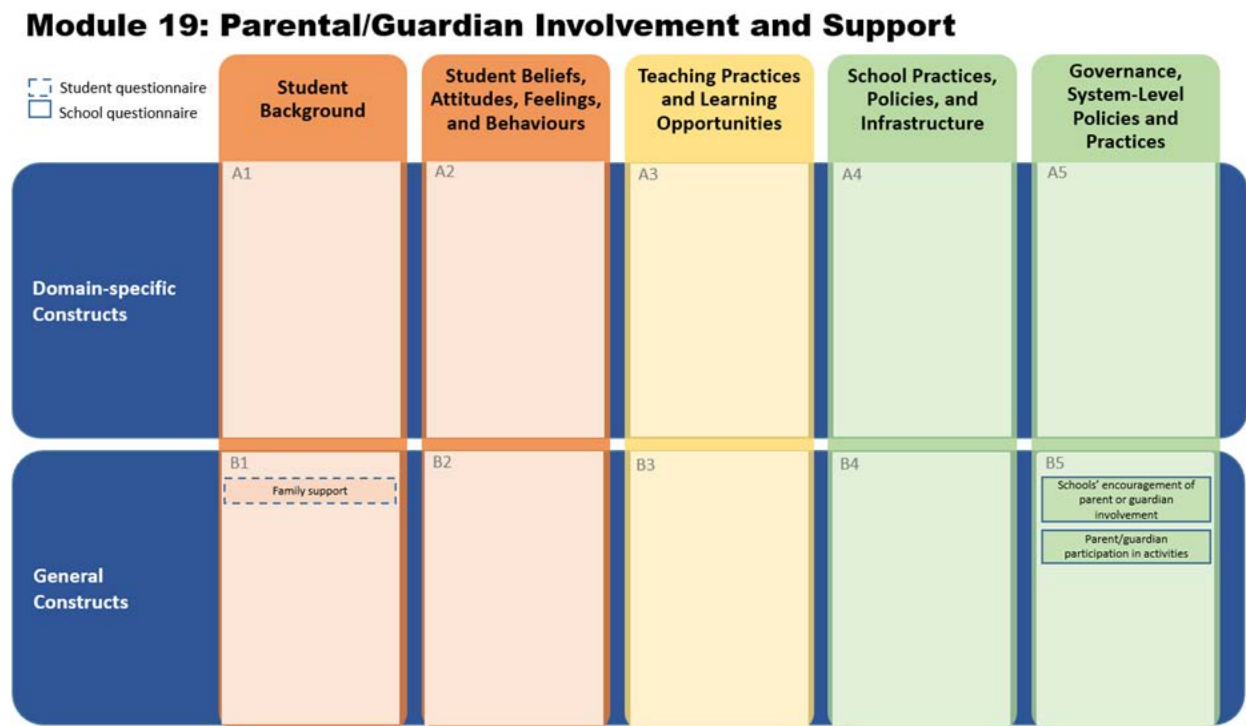
Creative thinking

For PISA 2022, creative thinking is defined as "the competence to engage productively in the generation, evaluation and improvement of ideas, that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination" (OECD, 2019[151]). Creative thinking is a necessary competence that can benefit student learning by supporting the interpretation of experiences, actions and events in novel and personally meaningful ways; by facilitating understanding, even in the context of predetermined learning goals; and by promoting the acquisition of content knowledge through approaches that encourage exploration and discovery rather than rote learning and automation (Beghetto, Baer and Kaufman, 2015[152]; Beghetto and Kaufman, 2007[153]; Beghetto and Plucker, 2006[154]). It can help students adapt and contribute to a rapidly changing world that requires flexibility and 21st century skills that go beyond core literacy and numeracy (OECD, 2019[151]).

According to the creative thinking framework for PISA 2022, this competence is fostered by a combination of internal resources, or "individual enablers", and influenced by features of the students' social environment, or "social enablers" (OECD, 2019[151]). Schools are settings in which students' manifestations of creative thinking can be observed. Individual enablers of creative thinking include students' cognitive skills, domain readiness (i.e., domain-specific knowledge and experience), openness to experience and intellect, goal orientation and creative self-beliefs (i.e., willingness to persist towards one's goals in the face of difficulty and beliefs about one's own ability to be creative), collaborative engagement (i.e., willingness to work with others and build upon others' ideas), and task motivation. Social enablers of creative thinking include cultural norms and expectations, educational approaches (e.g., the outcomes an

education system values for its students and the content it prioritises in the curriculum), and classroom climate (e.g., teacher and school practices that encourage or stifle creative thinking). The PISA contextual questionnaires aim to collect information on those enablers and drivers that are not directly assessed in the cognitive test of creative thinking.

Constructs in the STQ focus on Creative self-efficacy, Creative school and out-of-school activities, Creative environments, and various facets of openness (e.g., openness to intellect, openness to arts, ingenuity). Constructs in the SCQ focus on school administrators' beliefs about creativity, creative school environment, creative activities offered by the school, and the school's culture or climate of openness. Creative thinking constructs are also included in the optional teacher and parent questionnaires to gather additional information on the beliefs about creativity and creative social environments promoted by these sources.

Figure 5.23. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.23. Constructs in creative thinking module
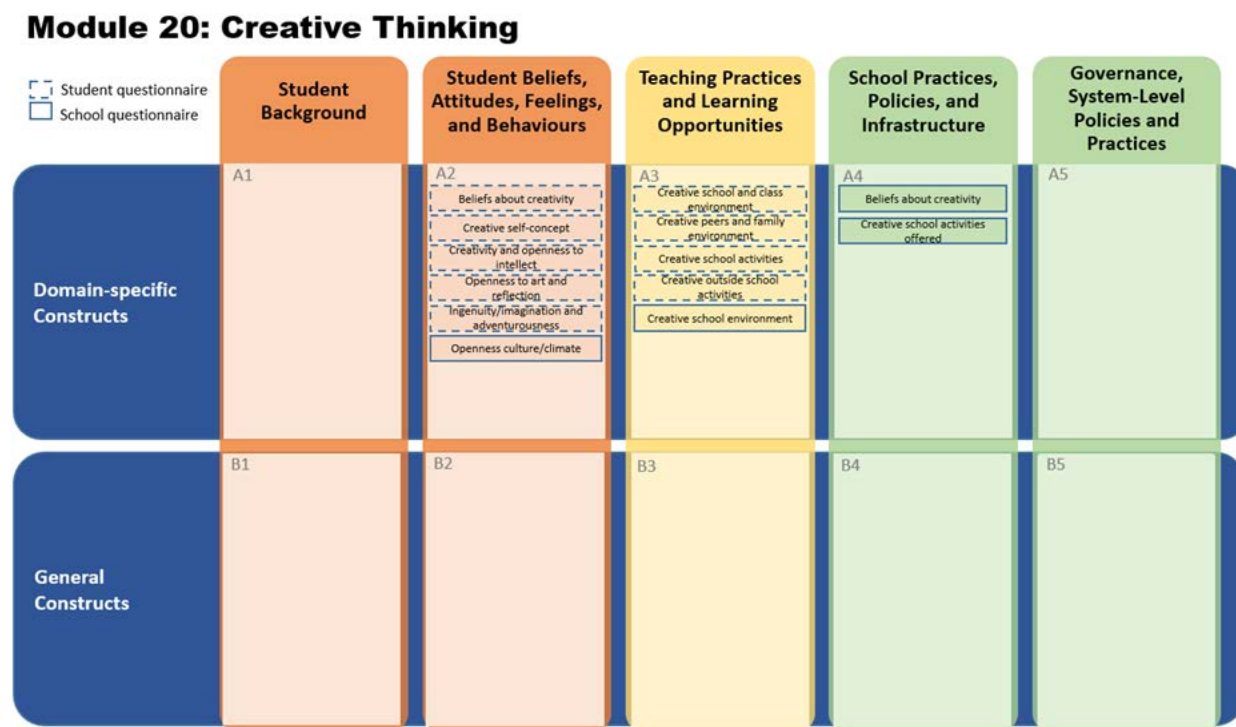
Global crises

The global spread of COVID-19 in 2020 has created unprecedented academic disruptions around the world, with the closures of schools leading to widespread losses in instructional time for students and the need for schools and education systems to turn to alternative learning opportunities that might mitigate these losses (OECD, 2020[155]). The disruption to schooling will likely have profound short- and long-term impacts on student learning and well-being, especially among students from diverse backgrounds who are more likely to face additional barriers to physical learning opportunities as well as social and emotional support available in schools (OECD, 2020[155]).

In PISA 2022, the global crises module (GCM) includes questions for students and school administrators to capture the experiences of education stakeholders during the COVID-19 pandemic. Constructs in this

module are described in an OECD working paper (Bertling et al., 2020[156]). As the module aims to collect information about the educational responses to the pandemic, the questions target one of the most widely implemented responses: the closure of school buildings to students.

Constructs in the STQ focus on the duration of school closure, resources during school closure, subjective perceptions of learning (and learning loss) during school closure, family and teacher support during school closure, and preparedness for future school closures. Constructs in the SCQ capture the number of days that school buildings were closed to students, organisation of instruction, resources available to students, factors hindering remote instruction, teacher communication with students, student attendance in distance learning activities, resources used to support teachers in providing remote instruction, stakeholder support for schools, school preparations for remote instruction, and school's preparedness for digital learning.

Figure 5.24. below illustrates how all constructs in this module map onto the taxonomy.

Figure 5.24. Constructs in global crises module

PISA 2022 Survey Design Principles

PISA has made significant contributions to the enhancement and refinement of survey design principles. However, its previous contextual questionnaire frameworks have not systematically evaluated different methodological approaches or described a comprehensive list of best practices and survey design principles to guide item development. For example, across PISA cycles there have been frequent changes to the number of response options, response option labels, the number of items within scaled indices, or the use of reversed keyed items. Moreover, lack of cross-cultural comparability of questionnaire scales partly due to response styles in PISA is a well-known challenge. While potential strategies for alternative item types (e.g., (Kyllonen, 2013[157])) as well as statistical approaches (e.g., (He et al., 2017[158])) have been explored, these have not always had the expected impact or have not led to noticeable shifts in how PISA data is reported and used. Lastly, measuring the above outlined constructs in PISA 2022 further

faces the challenges of implementing robust measurement approaches while keeping student burden low. This framework section presents a clear set of survey design principles that were applied to further enhance construct validity of the questionnaire measures in PISA 2022 and beyond, thereby strengthening the basis for cross-national and cross-cycle comparisons.

Table 5.8. below gives an overview of all principles, each of which will be described in more detail below.

Table 5.8. PISA 2022 Survey Design Principles

	PISA 2022 Survey Design Principles
Question types	1. Continue administering rating-scale type items with common response options grouped into matrix questions but harmonize the length of matrix questions to include fewer items per respondent per screen to balance efficiency with controlling cognitive load and respondent fatigue due to per-screen burden.
	2. Continue using alternative item formats introduced during the 2012-2018 cycles (e.g., anchoring vignettes, forced choice, situational judgments tests, slider bars) in cases where there is clear empirical evidence that these methods improve measurement.
	3. Minimize the use of fill-in/free response type questions if multiple-choice questions can yield acceptable data to reduce risk of coding inconsistencies and burden on countries for human coding.
Question Wording	4. For new item development, develop balanced scales with both positively and negatively framed items while avoiding double inversions or negations.
	5. **For new item development, place contextual cues (e.g., "outside of school", "during mathematics lessons"), if applicable, directly in the item rather than the question st**em to improve clarity and reduce wordiness and complexity of question stems.
	6. For new item development, avoid double- or multi-barrelled questions.
	7. For new item development, harmonize the number of examples in question or balance reading load while avoiding potential student misinterpretations of examples as definitions.
	8. For new matrix questions developed to capture reflective constructs, ensure sufficient distinctness of items by avoiding including items that are too similar (e.g., items that share a substantial number of words or repeat phrases used in other items).
Response Options	9. For new item development, use quantifiable frequency response options where possible instead of vague options to facilitate consistent interpretation across all respondents.
	10. For new item development, increase the number of response options from 4 to 5 where feasible to allow for more differentiation of responses across students.
	11. For trend scales re-administered from previous cycles, do not change response options unless there is fundamental issue with previous options to maximize chances for data-based comparisons across cycles.
	12. Display response options in ascending (lowest – highest) order for new questions but retain original order of intact scales retained from previous PISA cycles to facilitate cross-cycle comparisons.
Scaled Indices	13. Continue measuring reflective constructs with multi-item indices scaled based on Item Response Theory (IRT).
	14. For new development, establish a minimum number of five items per index to ensure at least a basic level of construct representation.
Routing	15. Use the affordances of the digital delivery platform to use deterministic routing for those questions where collection of more detailed information can be limited to a defined subset of students based on their responses to a previous question that defines a clear routing path.

Matrix Sampling	16. For questions reflective of latent constructs and designed for scaled indices, use a within-construct questionnaire matrix sampling design whereby individual students answer a subset of five items from a larger set of items for each construct.
	17. For questions representing manifest or formative constructs and designed for simple indices, collect data on each question from every student.
	18. Due not use matrix sampling for any question associated with the Index of ESCS to maximize comparability with previous cycles.
Use of log file data	19. Make questionnaire assembly decisions informed by timing data, to the extent that data from previous cycles or other testing programs is available.
	20. Utilize log file data to detect response patterns that may impact the quality of collected survey data (e.g., straight lining, rapid responding).
	21. Explore use of log file data to enhance survey-based measures of student test-taking motivation (e.g., timing data may be utilized to add to a measure of effort during the PISA test).

Question types

Use of Matrix Questions

Table 5.9. below provides an overview of the number of items included in matrix questions across past PISA cycles. On average matrix questions have included between 3 and 6 items, with some exceptions of questions with just two items, as well as a notable number of questions with 7 or more items.

For PISA 2022, the number of items in a matrix across questions will be harmonized to optimize the costs and benefits of using matrix questions over discrete single items. Recent research in the context of the *National Assessment of Educational Progress* (NAEP) showed that data quality of matrix questions is comparable to quality of discrete items, with the main difference that matrix questions take much less time to answer (Almonte and Bertling, 2018[159]). The response time benefit plays out especially the longer the matrix is, given that students have to read the stem initially and that time will be added to the first item response. At the same time, data quality suffers if matrices become too long. For example, findings from NAEP show that missing data rates increase if matrices become too long to fit on one screen without scrolling (i.e., higher missing rates are found particularly for those items at the end of a matrix that are not visible without scrolling). While reminders in the digital platform (e.g., prompts alerting respondents when an item on a page has not been answered) may help remedy these effects, it is not clear whether such reminders are equally well understood by test takers across the wide range of the PISA population.

Table 5.9. Number of Items in matrix questions across PISA Cycles

Number of Subitems	YR2000	YR2003	YR2006	YR2009	YR2012	YR2015	YR2018	Year PISA-D	Total	Average
2 sub-items	0	0	0	0	4	0	0	3	7	0.9
3 sub-items	5	2	3	2	4	6	14	5	41	5.1
4 sub-items	1	1	3	4	8	5	14	2	38	4.8
5 sub-items	2	3	4	5	6	9	10	2	41	5.1
6 sub-items	4	3	6	0	5	4	9	4	35	4.4
7 sub-items	4	0	1	3	0	2	3	1	14	1.8
8 sub-items	2	2	4	0	3	4	2	2	19	2.4
9 sub-items	1	0	0	2	5	2	2	1	13	1.6
10 sub-items	0	2	1	0	1	0	1	3	8	1.0
11 sub-items	1	0	0	1	0	2	0	2	6	0.8
12 sub-items	0	1	1	0	0	0	0	0	2	0.3
13 sub-items	0	0	0	1	1	0	0	0	2	0.3
14 sub-items	0	1	0	0	0	0	0	0	1	0.1
15 sub-items	0	0	0	0	0	0	0	1	1	0.1
16 sub-items	0	1	0	0	1	1	1	0	4	0.5
17 sub-items	0	0	2	1	1	0	0	0	4	0.5
18 sub-items	1	0	0	0	0	0	0	0	1	0.1
24 sub-items	1	0	0	0	0	0	0	0	1	0.1
28 sub-items	1	0	0	0	0	0	0	0	1	0.1
Total	23	16	25	19	39	35	56	26	239	

Note. Green bars denote frequency distributions for individual PISA years, orange bars denote frequencies of total counts across all years, and blue bars denote average frequencies across all PISA years.

Use of alternative item formats

Innovative item formats have been explored extensively across the PISA 2012 and 2015 PISA cycles. For instance, PISA 2012 explored the use of anchoring vignettes, situational judgment test items, overclaiming items, and forced choice (Kyllonen, 2013[157]). PISA 2015 continued using anchoring vignettes and introduced slider bars to take full advantage of the digital delivery platform.

Since the introduction of alternative items formats to PISA in 2012, their use in other LSA context questionnaires has so far found rather limited applications and validity studies have resulted in mixed results (e.g., (Bertling and Kyllonen, 2014[160]; Primi et al., 2018[161]; Stankov, Lee and von Davier, 2017[162]). Anchoring vignettes and situational judgment tests come with the added complexity that they pose greater demands on respondent time than more traditional rating-scale multiple-choice questions in order to fully exercise the benefits of these techniques. For instance, research with PISA 2012 anchoring vignettes showed that the technique could improve cross-cultural comparability of resulting scales when vignettes were applied to self-report items designed to measure the same construct (Bertling and Kyllonen, 2014[160]), which corresponds to the originally proposed application of the technique (King and Wand, 2007[163]), but the application of one or few sets of vignettes to multiple distinct scales capturing entirely different constructs may be problematic from a validity perspective (e.g., (Stankov, Lee and von Davier, 2017[162]; von Davier et al., 2017[164])). Including customized vignettes for every construct in the questionnaire, on the other hand, is not feasible within the time constraints of the PISA STQ administration. The most promising use of vignettes in the context of PISA may not be to recode original student responses but rather consider student responses to vignettes as additional complementary information on students' interpretations of the response options across countries and their use of the entire range of the offered scales (Bertling, 2018[165]).

Situational judgment tests are known for their relatively lower internal consistencies (a finding confirmed by PISA 2012 data; (Bertling, 2012[166])) calling for longer scales in order to meet reliability standards for LSAs. Forced choice items have a similar problem. While promising psychometric models are available that allow for the derivation of normative scales through ipsative data (Brown and Maydeu-Olivares, 2013[167]; Stark, Chernyshenko and Drasgow, 2005[168]), these methods require large numbers of items and pairing of many constructs in order to yield robust results. These conditions are typically not met in LSAs where most constructs are operationalized only through a few items and limited time is available. Mixed results have also been reported regarding test-taker perceptions of forced choice items, with sometimes negative impressions of forced choice items.

The most promising technique so far among the innovative item formats explored in PISA 2012 is the use of overclaiming items to adjust subjective topic familiarity ratings for students' tendencies to overclaim what they know and can do. The technique has been widely used in psychological and educational research ((Bensch et al., 2017[169]; Ziegler, Kemper and Rammstedt, 2013[170]), and recent applications in the context of the NAEP program in the United States, for instance, confirmed promising findings found in the context of PISA 2012. Another benefit of the overclaiming technique is that it comes at a relatively low cost – only few items need to be added to existing scales. Despite these benefits, an important caveat is that the overclaiming technique lends itself only to a very limited number of constructs (i.e., subjective ratings of familiarity with a topic), which makes it less promising as a technique to address cross-cultural equivalence concerns more broadly across a larger range of constructs (e.g., attitudinal or behavioural constructs).

In light of these considerations, the number of innovative item formats in PISA 2022 FT instruments was kept small and limited to those formats for which gains in validity are expected and/or additional relevant information about students' response behaviours can be collected.

Minimize use of open-ended fill-in-the-blank questions

Open-ended questions that ask the respondent to fill-in a response using constrained or unconstrained free text entry may be problematic for several reasons. In addition to concerns about potentially larger response time burden for the respondent, one of the main challenges in the context of PISA is that analysis of resulting data requires an initial step of coding student responses into quantifiable categories, as well as the necessary quality control steps to ensure coding accuracy. Accuracy of open-ended student responses is a well-known issue with regard to the coding of open-ended responses specifically for parental occupation questions (Kaplan and Kuger, 2016[32]; Tang, 2017[33]). PISA 2022 aimed to minimize the use of fill-in/free response type questions except for cases where text entry is limited to a small number of digits (e.g., questions about the number of days per week) to reduce risk of coding inconsistencies and burden on countries for human coding.

Question wording

Use of positive and negative statements

Balancing positively with negatively framed statements in questionnaire items designed to measure bipolar latent constructs is an established tradition in psychological measurement. For bipolar constructs, including both positively and negatively framed statements helps ensure that the entire range of a given construct from both poles of the theoretically defined construct is well represented. For unipolar constructs, which are defined theoretically only with regard to one pole, balancing statements might be less necessary. Balancing statements, however, may be still useful in these cases to minimize the risk of inviting undesired survey responding behaviours, such as "straightlining" (i.e., a response pattern where respondents chose options regardless of their content by creating a straight line across options chosen for several items in a matrix question), and it bears the chance to explore whether additional data cleaning steps may improve the validity and reliability of scales based on such items (Primi et al., 2019[171]; 2019[172]).

On the flipside, researchers have reported that respondents with poor reading proficiency may have difficulty responding accurately to scales that combine both positively and negatively worded items, specifically when negations are used, potentially leading to double-negatives (e.g., "I strongly disagree that mathematics is not one of my favourite subjects."). This problem may be minimized by refraining from using simple negations of positive statements when writing negatively framed statements (but see (Cacioppo and Berntson, 1994[173])). Table 5.10 below illustrates how negatively framed items can be written without the need to include negations.

Table 5.10. Examples of positively and negatively framed statements

Positively framed statement (examples)	Reversed keyed with negations (examples)	Negatively framed without negations (examples)
I am full of energy.	I am not full of energy.	I tire out quickly.
I finish things I start.	I don't finish things I start.	I leave things unfinished.
I finish things I start.	I don't finish things I start.	I leave things unfinished.

Another alternative approach that has been proposed is to present respondents with questions that intersperse items from scales of more or less socially desirable traits, rather than using reverse-scored items (e.g., (Gehlbach and Barge, 2012[174])). Interspersing items from different constructs in one matrix has been implemented in PISA only in a few select cases (e.g., assessment of mathematics anxiety and mathematics self-concept in a combined matrix question in PISA 2012) with the overarching number of items designed to represent a scale being grouped into one single matrix. The idea of interspersing items from different constructs in a common matrix has been recently explored in NAEP with findings pointing to only little differences in the factor structure and reliability of resulting indices. Potential benefits of creating construct heterogenous matrices should be carefully weighed against potential risks, including potentially increased cognitive load due to content variation across items in a matrix.

Contextual cue placement

Questionnaire items often ask students to report a behavioural frequency or indicate agreement with a statement when considering a specific contextual cue that may be provided in the question stem or in each individual item (see Table 5.11. below for an example). While placement of contextual cues in the question stem may seem somewhat more efficient from a reading load perspective, it may be less advisable considering research findings that respondents often place only little attention on reading information in the question stem. Placing an important contextual cue in the question stem bears the risk of students missing this piece of information and, consequentially, providing general rather than specific responses to each item. Recent findings from a large-scale pilot in the context of the United States' NAEP assessment are in line with this assumption (Qureshi, Alegre and Bertling, 2018[175])

Table 5.11. Examples of contextual cue placement in question Stem vs. Item

Contextual cue placement in stem only (example)	Contextual cue placement in each item (example)
Thinking about your mathematics class, how much do you agree or disagree with each of the following statements? a) I come to class prepared. b) I finish my homework right away. c) I enjoy participating in group activitites.	How much do you agree or disagree with each of the following statements? a) I come to my mathematics class prepared. b) I finish my mathematics homework right away. c) I enjoy participating in group activities in my mathematics class.
Word count: 38	Word count: 40

Avoid multi-barrelled statements

An established key principle in survey methodology is not to combine multiple ideas or statements into a single item because of the resulting multi-barrelledness and statistical confounding of student responses (Dillman, Smyth and Christian, 2014[176]; Gehlbach and Artino, 2018[177]). Table 5.12. below shows examples of double- or multi-barrelled items, alongside alternative wording as single statement items.

Table 5.12. Examples of single- vs. multi-barrelled statements

Double- or Multi-barrelled statement (examples)	Alternative wording as multiple single statement items.
I am relaxed and handle stress well.	Statement 1: I am relaxed. Statement 2: I handle stress well.
I am helpful and unselfish with others.	Statement 1: I am helpful to others. Statement 2: I am unselfish with others.

Choose a meaningful number of examples

A notable number of questions used in previous PISA STQ and SCQ include examples. These examples are necessary to convey what information the respondent is asked to provide and to clarify potential ambiguities of broad terms, such as "classical literature" or "digital devices". Table 5.13. illustrates that items may differ with regard to the number of examples used and outlines potential validity concerns related to the use of too few or too many examples in an item. In order to maximize the utility of examples in PISA 20221 examples were harmonized to a range of 2-5, where feasible. In addition, country-specific examples should be allowed for inclusion.

Table 5.13. Illustration of questions with different numbers of examples

Number of examples provided	Example item	Potential validity concern(s)
Single example	Which of the following are in your home? Classical literature (e.g., <Shakespeare>) *(from PISA 2018)*	• Students may misinterpret the parenthetical as a definition rather than an example if only a single example is provided.
More than 5 examples	What kind of job does your father have? Machine Operator (e.g., dry-cleaner, worker in clothing or shoe factory, sewing machine operator, paper products machine operator, crane operator, bus driver, truck driver) *(from PISA-D)*	• The long list of examples increases the reading load of the question, and consequentially the cognitive load, which may affect understanding particularly for respondents with lower proficiency levels. • Some respondents may also misinterpret the long list of terms in the parenthetical as a full list of possible exemplars of a larger category rather than as examples.

Minimise surface-level similarities in wording across matrix question items

While most matrix questions used in the PISA STQ and SCQ are designed to measure latent constructs by asking respondents a range of similar, yet related questions, it is important that statements are sufficiently distinct to avoid issues of co-linearity between data collected on each item, which may complicate IRT-scaling and inflated internal consistencies. Moreover, including statements that are too similar in the questionnaire may limit the value of the questions for reporting, unless there is strong reason to keep item wording consistent with previously used items' wording or for comparability with other studies.

Table 5.14 below provides an example of statements deemed potentially too similar alongside an illustration how surface-level similarities between the items in the same matrix may be reduced.

Table 5.14. Example of questions with surface-level similarities

Scale with potentially too similar items	Rewording of items to reduce surface-level similarities

To what extent do you agree or disagree with the following statements?
    a) I finish what I start.
    b) I finish tasks despite difficulties in the way.
	To what extent do you agree or disagree with the following statements?
	    a) I finish what I start.
	    b) I complete tasks despite difficulties in the way.

Response options

Number of options

Across the past seven PISA cycles, the STQ and SCQ have used a broad range of rating scale response option sets, most of which included four response options (see Figure 5.25 for an overview).

Based on current knowledge in survey method research, five response options have been proposed as an optimal number for any survey question to collect data of sufficient variability (Revilla, Saris and Krosnick, 2014[178]) and researchers have cautioned against using response options with too few (i.e., two or three; (Lee and Paek, 2014[179])) or too many categories as well as neutral middle categories (Alwin, Baumgartner and Beattie, 2018[180]). PISA 2022 questionnaires will balance the need to have sufficiently many data points along which student responses can be distinguished with the respondents' inability to distinguish too many response options and the desire to keep response options as simple as possible to facilitate translations and adaptations. For new item development, it is recommended to increase the number of response options from four to five where feasible to allow for more differentiation of responses across students and more advanced statistical modelling. At the same time, it is *not* recommended to introduce a fifth (middle) category to the established PISA 4-point agreement scale given its longstanding use in PISA, unless there is a specific reason for the particular construct why a middle category would improve validity or cross-cultural comparability.

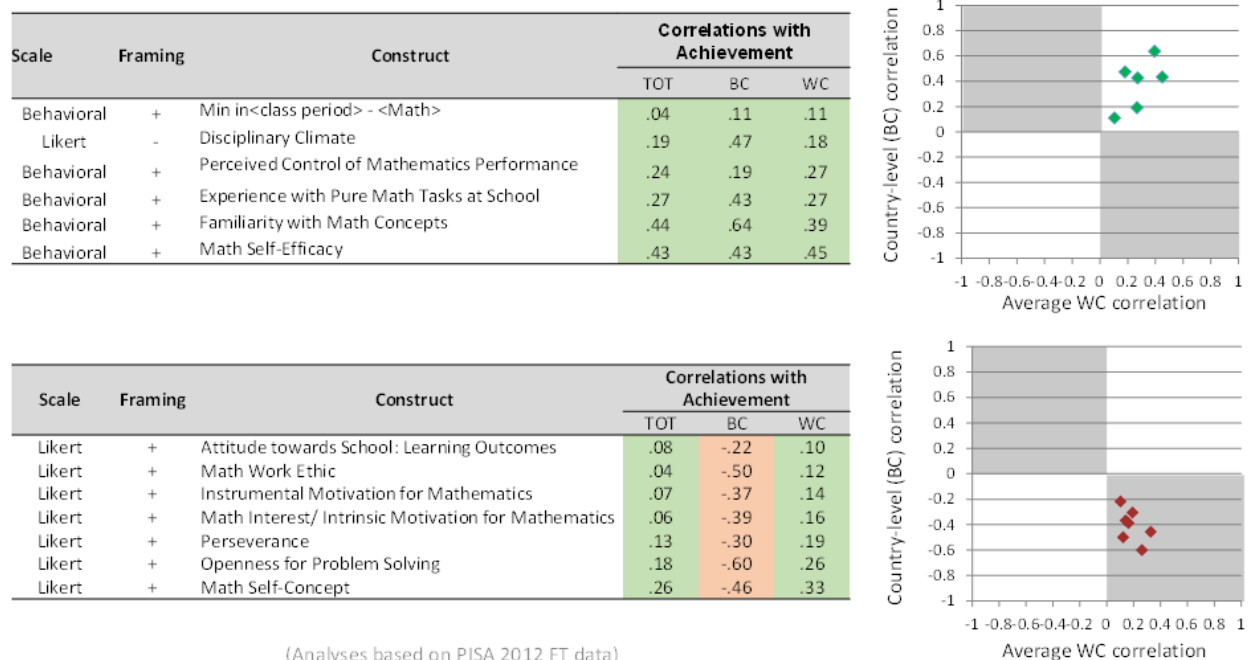## Figure 5.25. Previously used rating-scale response options in PISA 2000-2018

Type of response option	Order	Number of Options	Response Label	2000	2003	2006	2009	2012	2015	2018	Total
Agreement	Decrease	4	Strongly agree / Agree / Disagree / Strongly disagree	0	7	5	0	13	2	2	29
Agreement	Increase	4	Disagree/Disagree somewhat/Agree somewhat/Agree	1	0	0	0	0	0	0	1
Agreement	Increase	4	Strongly disagree/Disagree/Agree/ Strongly agree	3	0	0	3	0	7	16	29
Amount	Decrease	4	Very confident/Confident/Not very confident/Not at all confident	0	1	0	0	1	0	0	2
Amount	Decrease	4	Very important/Important/Of little importance/Not important at all	0	0	1	0	0	0	0	1
Amount	Decrease	4	Very likely/Likely/Slightly likely/Not at all likely	0	0	0	0	1	0	0	1
Frequency	Decrease	4	Always or almost always/Often/Sometimes/Never or rarely	0	0	0	0	2	0	0	2
Frequency	Decrease	4	Frequently / Sometimes / Rarely / Never	0	0	0	0	5	0	0	5
Frequency	Decrease	4	In all lessons/In most lessons/In some lessons/Never or hardly ever	0	0	1	0	0	1	0	2
Frequency	Decrease	4	Very often/Regularly/Sometimes/Never or hardly ever	0	0	1	0	0	1	0	2
Frequency	Increase	4	Almost never/Sometimes/Often/Almost always	1	0	0	1	0	0	0	2
Frequency	Increase	4	Never or almost never/A few times a year/A few times a month/Once a week or more	0	0	0	0	0	2	1	3
Frequency	Increase	4	Never or almost never/About once a week/2 to 3 times a week/Almost every day	0	0	0	0	0	0	0	0
Frequency	Increase	4	Never or almost never/Some lessons/Many lessons/Every lesson or almost every lesson	0	0	0	0	0	3	2	5
Frequency	Increase	4	Never or hardly ever/A few times per year/About once a month/Several times a month	1	0	0	0	0	0	0	1
Frequency	Increase	4	Never or hardly ever/In some lessons/In most lessons/In all lessons	0	0	0	3	0	0	1	4
Frequency	Increase	4	Never or hardly ever/Once or twice a year/About 3 or 4 times a year/More than 4 times a year	1	0	0	0	0	0	0	1
Frequency	Increase	4	Never/One or two times/Three or four times/Five or more times	0	0	0	0	0	1	1	2
Frequency	Increase	4	Never/Rarely/Sometimes/Always	0	0	0	0	0	0	1	1
Frequency	Increase	4	Never/Some lessons/Most lessons/Every lesson	2	0	0	0	0	0	0	2
Frequency	Increase	4	Never/Sometimes/Most of the time/Always	1	0	0	0	0	0	0	1
Amount	Decrease	5	Very much like me/Mostly like me/Somewhat like me/Not much like me/Not at all like me	0	0	0	0	2	0	4	6
Amount	Increase	5	Not at all true of me/Slightly true of me/Moderately true of me/Very true of me/Extremely true of me	0	0	0	0	0	0	1	1
Amount	Increase	5	Not at all true/Slightly true//Very true /Extremely true	0	0	0	0	0	0	2	2
Frequency	Increase	5	Never or almost never/A few times a year/About once a month/Several times a month/Several times a week	0	0	0	1	0	0	1	2
Frequency	Increase	5	Never or hardly ever/A few times a year/About once a month/Several times a month/Several times a week	4	0	0	0	0	0	0	4
Frequency	Increase	5	Never/A few times a year/About once a month/Several times a month/Several times a year	0	0	0	1	0	0	0	1

Note: Green bars denote frequency distributions for individual PISA years; orange bars denote frequencies of total counts across all years.

Use of agreement and frequency scales

While the overwhelming majority of questions in past PISA cycles have used agreement-type response options (see Figure 5.25), decades of survey methodological research have demonstrated a range of issues with this type of verbal framing of questions, including their proneness to acquiescence response bias and high cognitive burden (e.g., (Revilla, Saris and Krosnick, 2014[178])). Bertling and Kyllonen (2014[160]) have shown that scales in the PISA 2012 STQ were especially prone to the so-called "Attitude-Achievement Paradox" (i.e., a phenomenon whereby scales correlate positively with achievement within a group [e.g., country] but correlations flip to the negative when aggregated group-level data [e.g., country-level data] is considered) when positively framed agreement response options were used. In contrast, scales using negatively framed agreement response options or behavioural frequency response options were not affected by the phenomenon (see Figure 5.26 below). These findings seem to indicate that frequency-based response options may be preferable over agreement-type options.

Figure 5.26. Scales affected vs. not affected by attitude achievement paradox in PISA 2012

Scale	Framing	Construct	Correlations with Achievement		
			TOT	BC	WC
Behavioral	+	Min in<class period> - <Math>	.04	.11	.11
Likert	-	Disciplinary Climate	.19	.47	.18
Behavioral	+	Perceived Control of Mathematics Performance	.24	.19	.27
Behavioral	+	Experience with Pure Math Tasks at School	.27	.43	.27
Behavioral	+	Familiarity with Math Concepts	.44	.64	.39
Behavioral	+	Math Self-Efficacy	.43	.43	.45

Scale	Framing	Construct	Correlations with Achievement		
			TOT	BC	WC
Likert	+	Attitude towards School: Learning Outcomes	.08	-.22	.10
Likert	+	Math Work Ethic	.04	-.50	.12
Likert	+	Instrumental Motivation for Mathematics	.07	-.37	.14
Likert	+	Math Interest/ Intrinsic Motivation for Mathematics	.06	-.39	.16
Likert	+	Perseverance	.13	-.30	.19
Likert	+	Openness for Problem Solving	.18	-.60	.26
Likert	+	Math Self-Concept	.26	-.46	.33

(Analyses based on PISA 2012 FT data)

Source: (Bertling & Kyllonen, 2014)
Note: TOT= correlation with achievement for total (pooled) sample across all countries; BC= between-country correlation based on aggregated country-level data; WC= average within-country correlation across all countries.

It should be noted that response option type and construct were confounded in the aforementioned analyses in PISA 2012, which is why additional research in the specific context of PISA is recommended prior to considering *replacement* of agreement-type questions with frequency-type questions across all constructs. Please note, many of the constructs described in the previous sections, especially the outlined social and emotional characteristics, by definition entail a subjective (and possibly culturally dependent) component and metric or scalar invariance across different cultural groups may therefore be unwarranted. While most of these subjective constructs have traditionally been assessed with agreement type scales, different possible response option sets for PISA 2022 have been explored in cognitive interviews and the international FT. For PISA 2022 response options that allow for more informative and less ambiguous

reporting to technical and non-technical audiences were given priority. Table 5.15. summarises which response option sets are used in PISA 2022).
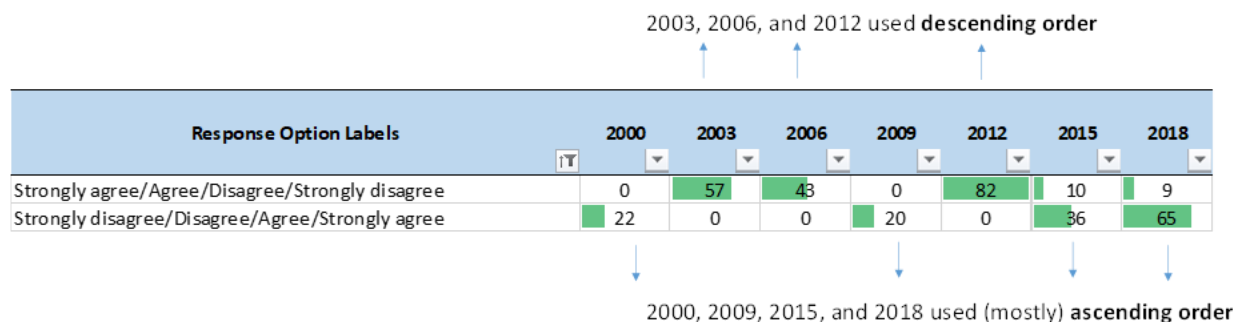
Table 5.15. Response option sets in PISA 2022

Scale type	Response options	Use in PISA 2022
Agreement	Strongly disagree – Disagree – Agree – Strongly agree OR Strongly disagree – Disagree – Neither disagree nor agree – Agree – Strongly Agree	• Retained for re-administration of previously used PISA questions or questions from other OECD surveys (e.g., SSES) • Avoided in new questions
Abstract Frequency	Never– Rarely – Sometimes – Often – Always	• Retained for re-administration of previously used PISA questions only • Avoided in new questions
Absolute Approximate Frequency	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	• Used for new questions
Relative Approximate Frequency	Never or almost never – Less than half of the lessons – About half of the lessons – More than half of the lessons – Every lessons or almost every lesson	• Used for new questions

Harmonizing directionality of response options

Figure 5.27 below shows how the directionality of response options for the most commonly used PISA questionnaire response options changed since the first PISA cycle in 2000. While response options were administered strictly in ascending order in 2000 and 2009, response options were administered strictly in descending order in 2003, 2006, and 2012. The 2015 and 2018 cycles used a hybrid approach where most questions used ascending order but some questions introduced in earlier cycles were kept in descending order. While harmonizing the directionality of response options in PISA 2022 would likely improve the student experience by making it more consistent across the questionnaire, statistical concerns about backwards comparability of data need to be taken into account. Past FT experiments for PISA 2015 had shown notable effects on item parameters of the direction of response options, and therefore the direction of response options for scales retained from previous PISA cycles will remain unchanged.

Figure 5.27. Variation in directionality of response options from PISA 2000-2018

2003, 2006, and 2012 used **descending order**

Response Option Labels	2000	2003	2006	2009	2012	2015	2018
Strongly agree/Agree/Disagree/Strongly disagree	0	57	43	0	82	10	9
Strongly disagree/Disagree/Agree/Strongly agree	22	0	0	20	0	36	65

2000, 2009, 2015, and 2018 used (mostly) **ascending order**

Note: Green bars denote frequency distributions for individual PISA years.

Scaled indices

Distinguishing manifest, reflective, and formative constructs

The constructs outlined in this module can be distinguished into constructs that are manifest in nature (i.e., are directly observable and reportable based on respondent answers to a single question) and constructs that are not directly observed and cannot be reported on based on respondent answers to a single question but require the creation of indices for reporting. The latter category can be further differentiated into reflective constructs and formative constructs (for an overview see (Bollen and Lennox, 1991[181]); (Edwards and Bagozzi, 2000[182])). Table 5.16. below lists some examples from the PISA context for manifest, reflective, and formative constructs.

Table 5.16. Examples of manifest, reflective, and formative constructs in PISA

Manifest constructs in PISA (examples)	Reflective constructs in PISA (examples)	Formative constructs in PISA (examples)
Parental Education	Sense of Belonging	ESCS
Parental Occupation	Mathematics Self-efficacy	
Mathematics Class Periods per Week	Perseverance	

Reflective constructs can be formalized into latent variable models, which often make a unidimensionality assumption of a single statistical cause that determines responses on the items reflective of the construct (MacCallum and Browne, 1993[183]). Social science usually assumes constructs are reflective (Bollen, 2002[184]), and most of the student attitudes, values, and beliefs constructs described in this framework fall into this category: the underlying trait determines how students think, feel, and behave in certain situations.

In contrast, formative constructs are theoretically inconsistent with latent variable models. Socioeconomic status is often considered the archetypical example of a formative construct (e.g., (Bollen and Lennox, 1991[181])). Another example from PISA that would classify as formative are students' OTL in Mathematics. Unlike the case with reflective constructs, indicators such as parental education or students' exposure to certain type of mathematics problems are not assumed to be caused by ESCS or OTL respectively. Instead, different levels of ESCS or OTL are assumed to emerge when a set of theoretically defined components are combined together. As a result, changes to the item composition necessarily changes the construct. Formative constructs therefore are less suitable for the use of IRT modelling (Howell, Breivik and Wilcox, 2007[185]).

Number of Items per Scaled Index

Despite the consistency in scaling indices based on IRT, there is considerable variation with regard to the number of items used in scaled indices across the questionnaires from PISA 2000 – PISA 2018. Some reflective constructs have been targeted with a single item (e.g., Growth mindset) or very few items (e.g., Sense of purpose) whereas 10 or more items were used to scale other constructs (e.g., Familiarity with Mathematical Concepts). Table 5.17 lists additional examples for short and long questionnaires scales across the last three PISA cycles.

Table 5.17. Examples of short and long questionnaire scales in PISA

Examples of short student questionnaire scales (5 items or less)		Examples of long student questionnaire scales (8 or more items)	
Less than 4 items	• PISA 2018: Growth Mindset (1 item) • PISA 2015 and 2018: Life Satisfaction (1 item) • PISA 2018: Attitude Towards School; Sense of Purpose (3 items)	8 items	• PISA 2012: Mathematics Self-Efficacy • PISA 2015: Science Self-Efficacy; Collaboration
4 items	• PISA 2015: Interest and Valuing of Science • PISA 2018: Motivation, Performance Anxiety	9 items	• PISA 2012: Sense of Belonging; Cognitive Activation; Mathematics Work Ethic • PISA 2018: Positive and Negative Affect; Test Language Reading Activities
5 items	• PISA 2012: Perseverance; Enjoyment of Problem Solving • PISA 2015: Test Anxiety • PISA 2018: Enjoyment of Reading; Perspective Taking	10 or more items	• PISA 2012: Familiarity with Mathematical Concepts (13 items); Home Possessions (17 items) • PISA 2015: Science Learning Activities (10 items) • PISA 2018: Engagement in Global Issues (10 items)

While including single items or very short scales may be appealing from the administration perspective, it might be problematic from the measurement perspective. In order to provide valid and reliable measurement of most contextual variables and additional constructs across participating education systems, it is crucial to rely on multiple indicators for the construct at hand.

PISA 2022 will continue measuring reflective constructs with multi-item indices scaled based on Item Response Theory. For new development, a new lower bound of five for the number of questions in any index was established to ensure reasonable levels of internal consistency and construct representation.

Routing

When PISA questionnaires were delivered on paper, the possibilities to customize individual student experiences through routing were extremely limited. The transition to the digital delivery platform in 2015 opened new possibilities for a routing approach, whereby respondents receive different questions based on their responses to previous survey questions. The approach has been used for specific questions, such as to administer follow-up questions, but it has so far not been widely used to increase the efficiency of collecting data for key PISA constructs, such as ESCS.

Beyond the measurement of ESCS, deterministic routing and skip patterns will be used for manifest constructs where a clear path can be specified a priori. When introducing routing, an additional important consideration is how to provide countries that will administer questionnaires on paper with an as seamless as possible respondent experience.

Matrix Sampling

Constraints of overall testing time and the large sample sizes in large-scale assessments make matrix-sampling approaches, whereby different respondents receive different sets of items, a viable option to reduce burden while maintaining content coverage across relevant areas. Matrix-sampling approaches are the standard practice for the subject-area tests in educational large-scale assessments (Comber and Keeves, 1973[186]; OECD, 2013[187]) and have more recently been used as an alternative to single-form questionnaire designs.

PISA 2012 utilized a three-booklet questionnaire matrix sampling design whereby individual students received one of three possible booklets containing only a subset of all survey questions administered. This

approach, which is illustrated in Figure 5.28., allowed for testing a total of 41 minutes of questionnaire material in the main survey with each individual student's time limited to 30 minutes, i.e., the design allowed for collection of data on 33 percent more questions than in previous cycles without increasing individual student burden (Adams, Lietz and Berezner, 2013[188]).

Figure 5.28. Schematic illustration of the PISA 2012 3-booklet matrix sampling design
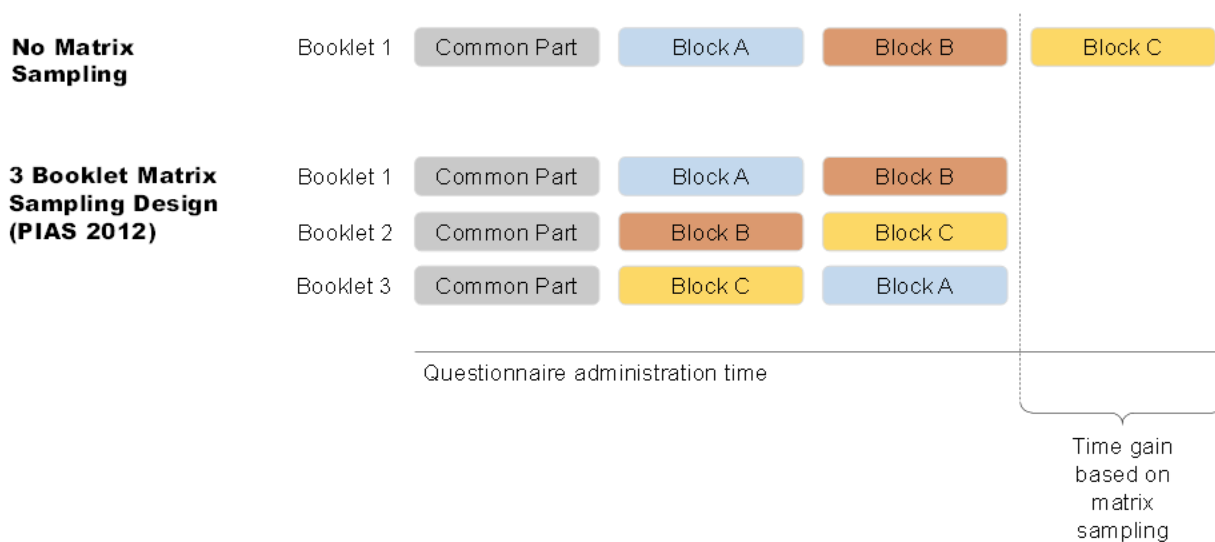
A disadvantage of the 2012 three-form design was that entire constructs were rotated rather than rotating individual items within constructs. Thus, one student might answer questions on certain constructs while another student might answer questions on entirely different constructs, but no student answered questions on all constructs. While many researchers reported very small to negligible impact on the overall measurement model, including conditioning and estimation of plausible values (Adams, Lietz and Berezner, 2013[188]; Almonte, 2014[189]; Kaplan and Su, 2014[190]; Monseur and Bertling, 2014[191]), methodological concerns about possible attrition in sample size when conducting multivariate regression models and biases in the estimation of plausible values under the construct-level 2012 rotation design have also been raised (von Davier, 2013[192]).

PISA 2015 and 2018 reverted back to a single questionnaire form and extended the questionnaire time from 30 to 35 minutes to find a compromise between providing a non-matrix sampled data set and including more variables than feasible to include in a 30-minute booklet.

Over the past five years, research has advanced and brought forward new insights about risks and benefits of using matrix sampling for questionnaires, including the exploration of alternative approaches that may prevent the challenges encountered with the 2012 design (e.g., (Bertling and Weeks, 2018[193]; Bertling and Weeks, 2018[194]; Kaplan and Su, 2016[195])).

PISA 2022 will utilize an alternative matrix sampling design to the one used in PISA 2012, which would rotate questions within constructs instead of across constructs.

Figure 5.29. Comparison of construct-level missing data structure resulting from alternative matrix sampling approaches
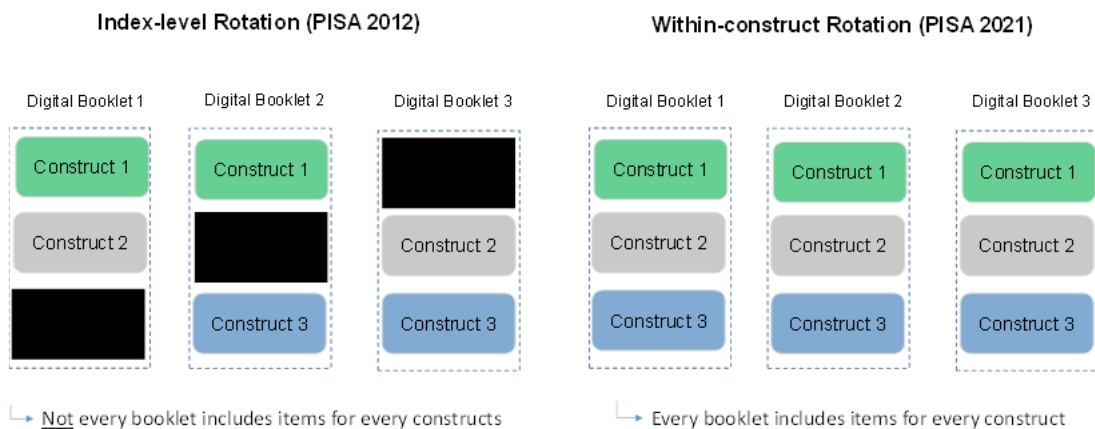

Figure 5.29. illustrates the differences between index-level and within-construct matrix sampling designs in terms of construct-level missing data. Unlike the PISA 2012 design, in the PISA 2022 design, every student will receive questions on all constructs but only answer a subset of all questions for each construct.

Bertling and Weeks (2018[193]; 2018[194]), presented findings from a series of simulation studies using PISA 2012 and PISA 2015 data to the PISA TAG and QEG and concluded that there is no statistical reason to rule out within-construct matrix sampling as a potential operational design for the PISA 2022 MS. Differences found in a first study between fixed vs. random selection of anchor items and rotated items were practically negligible, suggesting that both designs would be feasible in PISA (Bertling and Weeks, 2018[193]). Results from a second study (Bertling and Weeks, 2018[194]) clearly indicated that within-construct matrix sampling with a random choice of rotated items offers the best results among different matrix sampling approaches. Moreover, findings are in strong support that a design where five items are randomly selected from each item matrix will offer superior data for backwards trend analyses than a single form shortened five item scale or designs with anchor items.

Based on discussing a range of possible alternative designs with the PISA TAG and evaluating the new design during the FT stage, PISA 2022 will utilize a design where a random set of five items per construct (drawn from a set of 8-10 items total for each construct) is administered to each student for those questions designed for the creation of scaled indices.

Log file data

Since 2015, the PISA assessment has made the transition to computer-based formats. Besides the answers to cognitive and context questionnaire material, the electronic assessment platform captures basic test takers' behavioural data, also known as log-file data (OECD, 2017[196]). These log-file data can be used for various purposes. For instance, in PISA 2015 and 2018, the answering time was used to guide content selection after the FT. This practice continued during the analysis of the PISA 2022 FT data.

Survey response behaviours captured by log-file data may also be used to relate to cognitive processes (Almond et al., 2012[197]; Couper and Kreuter, 2013[198]; Naumann, 2015[199]; Yan and Tourangeau, 2007[200]). In recent studies, log-file analysis has been used to measure motivation (Hershkovitz and Nachmias, 2009[201])), or to link answering behaviour to aspects of personality (Papamitsiou and Economides, 2017[202]) or students' learning styles (Agudo-Peregrina et al., 2014[203]; Efrati, Limongelli and Sciarrone, 2014[204]).

Accordingly, research interest in this area is growing rapidly. While the Programme for the International Assessment of Adult Competencies (PIAAC) study has published an online LogData analyzer tool that allows for easy access to these data for secondary analyses, open access to PISA log data is still missing. The PISA questionnaires in 2022 will once again be assessed via a CBA platform, thus the captured log-file data could potentially be used in secondary research analyses to explore relationships between answering behaviour and outcomes, in addition to informing content selection post-FT.

As fundamental research is missing on the relationship between context indicators as assessed by tests and questionnaires and corresponding data from log-files, making the PISA data accessible for further research seems to be a promising starting point. Although (Jude and Kuger, 2018[205])point out that currently "no theoretical frameworks exist specifying which kind of log-file data would be the most promising to contribute additional information in ILSAs," making these data accessible could help researchers explore theories, compare relationships in different countries, and help to identify new item types that would yield useful log-file data in future PISA cycles.

References

Abedi, J. (2006), *English language learners and math achievement: A study of opportunity to learn and language accommodation*, University of California.   [15]

Abrahams, L. et al. (2019), "Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts.", *Psychological Assessment*, Vol. 31/4, pp. 460-473, https://doi.org/10.1037/pas0000591.   [75]

Adams, R., P. Lietz and A. Berezner (2013), "On the use of rotated context questionnaires in conjunction with multilevel item response models", *Large-scale Assessments in Education*, Vol. 1/1, https://doi.org/10.1186/2196-0739-1-5.   [188]

Agudo-Peregrina, Á. et al. (2014), "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning", *Computers in Human Behavior*, Vol. 31, pp. 542-550, https://doi.org/10.1016/j.chb.2013.05.031.   [203]

Almlund, M. et al. (2011), "Personality Psychology and Economics", in *Handbook of The Economics of Education, Handbook of the Economics of Education*, Elsevier, https://doi.org/10.1016/b978-0-444-53444-6.00001-8.   [8]

Almond, R. et al. (2012), "A PRELIMINARY ANALYSIS OF KEYSTROKE LOG DATA FROM A TIMED WRITING TASK", *ETS Research Report Series*, Vol. 2012/2, pp. i-61, https://doi.org/10.1002/j.2333-8504.2012.tb02305.x.   [197]

Almonte, D. (2014), *Spiraling of contextual questionnaires in the NAEP TEL pilot assessment. In Bertling, J. P. (Chair), Spiraling contextual questionnaires in educational large-scale assessments.*.   [189]

Almonte, D. and J. Bertling (2018), *Effects of item format (discrete vs. matrix) on grade four student responses. In New insights on survey questionnaire context effects from multiple large-scale assessments.*.   [159]

Alwin, D., E. Baumgartner and B. Beattie (2018), "Number of Response Categories and Reliability in Attitude Measurement†", *Journal of Survey Statistics and Methodology*, Vol. 6/2, https://doi.org/10.1093/jssam/smx025.   [180]

Alwin, D., E. Baumgartner and B. Beattie (2017), "Number of Response Categories and Reliability in Attitude Measurement†", *Journal of Survey Statistics and Methodology*, Vol. 6/2, pp. 212-239, https://doi.org/10.1093/jssam/smx025.   [219]

Astor, R., R. Benbenishty and J. Estrada (2009), "School Violence and Theoretically Atypical Schools: The Principal's Centrality in Orchestrating Safe Schools", *American Educational Research Journal*, Vol. 46/2, pp. 423-461, https://doi.org/10.3102/0002831208329598.   [62]

Astor, R., N. Guerra and R. Van Acker (2010), "How Can We Improve School Safety Research?", *Educational Researcher*, Vol. 39/1, pp. 69-78, https://doi.org/10.3102/0013189x09357619.   [65]

Babcock, P. and K. Bedard (2011), "The Wages of Failure: New Evidence on School Retention and Long-Run Outcomes", *Education Finance and Policy*, Vol. 6/3, pp. 293-322, https://doi.org/10.1162/edfp_a_00037.   [89]

Baird, J. et al. (2014), *State of the Field Review: Assessment and Learning.*.   [126]

Bansak, K., J. Hainmueller and D. Hangartner (2016), "How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers", *Science*, Vol. 354/6309, pp. 217-222, https://doi.org/10.1126/science.aag2147.   [48]

Beal, S. and L. Crockett (2010), "Adolescents' occupational and educational aspirations and expectations: Links to high school activities and adult educational attainment.", *Developmental Psychology*, Vol. 46/1, pp. 258-265, https://doi.org/10.1037/a0017416.   [47]

Beghetto, R., J. Baer and J. Kaufman (2015), *Teaching for creativity in the common core classroom*, Teachers College Press.   [152]

Beghetto, R. and J. Kaufman (2007), "Toward a broader conception of creativity: A case for "mini-c" creativity.", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 1/2, pp. 73-79, https://doi.org/10.1037/1931-3896.1.2.73.   [153]

Beghetto, R. and J. Plucker (2006), "The Relationship Among Schooling, Learning, and Creativity: "All Roads Lead to Creativity" or "You Can't Get There from Here"?", in *Creativity and Reason in Cognitive Development*, Cambridge University Press, https://doi.org/10.1017/cbo9780511606915.019.   [154]

Bensch, D. et al. (2017), "Teasing Apart Overclaiming, Overconfidence, and Socially Desirable Responding", *Assessment*, Vol. 26/3, pp. 351-363, https://doi.org/10.1177/1073191117700268.   [169]

Berkenmeyer, N. and S. Müller (2010), "Schulinterne Evaluation - nur ein Instrument zur Selbststeuerung von Schulen? / School-internal evaluation-only an instrument for self-control of schools?", *H. Altrichter & K. Maag Merki (Eds.), Handbuch neue Steuerung im Schulsystem / Handbook New Control in the School System*, Wiesbaden: VS Verlag für Sozialwissenschaften., pp. 195 - 218.   [133]

Bertling, J. (2018), *Anchoring Vignettes and Situational Judgment Tests in PISA 2012.*.   [165]

Bertling, J. (2012), *Students' approaches to problem solving-comparison of MTMM models for SJT data from PISA 2012.*   [166]

Bertling, J. and P. Kyllonen (2014), *Improved measurement of noncognitive constructs with anchoring vignettes.*. [160]

Bertling, J., T. Marksteiner and P. Kyllonen (2016), "General Noncognitive Outcomes", in *Methodology of Educational Measurement and Assessment, Assessing Contexts of Learning*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-45357-6_10. [9]

Bertling, J. et al. (2020), "A tool to capture learning experiences during COVID-19: The PISA Global Crises Questionnaire Module", *OECD Education Working Papers*, No. 232, OECD Publishing, Paris, https://doi.org/10.1787/9988df4e-en. [156]

Bertling, J. and J. Weeks (2018), *Plans for Within-construct Questionnaire Matrix Sampling in PISA 2021.*. [193]

Bertling, J. and J. Weeks (2018), *Within-construct Questionnaire Matrix Sampling: Comparison of Different Approaches for PISA 2021.*. [194]

Binder, M. (2009), "Why Are Some Low-Income Countries Better at Providing Secondary Education?", *Comparative Education Review*, Vol. 53/4, pp. 513-534, https://doi.org/10.1086/603641. [91]

Black, P. (2015), "Formative assessment – an optimistic but incomplete vision", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 161-177, https://doi.org/10.1080/0969594x.2014.999643. [127]

Black, P. and D. Wiliam (1998), "Assessment and Classroom Learning", *Assessment in Education: Principles, Policy & Practice*, Vol. 5/1, pp. 7-74, https://doi.org/10.1080/0969595980050102. [137]

Blau, D. and J. Currie (2006), "Chapter 20 Pre-School, Day Care, and After-School Care: Who's Minding the Kids?", in *Handbook of the Economics of Education*, Elsevier, https://doi.org/10.1016/s1574-0692(06)02020-4. [43]

Blum, W. and D. Leiss (2007), "Investigating quality mathematics teaching: The DISUM Projec", *C. Bergsten & B. Grevholm (Eds.), Developing and researching quality in mathematics teaching and learning. Proceedings of MADIF 5, SMDF, Linköping*, pp. 3-16. [117]

Bollen, K. (2002), "Latent Variables in Psychology and the Social Sciences", *Annual Review of Psychology*, Vol. 53/1, pp. 605-634, https://doi.org/10.1146/annurev.psych.53.100901.135239. [184]

Bollen, K. and R. Lennox (1991), "Conventional wisdom on measurement: A structural equation perspective.", *Psychological Bulletin*, Vol. 110/2, pp. 305-314, https://doi.org/10.1037/0033-2909.110.2.305. [181]

Bowman, N. (2010), "College Diversity Experiences and Cognitive Development: A Meta-Analysis", *Review of Educational Research*, Vol. 80/1, pp. 4-33, https://doi.org/10.3102/0034654309352495. [66]

Boyd, D. et al. (2009), "Teacher Preparation and Student Achievement", *Educational Evaluation and Policy Analysis*, Vol. 31/4, pp. 416-440, https://doi.org/10.3102/0162373709353129. [120]

Broh, B. (2002), "Linking Extracurricular Programming to Academic Achievement: Who Benefits and Why?", *Sociology of Education*, Vol. 75/1, p. 69, https://doi.org/10.2307/3090254. [82]

Brown, A. and A. Maydeu-Olivares (2013), "How IRT can solve problems of ipsative data in forced-choice questionnaires.", *Psychological Methods*, Vol. 18/1, pp. 36-52, https://doi.org/10.1037/a0030641.

[167]

Bryk, A. et al. (2009), *Organizing Schools for Improvement*, University of Chicago Press, https://doi.org/10.7208/chicago/9780226078014.001.0001.

[25]

Buchmann, C. and B. Dalton (2002), "Interpersonal Influences and Educational Aspirations in 12 Countries: The Importance of Institutional Context", *Sociology of Education*, Vol. 75/2, p. 99, https://doi.org/10.2307/3090287.

[46]

Butler, J. and R. Adams (2007), "The Impact of Differential Investment of Student Effort on the Outcomes of International Studies, Journal of Applied Measurement", *Journal of Applied Measurement*, Vol. 8, pp. 279-304.

[55]

Cacioppo, J. and G. Berntson (1994), "Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates.", *Psychological Bulletin*, Vol. 115/3, pp. 401-423, https://doi.org/10.1037/0033-2909.115.3.401.

[173]

Callahan, R. (2005), "Tracking and High School English Learners: Limiting Opportunity to Learn", *American Educational Research Journal*, Vol. 42/2, pp. 305-328, https://doi.org/10.3102/00028312042002305.

[18]

Cantril, H. (1965), *The pattern of human concerns.*, New Brunswick, NJ: Rutgers University Press.

[41]

Carroll, J. (1963), *A model of school learning*, Teachers College Record.

[14]

Chernyshenko, O., M. Kankaraš and F. Drasgow (2018), "Social and emotional skills for student success and well-being: Conceptual framework for the OECD study on social and emotional skills*, OECD Education Working Papers*, No. 173, OECD Publishing, Paris, https://doi.org/10.1787/db1d8e59-en.

[209]

Citro, C. (1995), *Measuring Poverty*, National Academies Press, Washington, D.C., https://doi.org/10.17226/4759.

[36]

Cogan, L. and W. Schmidt (2014), "The Concept of Opportunity to Learn (OTL) in International Comparisons of Education", in *Assessing Mathematical Literacy*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-10121-7_10.

[24]

Cogan, L., W. Schmidt and S. Guo (2018), "The role that mathematics plays in college- and career-readiness: evidence from PISA", *Journal of Curriculum Studies*, Vol. 51/4, pp. 530-553, https://doi.org/10.1080/00220272.2018.1533998.

[92]

Comber, L. and J. Keeves (1973), *Science education in nineteen countries: International studies in evaluation.*, New York: John Wiley and Sons.

[186]

Cooper, H., J. Robinson and E. Patall (2006), "Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003", *Review of Educational Research*, Vol. 76/1, pp. 1-62, https://doi.org/10.3102/00346543076001001.

[113]

Couper, M. and F. Kreuter (2013), "Using paradata to explore item level response times in surveys", *Journal of the Royal Statistical Society. Series A: Statistics in Society*, Vol. 176/1, https://doi.org/10.1111/j.1467-985X.2012.01041.x.

[198]

Couper, M. and F. Kreuter (2012), "Using paradata to explore item level response times in surveys", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 176/1, pp. 271-286, https://doi.org/10.1111/j.1467-985x.2012.01041.x. [218]

Cowan, C. (2012), *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation.*, National Center for Education Statistics. [30]

Creemers, B. and L. Kyriakides (2007), *The Dynamics of Educational Effectiveness*, Routledge, https://doi.org/10.4324/9780203939185. [13]

Crothers, L. et al. (2010), "A Preliminary Study of Bully and Victim Behavior in Old-for-Grade Students: Another Potential Hidden Cost of Grade Retention or Delayed School Entry", *Journal of Applied School Psychology*, Vol. 26/4, pp. 327-338, https://doi.org/10.1080/15377903.2010.518843. [86]

Cunha, F. et al. (2006), "Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation", in *Handbook of the Economics of Education, Handbook of the Economics of Education Volume 1*, Elsevier, https://doi.org/10.1016/s1574-0692(06)01012-9. [44]

Darling-Hammond, L. et al. (2005), "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness.", *Education Policy Analysis Archives*, Vol. 13, p. 42, https://doi.org/10.14507/epaa.v13n42.2005. [121]

Debeer, D. et al. (2014), "Student, School, and Country Differences in Sustained Test-Taking Effort in the 2009 PISA Reading Assessment", *Journal of Educational and Behavioral Statistics*, Vol. 39/6, pp. 502-523, https://doi.org/10.3102/1076998614558485. [50]

DeLuca, C. et al. (2015), "Instructional Rounds as a professional learning model for systemic implementation of Assessment for Learning", *Assessment in Education: Principles, Policy and Practice*, Vol. 22/1, https://doi.org/10.1080/0969594X.2014.967168. [138]

DeLuca, C. et al. (2014), "Instructional Rounds as a professional learning model for systemic implementation of Assessment for Learning", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 122-139, https://doi.org/10.1080/0969594x.2014.967168. [214]

Demakakos, P. et al. (2008), "Socioeconomic status and health: The role of subjective social status", *Social Science & Medicine*, Vol. 67/2, pp. 330-340, https://doi.org/10.1016/j.socscimed.2008.03.038. [37]

Dienlin, T. and N. Johannes (2020), "The impact of digital technology use on adolescent well-being", *Dialogues in Clinical Neuroscience*, Vol. 22/2, pp. 135-142, https://doi.org/10.31887/dcns.2020.22.2/tdienlin. [78]

Dillman, D., J. Smyth and L. Christian (2014), *Internet, phone, mail, and mixed-mode surveys: The tailored design method (4th ed.).*, Hoboken, NJ: John Wiley. [176]

Duncan, G. and R. Murnane (2011), *Whither opportunity? Rising inequality, schools, and children's life chances.*, New York: Russell Sage Foundation. [20]

Eccles, J. et al. (1993), "Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families.", *American Psychologist*, Vol. 48/2, pp. 90-101, https://doi.org/10.1037/0003-066x.48.2.90. [61]

Edwards, J. and R. Bagozzi (2000), "On the nature and direction of relationships between constructs and measures.", *Psychological Methods*, Vol. 5/2, pp. 155-174, https://doi.org/10.1037/1082-989x.5.2.155. [182]

Efrati, V., C. Limongelli and F. Sciarrone (2014), "A Data Mining Approach to the Analysis of Students' Learning Styles in an e-Learning Community: A Case Study", in *Lecture Notes in Computer Science, Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-07440-5_27. [204]

Eklöf, H., B. Pavešič and L. Grønmo (2014), "A Cross-National Comparison of Reported Effort and Mathematics Performance in TIMSS Advanced", *Applied Measurement in Education*, Vol. 27/1, pp. 31-45, https://doi.org/10.1080/08957347.2013.853070. [51]

Epstein, J. (2001), *School, family, and community partnerships: Preparing educators, and improving schools*, Boulder, CO: Westview Press. [145]

Fan, X. and M. Chen (2001), "Parental involvement and students' academic achievement: A meta-analysis.", *Educational Psychology Review*, Vol. 13, pp. 1-22. [149]

Faubert, V. (2009), "School Evaluation: Current Practices in OECD Countries and a Literature Review"*, OECD Education Working Papers*, No. 42, OECD Publishing, Paris, https://doi.org/10.1787/218816547156. [131]

Fauth, B. et al. (2014), "Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes", *Learning and Instruction*, Vol. 29, pp. 1-9, https://doi.org/10.1016/j.learninstruc.2013.07.001. [103]

Fuligni, A. and H. Stevenson (1995), "Time Use and Mathematics Achievement among American, Chinese, and Japanese High School Students", *Child Development*, Vol. 66/3, pp. 830-842, https://doi.org/10.1111/j.1467-8624.1995.tb00908.x. [81]

Gehlbach, H. and A. Artino (2018), "The Survey Checklist (Manifesto)", *Academic Medicine*, Vol. 93/3, pp. 360-366, https://doi.org/10.1097/acm.0000000000002083. [177]

Gehlbach, H. and S. Barge (2012), "Anchoring and Adjusting in Questionnaire Responses", *Basic and Applied Social Psychology*, Vol. 34/5, pp. 417-433, https://doi.org/10.1080/01973533.2012.711691. [174]

Ghuman, S. and C. Lloyd (2010), "Teacher Absence as a Factor in Gender Inequalities in Access to Primary Schooling in Rural Pakistan", *Comparative Education Review*, Vol. 54/4, pp. 539-554, https://doi.org/10.1086/654832. [94]

Goodman, E. et al. (2001), "Adolescents' Perceptions of Social Status: Development and Evaluation of a New Indicator", *Pediatrics*, Vol. 108/2, pp. e31-e31, https://doi.org/10.1542/peds.108.2.e31. [38]

Greene, J. and M. Winters (2009), "The effects of exemptions to Florida's test-based promotion policy: Who is retained?", *Economics of Education Review*, Vol. 28/1, pp. 135-142, https://doi.org/10.1016/j.econedurev.2008.02.002. [88]

Griffith, C. et al. (2010), "Grade retention of students during grades K-8 predicts reading achievement and progress during secondary schooling", *Reading and Writing Quarterly*, Vol. 26/1, https://doi.org/10.1080/10573560903396967. [84]

Griffith, C. et al. (2009), "Grade Retention of Students During Grades K–8 Predicts Reading Achievement and Progress During Secondary Schooling", *Reading & Writing Quarterly*, Vol. 26/1, pp. 51-66, https://doi.org/10.1080/10573560903396967. [210]

Gurin, P. et al. (2004), "The educational value of diversity", *P. Gurin, J. S. Lehman, & E. Lewis (Eds.), Defending diversity: Affirmative action at the University of Michigan*, Vol. Ann Arbor: University of Michigan, pp. 97-188. [67]

Gurin, P. et al. (2002), "Diversity and Higher Education: Theory and Impact on Educational Outcomes", *Harvard Educational Review*, Vol. 72/3, pp. 330-367, https://doi.org/10.17763/haer.72.3.01151786u134n051. [68]

Hanushek, E. and L. Woessmann (2011), "The Economics of International Differences in Educational Achievement", in *Handbook of the Economics of Education*, Elsevier, https://doi.org/10.1016/b978-0-444-53429-3.00002-8. [27]

Harris, A. (2002), *School Improvement: What's In It For Schools?*, London: Routledge. [122]

Hattie, J. (2009), *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.*, London & New York: Routledge. [100]

Hattie, J. and H. Timperley (2007), "The Power of Feedback", *Review of Educational Research*, Vol. 77/1, pp. 81-112, https://doi.org/10.3102/003465430298487. [110]

Hayward, L. (2015), "Assessment is learning: the preposition vanishes", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 27-43, https://doi.org/10.1080/0969594x.2014.984656. [141]

Heckman, J., J. Stixrud and S. Urzua (2006), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior", *Journal of Labor Economics*, Vol. 24/3, pp. 411-482, https://doi.org/10.1086/504455. [10]

He, J. et al. (2017), "On Enhancing the Cross–Cultural Comparability of Likert–Scale Personality and Value Measures: A Comparison of Common Procedures", *European Journal of Personality*, Vol. 31/6, pp. 642-657, https://doi.org/10.1002/per.2132. [158]

Hershkovitz, A. and R. Nachmias (2009), "Learning about Online Learning Processes and Students' Motivation through Web Usage Mining", *Interdisciplinary Journal of e-Skills and Lifelong Learning*, Vol. 5, pp. 197-214, https://doi.org/10.28945/73. [201]

Hill, N. and D. Tyson (2009), "Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement.", *Developmental Psychology*, Vol. 45/3, pp. 740-763, https://doi.org/10.1037/a0015362. [144]

Hopfenbeck, T., M. Flórez Petour and A. Tolo (2015), "Balancing tensions in educational policy reforms: large-scale implementation of Assessment for Learning in Norway", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 44-60, https://doi.org/10.1080/0969594x.2014.996524. [139]

Hopfenbeck, T. and M. Kjærnsli (2016), "Students' test motivation in PISA: the case of Norway", *The Curriculum Journal*, Vol. 27/3, pp. 406-422, https://doi.org/10.1080/09585176.2016.1156004. [52]

Hospel, V. and B. Galand (2016), "Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis", *Learning and Instruction*, Vol. 41, pp. 1-10, https://doi.org/10.1016/j.learninstruc.2015.09.001. [107]

Howell, R., E. Breivik and J. Wilcox (2007), "Reconsidering formative measurement.", *Psychological Methods*, Vol. 12/2, pp. 205-218, https://doi.org/10.1037/1082-989x.12.2.205. [185]

Hoy, W., J. Hannum and M. Tschannen-Moran (1998), "Organizational Climate and Student Achievement: A Parsimonious and Longitudinal View", *Journal of School Leadership*, Vol. 8/4, pp. 336-359, https://doi.org/10.1177/105268469800800401. [60]

Jerrim, J. (2015), "Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent", *Oxford Review of Education*, Vol. 41/3, pp. 310-333, https://doi.org/10.1080/03054985.2015.1028525. [53]

Jeynes, W. (2007), "The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement", *Urban Education*, Vol. 42/1, pp. 82-110, https://doi.org/10.1177/0042085906293818. [150]

Jia, Y. et al. (2009), "The Influence of Student Perceptions of School Climate on Socioemotional and Academic Adjustment: A Comparison of Chinese and American Adolescents", *Child Development*, Vol. 80/5, pp. 1514-1530, https://doi.org/10.1111/j.1467-8624.2009.01348.x. [63]

Jonsson, A., C. Lundahl and A. Holmgren (2015), "Evaluating a large-scale implementation of Assessment for Learning in Sweden", *Assessment in Education: Principles, Policy and Practice*, Vol. 22/1, https://doi.org/10.1080/0969594X.2014.970612. [140]

Jonsson, A., C. Lundahl and A. Holmgren (2014), "Evaluating a large-scale implementation of Assessment for Learning in Sweden", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 104-121, https://doi.org/10.1080/0969594x.2014.970612. [215]

Jude, N. and S. Kuger (2018), *Questionnaire Development and Design for International Large-Scale Assessments (ILSAs): Current Practice, Challenges, and Recommendations.*. [205]

Kane, T. and S. Cantrell (2010), *Learning about teaching: Initial findings from the measures of effective teaching project*. [104]

Kaplan, D. and S. Kuger (2016), "The Methodology of PISA: Past, Present, and Future", in *Methodology of Educational Measurement and Assessment, Assessing Contexts of Learning*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-45357-6_3. [32]

Kaplan, D. and D. Su (2016), "On Matrix Sampling and Imputation of Context Questionnaires With Implications for the Generation of Plausible Values in Large-Scale Assessments", *Journal of Educational and Behavioral Statistics*, Vol. 41/1, pp. 57-80, https://doi.org/10.3102/1076998615622221. [195]

Kaplan, D. and D. Su (2014), *Imputation issues relevant to context questionnaire rotation. In Bertling, J. P. (Chair), Spiraling contextual questionnaires in educational large-scale assessments.*. [190]

Kearney, C. et al. (2019), "Reconciling Contemporary Approaches to School Attendance and School Absenteeism: Toward Promotion and Nimble Response, Global Policy Review and Implementation, and Future Adaptability (Part 1)", *Frontiers in Psychology*, Vol. 10, https://doi.org/10.3389/fpsyg.2019.02222. [45]

Keles, B., N. McCrae and A. Grealish (2019), "A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents", *International Journal of Adolescence and Youth*, Vol. 25/1, pp. 79-93, https://doi.org/10.1080/02673843.2019.1590851. [79]

Kim, S. and N. Hill (2015), "Including fathers in the picture: A meta-analysis of parental involvement and students' academic achievement.", *Journal of Educational Psychology*, Vol. 107/4, pp. 919-934, https://doi.org/10.1037/edu0000023. [146]

King, G. and J. Wand (2007), "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes", *Political Analysis*, Vol. 15/1, pp. 46-66, https://doi.org/10.1093/pan/mpl011. [163]

Klieme, E. (2014), *PISA 2015 draft questionnaire framework.*, OECD Publishing, Paris. [3]

Klieme, E., C. Pauli and K. Reusser (2009), "The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms", *T. Janik & T. Seidel (Eds.) The power of video studies in investigating teaching and learning in the classroom*, ,, pp. 137-160. [106]

Kloosterman, R. and P. de Graaf (2010), "Non-promotion or enrolment in a lower track? The influence of social background on choices in secondary education for three cohorts of Dutch pupils", *Oxford Review of Education*, Vol. 36/3, pp. 363-384, https://doi.org/10.1080/03054981003775244. [90]

Kutsyuruba, B., D. Klinger and A. Hussain (2015), "Relationships among school climate, school safety, and student achievement and well-being: a review of the literature", *Review of Education*, Vol. 3/2, pp. 103-135, https://doi.org/10.1002/rev3.3043. [56]

Kyllonen, P. (2013), "Innovative Questionnaire Assessment Methods to Increase Cross-Country Comparability", in *Handbook of International Large-Scale Assessment*, Chapman and Hall/CRC, https://doi.org/10.1201/b16061-17. [157]

Kyriakides, L. and B. Creemers (2008), "Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model", *School Effectiveness and School Improvement*, Vol. 19/2, pp. 183-205, https://doi.org/10.1080/09243450802047873. [111]

Lareau, A. and E. Weininger (2003), "Cultural capital in educational research: A critical assessment", *Theory and Society*, Vol. 32/5/6, pp. 567-606, https://doi.org/10.1023/b:ryso.0000004951.04408.b0. [23]

Lee, J. (2021), "Teacher–student relationships and academic achievement in Confucian educational countries/systems from PISA 2012 perspectives", *Educational Psychology*, Vol. 41/6, pp. 764-785, https://doi.org/10.1080/01443410.2021.1919864. [64]

Lee, J. (2009), "Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries", *Learning and Individual Differences*, Vol. 19/3, pp. 355-365, https://doi.org/10.1016/j.lindif.2008.10.009. [74]

Lee, J. and F. Borgonovi (2022), "Relationships between Family Socioeconomic Status and Mathematics Achievement in OECD and Non-OECD Countries", *Comparative Education Review*, Vol. 66/2, pp. 199-227, https://doi.org/10.1086/718930. [35]

Lee, J. and I. Paek (2014), "In Search of the Optimal Number of Response Categories in a Rating Scale", *Journal of Psychoeducational Assessment*, Vol. 32/7, pp. 663-673, https://doi.org/10.1177/0734282914522200. [179]

LEE, J. and V. SHUTE (2010), "Personal and Social-Contextual Factors in K–12 Academic Performance: An Integrative Perspective on Student Learning", *Educational Psychologist*, Vol. 45/3, pp. 185-202, https://doi.org/10.1080/00461520.2010.493471. [26]

Lee, J. and L. Stankov (2018), "Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA", *Learning and Individual Differences*, Vol. 65, pp. 50-64, https://doi.org/10.1016/j.lindif.2018.05.009. [71]

Lee, J. and L. Stankov (2013), "Higher-order structure of noncognitive constructs and prediction of PISA 2003 mathematics achievement", *Learning and Individual Differences*, Vol. 26, pp. 119-130, https://doi.org/10.1016/j.lindif.2013.05.004. [73]

Lee, J., Y. Zhang and L. Stankov (2019), "Predictive Validity of SES Measures for Student Achievement", *Educational Assessment*, Vol. 24/4, pp. 305-326, https://doi.org/10.1080/10627197.2019.1645590. [29]

Lemeshow, A. et al. (2008), "Subjective Social Status in the School and Change in Adiposity in Female Adolescents", *Archives of Pediatrics & Adolescent Medicine*, Vol. 162/1, p. 23, https://doi.org/10.1001/archpediatrics.2007.11. [39]

Levin, K. and C. Currie (2014), "Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples", *Social Indicators Research*, Vol. 119/2, https://doi.org/10.1007/s11205-013-0507-4. [42]

Levin, K. and C. Currie (2013), "Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples", *Social Indicators Research*, Vol. 119/2, pp. 1047-1063, https://doi.org/10.1007/s11205-013-0507-4. [208]

Lipowsky, F. et al. (2009), "Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem", *Learning and Instruction*, Vol. 19/6, pp. 527-537, https://doi.org/10.1016/j.learninstruc.2008.11.001. [109]

Little, I. and C. Bell (2009), *A practical guide to evaluating teacher effectiveness.*, Washington, DC: National Comprehensive Centre for Teacher Quality. [21]

Loukas, A. (2007), "What is school climate? High-quality school climate is advantageous for all students and may be particularly beneficial for at-risk students", *Leadership Compass*, Vol. 5, pp. 1-3. [57]

MacCallum, R. and M. Browne (1993), "The use of causal indicators in covariance structure models: Some practical issues.", *Psychological Bulletin*, Vol. 114/3, pp. 533-541, https://doi.org/10.1037/0033-2909.114.3.533. [183]

Mahoney, J. and R. Cairns (1997), "Do extracurricular activities protect against early school dropout?", *Developmental Psychology*, Vol. 33/2, pp. 241-253, https://doi.org/10.1037/0012-1649.33.2.241. [83]

Marsh, H. et al. (2019), "The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies.", *Journal of Educational Psychology*, Vol. 111/2, pp. 331-353, https://doi.org/10.1037/edu0000281. [72]

Marsh, H. et al. (2017), "Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores.", *Journal of Educational Psychology*, Vol. 109/3, pp. 425-438, https://doi.org/10.1037/edu0000144. [87]

Martin, M., I. Mullis and P. Foy (2008), *TIMSS 2007 international science report: Findings from IEA's Trends in International Mathematics and Science Study at the eighth and fourth grades.*. [96]

McDonnell, L. (1995), "Opportunity to Learn as a Research Concept and a Policy Instrument", *Educational Evaluation and Policy Analysis*, Vol. 17/3, pp. 305-322, https://doi.org/10.3102/01623737017003305. [19]

McGee, J. (2013), "<i>Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances.</i>by Greg J. Duncan and Richard J. Murnane (Eds.)", *Journal of School Choice*, Vol. 7/1, pp. 107-110, https://doi.org/10.1080/15582159.2013.759850. [207]

Milem, J., M. Chang and A. Antonio (2005), *Making diversity work on campus: A research-based perspective.*, Washington, DC: Association of American Colleges and Universities. [69]

Minor, E. et al. (2015), "A New Look at the Opportunity-to-Learn Gap across Race and Income", *American Journal of Education*, Vol. 121/2, pp. 241-269, https://doi.org/10.1086/679392. [22]

Monseur, C. and J. Bertling (2014), *Questionnaire rotation in international surveys: Findings from PISA. In Bertling, J. P. (Chair), Spiraling contextual questionnaires in educational large-scale assessments.*. [191]

Mullis, I., M. Martin and L. Jones (2015), "Third International Mathematics and Science Study (TIMSS)", in *Encyclopedia of Science Education*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2150-0_515. [212]

Murayama, K. et al. (2016), "Don't aim too high for your kids: Parental overaspiration undermines students' learning in mathematics.", *Journal of Personality and Social Psychology*, Vol. 111/5, pp. 766-779, https://doi.org/10.1037/pspp0000079. [147]

Naumann, J. (2015), "A model of online reading engagement: Linking engagement, navigation, and performance in digital reading", *Computers in Human Behavior*, Vol. 53, pp. 263-277, https://doi.org/10.1016/j.chb.2015.06.051. [199]

OECD (2020), "Schooling disrupted, schooling rethought: How the Covid-19 pandemic is changing education"*, OECD Policy Responses to Coronavirus (COVID-19)*, OECD Publishing, Paris, https://doi.org/10.1787/68b11faf-en. [155]

OECD (2019), *PISA 2021 Creative Thinking Framework (Third Draft).*, Paris: OECD Publishing. [151]

OECD (2018), *PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264305274-en. [6]

OECD (2017), *PISA 2015 technical report: Chapter 17 questionnaire design and computer-based questionnaire platform.*, Paris: OECD Publishing. [196]

OECD (2017), *Social and emotional skills: Well-being, connectedness and success.*, Paris: OECD Publishing. [77]

OECD (2013), "Background Questionnaires", in *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris, https://doi.org/10.1787/9789264190511-9-en. [2]

OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264190511-en. [4]

OECD (2013), "PISA 2012 Technical background", in *PISA 2012 Results: What Students Know and Can Do (Volume I): Student Performance in Mathematics, Reading and Science*, OECD Publishing, Paris, https://doi.org/10.1787/9789264201118-10-en. [187]

OECD (2011), *Quality Time for Students: Learning In and Out of School*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264087057-en. [95]

OECD (2010), *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264091450-en. [125]

OECD (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/9789264068780-en. [114]

OECD (2004), *Learning for Tomorrow's World: First Results from PISA 2003*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264006416-en. [115]

Opdenakker, M. and J. Van Damme (2000), "Effects of Schools, Teaching Staff and Classes on Achievement and Well-Being in Secondary Education: Similarities and Differences Between School Outcomes", *School Effectiveness and School Improvement*, Vol. 11/2, pp. 165-196, https://doi.org/10.1076/0924-3453(200006)11:2;1-q;ft165. [58]

Ou, S. and A. Reynolds (2010), "Grade Retention, Postsecondary Education, and Public Aid Receipt", *Educational Evaluation and Policy Analysis*, Vol. 32/1, pp. 118-139, https://doi.org/10.3102/0162373709354334. [85]

Ozga, J. (2012), "Assessing PISA", *European Educational Research Journal*, Vol. 11/2, pp. 166-171, https://doi.org/10.2304/eerj.2012.11.2.166. [124]

Papamitsiou, Z. and A. Economides (2017), "Exhibiting achievement behavior during computer-based testing: What temporal trace data and personality traits tell us?", *Computers in Human Behavior*, Vol. 75, pp. 423-438, https://doi.org/10.1016/j.chb.2017.05.036. [202]

Penk, C. (2015), *Effekte von Testteilnahmemotivation auf Testleistung im Kontext von Large-Scale-Assessments.*, Berlin: Humboldt-University. [54]

Pettigrew, T. and L. Tropp (2006), "A meta-analytic test of intergroup contact theory.", *Journal of Personality and Social Psychology*, Vol. 90/5, pp. 751-783, https://doi.org/10.1037/0022-3514.90.5.751. [70]

PISA Governing Board (2017), *Report from PISA 2021 background questionnaire strategic advisory group*. [5]

Primi, R. et al. (2019), "True or False? Keying Direction and Acquiescence Influence the Validity of Socio-Emotional Skills Items in Predicting High School Achievement", *International Journal of Testing*, Vol. 20/2, pp. 97-121, https://doi.org/10.1080/15305058.2019.1673398. [171]

Primi, R. et al. (2019), "Comparison of classical and modern methods for measuring and correcting for acquiescence", *British Journal of Mathematical and Statistical Psychology*, Vol. 72/3, pp. 447-465, https://doi.org/10.1111/bmsp.12168. [172]

Primi, R. et al. (2021), "SENNA Inventory for the Assessment of Social and Emotional Skills in Public School Students in Brazil: Measuring Both Identity and Self-Efficacy", *Frontiers in Psychology*, Vol. 12, https://doi.org/10.3389/fpsyg.2021.716639. [76]

Primi, R. et al. (2018), "Dealing with Person Differential Item Functioning in Social-Emotional Skill Assessment Using Anchoring Vignettes", in *Springer Proceedings in Mathematics & Statistics, Quantitative Psychology*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-77249-3_23. [161]

Purves, A. (1987), "The Evolution of the IEA: A Memoir", *Comparative Education Review*, Vol. 31/1, pp. 10-28, https://doi.org/10.1086/446653. [7]

Quon, E. and J. McGrath (2014), "Subjective socioeconomic status and adolescent health: A meta-analysis.", *Health Psychology*, Vol. 33/5, pp. 433-447, https://doi.org/10.1037/a0033716. [40]

Qureshi, F., J. Alegre and J. Bertling (2018), *Effect of contextual cue placement on achievement goals item responses, response time, and scalability. In New insights on survey questionnaire context effects from multiple large-scale assessments.*. [175]

Rankin-Erickson, J. and M. Pressley (2000), "A Survey of Instructional Practices of Special Education Teachers Nominated as Effective Teachers of Literacy", *Learning Disabilities Research and Practice*, Vol. 15/4, pp. 206-225, https://doi.org/10.1207/sldrp1504_5. [123]

Ratnam-Lim, C. and K. Tan (2015), "Large-scale implementation of formative assessment practices in an examination-oriented culture", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 61-78, https://doi.org/10.1080/0969594x.2014.1001319. [142]

Revilla, M., W. Saris and J. Krosnick (2014), "Choosing the Number of Categories in Agree-Disagree Scales", *Sociological Methods and Research*, Vol. 43/1, https://doi.org/10.1177/0049124113509605. [178]

Revilla, M., W. Saris and J. Krosnick (2013), "Choosing the Number of Categories in Agree–Disagree Scales", *Sociological Methods & Research*, Vol. 43/1, pp. 73-97, https://doi.org/10.1177/0049124113509605. [217]

Rosenkvist, M. (2010), "Using Student Test Results for Accountability and Improvement: A Literature Review", *OECD Education Working Papers*, No. 54, OECD Publishing, Paris, https://doi.org/10.1787/5km4htwzbv30-en. [128]

Rumberger, R. and G. Palardy (2005), "Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance", *American Educational Research Journal*, Vol. 42/1, pp. 3-42, https://doi.org/10.3102/00028312042001003. [59]

Rutkowski, D. and L. Rutkowski (2013), "Measuring Socioeconomic Background in PISA: One Size Might not Fit all", *Research in Comparative and International Education*, Vol. 8/3, pp. 259-278, https://doi.org/10.2304/rcie.2013.8.3.259. [31]

Ryan, K., M. Chandler and M. Samuels (2007), "WHAT SHOULD SCHOOL-BASED EVALUATION LOOK LIKE?", *Studies in Educational Evaluation*, Vol. 33/3-4, pp. 197-212, https://doi.org/10.1016/j.stueduc.2007.07.001. [134]

Rychen, D. and L. Salganik (2003), *Key competencies for a successful life and a well-functioning society*, Goettingen: Hogrefe & Huber. [11]

Sammons, P. (2009), "The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools", *School Effectiveness and School Improvement*, Vol. 20/1, pp. 123-129, https://doi.org/10.1080/09243450802664321. [206]

Sanders, J. and E. Davidson (2003), "A Model for School Evaluation", in *International Handbook of Educational Evaluation*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-010-0309-4_46. [132]

Santiago, P. and F. Benavides (2009), *Teacher evaluation: A conceptual framework and examples of country practices*, Paris: OECD Publishing. [135]

Scheerens, J. (2016), *Educational Effectiveness and Ineffectiveness*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-017-7459-8. [112]

Scheerens, J. (2002), "School self-evaluation: Origins, definition, approaches, methods and implementation", *Advances in Program Evaluation*, Vol. 8, https://doi.org/10.1016/s1474-7863(02)80006-0. [129]

Scheerens, J. (n.d.), "School self-evaluation: Origins, definition, approaches, methods and implementation", in *School-Based Evaluation: An International Perspective, Advances in Program Evaluation*, Emerald (MCB UP), Bingley, https://doi.org/10.1016/s1474-7863(02)80006-0. [213]

Scheerens, J. and R. Bosker (1997), *The foundations of educational effectiveness.*, Oxford: Pergamon Press. [12]

Scherff, L. and C. Piazza (2008), "Why Now, More Than Ever, We Need to Talk About Opportunity to Learn", *Journal of Adolescent & Adult Literacy*, Vol. 52/4, pp. 343-352, https://doi.org/10.1598/jaal.52.4.7. [93]

Schleicher, A. (2014), *Equity, Excellence and Inclusiveness in Education: Policy Lessons from Around the World*, International Summit on the Teaching Profession, OECD Publishing, Paris, https://doi.org/10.1787/9789264214033-en. [118]

Schmidt, W. (2009), "Opportunity to Learn", in *Handbook of Education Policy Research*, Routledge, https://doi.org/10.4324/9780203880968-55. [17]

Schmidt, W. (2001), "Why schools matter: a cross-national comparison of curriculum and learning", *Choice Reviews Online*, Vol. 39/09, pp. 39-5318-39-5318, https://doi.org/10.5860/choice.39-5318. [16]

Schmidt, W. (n.d.), "Opportunity to Learn?", in *Handbook of Education Policy Research*, Routledge, https://doi.org/10.4324/9780203880968.ch44. [211]

Schmidt, W. and N. Burroughs (2016), "The Trade-Off between Excellence and Equality: What International Assessments Tell Us", *Georgetown Journal of International Affairs*, Vol. 17/1, pp. 103-109, https://doi.org/10.1353/gia.2016.0011. [97]

Schmidt, W. et al. (2016), "The role of subject-matter content in teacher preparation: an international perspective for mathematics", *Journal of Curriculum Studies*, Vol. 49/2, pp. 111-131, https://doi.org/10.1080/00220272.2016.1153153.

[119]

Schmidt, W. et al. (2015), "The Role of Schooling in Perpetuating Educational Inequality", *Educational Researcher*, Vol. 44/7, pp. 371-386, https://doi.org/10.3102/0013189x15603982.

[98]

Schønning, V. et al. (2020), "Social Media Use and Mental Health and Well-Being Among Adolescents – A Scoping Review", *Frontiers in Psychology*, Vol. 11, https://doi.org/10.3389/fpsyg.2020.01949.

[80]

Sebastian, J., J. Moon and M. Cunningham (2017), "The relationship of school-based parental involvement with student achievement: a comparison of principal and parent survey reports from PISA 2012", *Educational Studies*, Vol. 43/2, https://doi.org/10.1080/03055698.2016.1248900.

[148]

Sebastian, J., J. Moon and M. Cunningham (2016), "The relationship of school-based parental involvement with student achievement: a comparison of principal and parent survey reports from PISA 2012", *Educational Studies*, Vol. 43/2, pp. 123-146, https://doi.org/10.1080/03055698.2016.1248900.

[216]

Seidel, T., R. Rimmele and M. Prenzel (2005), "Clarity and coherence of lesson goals as a scaffold for student learning", *Learning and Instruction*, Vol. 15/6, pp. 539-556, https://doi.org/10.1016/j.learninstruc.2005.08.004.

[108]

Slavin, R. and C. Lake (2008), "Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis", *Review of Educational Research*, Vol. 78/3, pp. 427-515, https://doi.org/10.3102/0034654308317473.

[101]

Snilstveit, B. (2016), *The impact of education programmes on learning and school participation in low- and middle-income countries*, International Initiative for Impact Evaluation (3ie), https://doi.org/10.23846/srs007.

[136]

Stankov, L., J. Lee and M. von Davier (2017), "A Note on Construct Validity of the Anchoring Method in PISA 2012", *Journal of Psychoeducational Assessment*, Vol. 36/7, pp. 709-724, https://doi.org/10.1177/0734282917702270.

[162]

Stark, S., O. Chernyshenko and F. Drasgow (2005), "An IRT Approach to Constructing and Scoring Pairwise Preference Items Involving Stimuli on Different Dimensions: The Multi-Unidimensional Pairwise-Preference Model", *Applied Psychological Measurement*, Vol. 29/3, pp. 184-203, https://doi.org/10.1177/0146621604273988.

[168]

Stevens, F. (1993), "Applying an Opportunity-to-Learn Conceptual Framework to the Investigation of the Effects of Teaching Practices via Secondary Analyses of Multiple- Case-Study Summary Data", *The Journal of Negro Education*, Vol. 62/3, p. 232, https://doi.org/10.2307/2295463.

[99]

Tang, J. (2017), *The development and applications of alternative student socioeconomic status measures*.

[33]

UN (2015), "Transforming governance for the 2030 agenda for sustainable development", in *World Public Sector Report 2015: Responsive and Accountable Governance*, United Nations, New York, https://doi.org/10.18356/e5a72957-en.

[1]

UNESCO Institute for Statistics/OECD/Eurostat (2015), *ISCED 2011 Operational Manual : Guidelines for classifying national education programmes and related qualifications*, UNESCO Institute for Statistics/OECD/Eurostat, https://doi.org/10.15220/978-92-9189-174-0-en. [34]

van Tartwijk, J. and K. Hammerness (2011), "The neglected role of classroom management in teacher education", *Teaching Education*, Vol. 22/2, pp. 109-112, https://doi.org/10.1080/10476210.2011.567836. [105]

Vieluf, S. et al. (2012), *Teaching Practices and Pedagogical Innovations: Evidence from TALIS*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/9789264123540-en. [116]

von Davier, M. (2013), "Imputing Prociency Data under Planned Missingness in Population Models", in *Handbook of International Large-Scale Assessment*, Chapman and Hall/CRC, https://doi.org/10.1201/b16061-13. [192]

von Davier, M. et al. (2017), "The Effects of Vignette Scoring on Reliability and Validity of Self-Reports", *Applied Psychological Measurement*, Vol. 42/4, pp. 291-306, https://doi.org/10.1177/0146621617730389. [164]

Wang, M., G. Haertel and H. Walberg (1993), "Toward a Knowledge Base for School Learning", *Review of Educational Research*, Vol. 63/3, pp. 249-294, https://doi.org/10.3102/00346543063003249. [102]

Washington, D. (ed.) (2016), *Learning data for better policy: A global agenda. CGD policy paper, 2.*. [130]

Wike, R. (2016), *Europeans fear wave of refugees will mean more terrorism, fewer jobs.*, . Pew Research Center, 11, 2016. [49]

Willms, J. (2006), *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems.*, Montreal: UNESCO Institute for Statistics. [28]

Wylie, E. and C. Lyon (2015), "The fidelity of formative assessment implementation: issues of breadth and quality", *Assessment in Education: Principles, Policy & Practice*, Vol. 22/1, pp. 140-160, https://doi.org/10.1080/0969594x.2014.990416. [143]

Yan, T. and R. Tourangeau (2008), "Fast times and easy questions: The effects of age, experience and question complexity on web survey response times", *Applied Cognitive Psychology*, Vol. 22/1, https://doi.org/10.1002/acp.1331. [220]

Yan, T. and R. Tourangeau (2007), "Fast times and easy questions: the effects of age, experience and question complexity on web survey response times", *Applied Cognitive Psychology*, Vol. 22/1, pp. 51-68, https://doi.org/10.1002/acp.1331. [200]

Ziegler, M., C. Kemper and B. Rammstedt (2013), "The Vocabulary and Overclaiming Test (VOC-T)", *Journal of Individual Differences*, Vol. 34/1, pp. 32-40, https://doi.org/10.1027/1614-0001/a000093. [170]
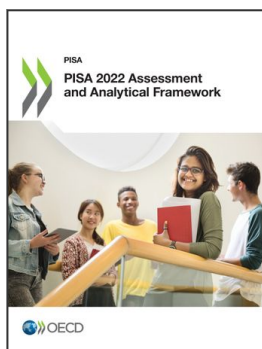
From:

PISA 2022 Assessment and Analytical Framework

Access the complete publication at:
https://doi.org/10.1787/dfe0bf9c-en