OECD SURVEY ON SOCIAL AND EMOTIONAL SKILLS

TECHNICAL REPORT

OECD Survey on Social and Emotional Skills

Technical Report

Table of contents

FIGURES

TABLES

Acronyms and Abbreviations

ACE	Australian Council for Educational Research
ARS	acquiescence response sets
AM	RM Assessment Master
ASDISM	Australian Signals Directorate Information Security Manual
AWS	Amazon Web Services
BIC	Bayesian Information Criterion
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
COB	country of birth
EFA	exploratory factor analysis
ENR	enrolment
FT	Field Test
GPCM	Generalised Partial Credit Model
HOMEPOS	home possessions index
ID	identification
IRT	Item Response Theory
ISEI	international socio-economic index of occupational status
ISO	International Organization for Standardization
IT	information technology
ISCED	International Standard Classification of Education
JRR	Jackknife repeated replication
MGCFA	multiple-group confirmatory factor analysis
MIMIC	multiple indicator multiple cause
ML	maximum likelihood
MLR	robust maximum likelihood
MML	marginal maximum likelihood
MOS	Measure of Size
MS	Main Study
OECD	Organisation for Economic Co-operation and Development
PA	Parent Questionnaire data file
PAREDYRS	years of formal parental education index
PCM	Partial Credit Model
PPS	probability proportional to size
PISA	*Programme for International Student Assessment*
PrQ	Principal Questionnaire data file
PSU	primary sampling unit
QM	Quality Monitor
RMSEA	Root Mean Square Error of Approximation
SAVF	Site Adaptation Verification Form
SEM	structural equation modelling
SEN	Special Education Needs
SES	socio-economic status
SA	Study Administrator
SC	School Coordinator
SGF	Student Grades Form
SPC	Site Project Centre
SPM	Site Project Manager
SPSS®	Statistical Package for the Social Sciences

SRMR	Standardised Root Mean Square Residual
SSES	Survey on Social and Emotional Skills
ST	Student Questionnaire data file
STA	Student Direct Assessment data item label
STF	Student Tracking Form
STLF	Student Teacher Linkage Form
STLC	Student Teacher Linkage Correspondence [form]
TC	Teacher Assessment Questionnaire data file
TcQ	Teacher Contextual Questionnaire data file
TCS	Target Cluster Size
TLI	Tucker-Lewis index
TMS	Translation Management Sheet
TRT	Technical Readiness Test
UAT	User Acceptance Test/Testing
URL	Uniform Resource Locator [web address]
WLE	weighted likelihood estimate
WLS	weighted least squares
WLSMV	weighted least squares mean and variance adjusted

Scale Labels

ASS	assertiveness
COO	cooperation
CRE	creativity
CUR	curiosity
EFF	self-efficacy
EMO	emotional control
EMP	empathy
ENE	energy
MOT	achievement motivation
OPT	optimism
PER	perseverance/persistence
RES	responsibility
SEL	self-control
SOC	sociability
STR	stress resistance/resilience
TOL	tolerance
TRU	trust

Participating Sites

BOG	Bogotá, Capital District (Colombia)
DAE	Daegu (Republic of Korea)
HEL	City of Helsinki (Finland)
HOU	Houston Independent School District (United States)
MAN	Manizales, Caldas (Colombia)
MOS	Moscow (Russian Federation)
OTT	Ottawa (Ontario, Canada)
SIN	Sintra (Portugal)
SUZ	Suzhou (Jiangsu Province, China)
IST	Istanbul (Turkey)

Chapter 1. Overview

This chapter provides an overview of the survey components and the preparation and implementation of the OECD Survey on Social and Emotional Skills 2019 (SSES), initiated and coordinated by the Organisation for Economic Co-operation and Development (OECD).

1.1. Introduction

The OECD Survey on Social and Emotional Skills (SSES) is an initiative which seeks further understanding of the characteristics and contextual factors that influence the formation and growth of students' social and emotional skills. The survey takes a snapshot of students' social and emotional skills and asesses the key factors that are thought to foster or hinder the development of these skills in schools and other settings.

The collaborative effort of the cities who participated in these initial asessments enabled a comprehensive set of information to be collected on these skills, both directly and indirectly, together with data about students' family, school and community learning contexts.

SSES is an age-based survey, asesssing two age cohorts – 10-year-old students in school in grade 2 or higher and 15-year-old students in school in grade 7 or higher. The younger cohort provides information on how students are developing in school and what may be needed in future school years. Students in the older cohort are approaching the end of compulsory schooling in most participating cities, so they provide information from the perspective of the output of schooling.

The first cycle of SSES commenced in 2016. After an extensive literature review, cognitive interviews were conducted with 10-year-old and 11- year-old students in the United States in December 2017, followed by an online survey in February 2018. In April and May 2018, item trials were conducted in six participating cities in two age cohorts. The SSES Field Test was implemented in late 2018 and the Main Study data collection was completed in late 2019.

Ten cities from nine countries participated in the SSES Main Study data collection:

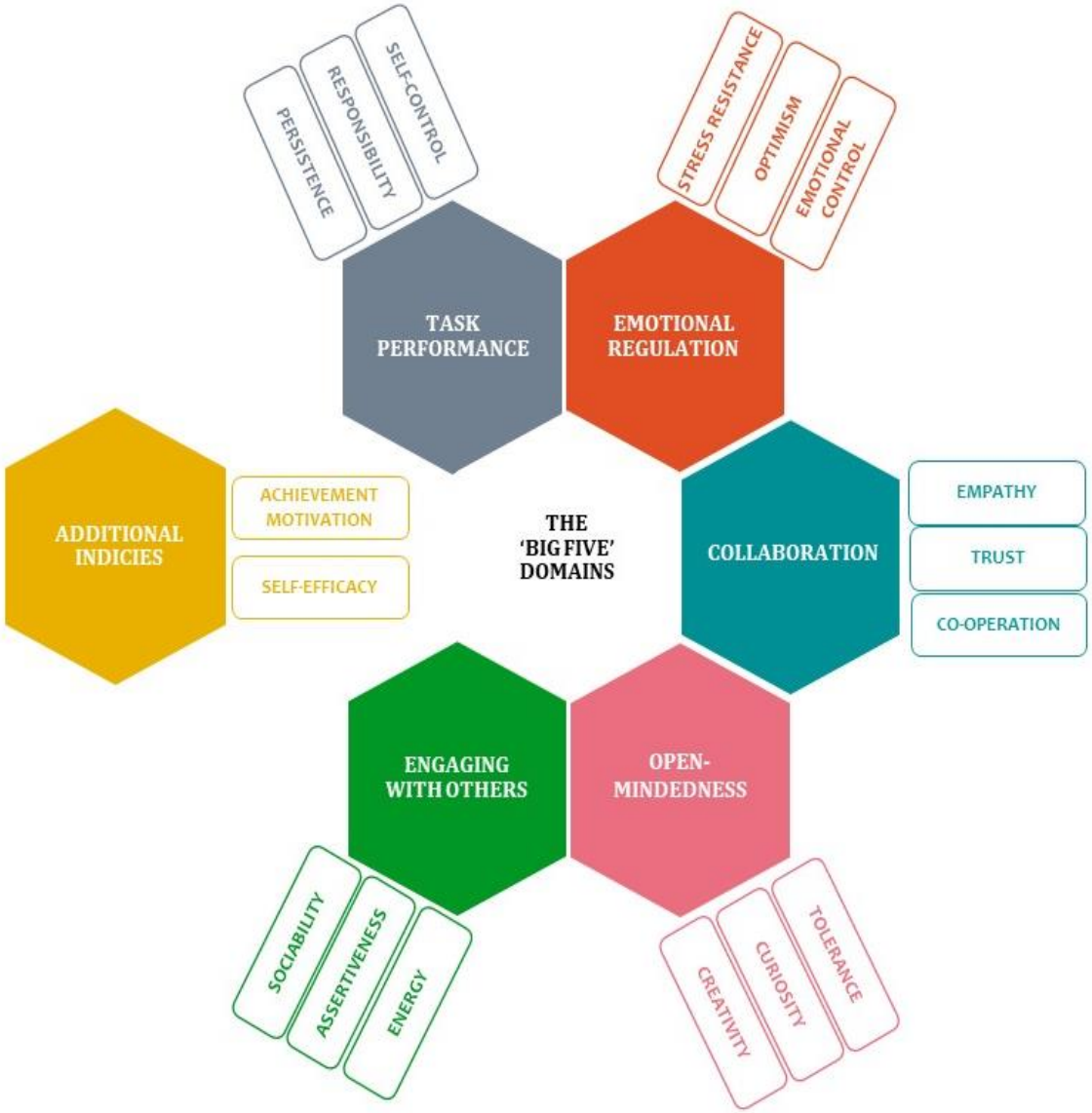
- Bogota, Colombia
- Daegu, Korea
- Helsinki, Finland
- Houston, Texas, United States
- Istanbul, Turkey
- Manizales, Colombia
- Moscow, Russian Federation
- Ottawa, Ontario, Canada
- Sintra, Portugal
- Suzhou, People's Republic of China

The Field Test collected information on 19 selected social and emotional skills of students. From this initial skill set, 15 social and emotional skills were selected for the assessment in the Main Study, from the Big Five model domains as shown in the survey's assessment framework.

- task performance – self-control, responsibility, persistence
- emotional regulation – stress resistance, optimism, emotional control
- collaboration – empathy, trust, co-operation
- open-mindedness – tolerance, curiosity, creativity
- engaging with others – sociability, assertiveness, energy

As well as these 15 skills, two additional indices (achievement motivation and self-efficacy) were created from a selection of items belonging to other skills.

Figure 1.1. The 'Big Five' domains



SSES gathers information about students' social and emotional skills through direct assessment of students via self-reports, and indirect assessment via parent and teacher reports of the student.

Collecting contextual information is critical for understanding how students' social and emotional skills have developed and how these skills may be improved. Students learn in many different settings, including within their families, schools and communities, with each context playing an important role throughout childhood and adolescence. Contextual information offers insight into the factors that help and those that hinder social and emotional skills development, including the policies and practices that support them.

SSES gathers this contextual information from students, and also from parents, teachers and school principals, through the background questionnaires. These questionnaires ask

about a broad range of contextual variables, including student socio-demographic background, and the family, school, peer and wider community environments.

SSES draws on this rich data to compare the predictive value of different skills and factors in students' environments relative to their social and emotional skills. The survey aims to:

- Provide an overview of the social and emotional skills of students.
- Deepen the understanding of the relationships between these skills factors in the individual, family, peer and school environment, as well as relationships between these skills and social and economic outcomes of students such as their academic performance, health and well-being.
- Demonstrate that valid, reliable, and comparable datasets on social and emotional skills can be produced across diverse student populations and settings.

1.2. Key Features of SSES

This report details the technical aspects underpinning the development, methodology and initial administration of the SSES survey, including:

- the design of the direct and indirect student assessment instruments developed specifically for SSES
- the requirements and procedures for the sampling of schools and students
- the measures undertaken to ensure equivalence between the different language versions of the survey instruments used in participating cities, while taking into account the diverse cultural contexts of those cities
- the technical requirements and procedures for administering online assessments in schools
- various operational procedures, including test administration arrangements, data collection and processing, quality control and monitoring measures developed to ensure comparable data is collected from all participating cities
- data scaling and analysis, and subsequent reporting
- quality assurance processes and procedures, including the quality monitoring school visits that enable SSES to provide high quality data outputs.

1.3. Managing and Implementing SSES

SSES is implemented collaboratively by Site Project Managers (SPM) in participating cities, the International Contractor and the OECD Secretariat.

The SPMs are responsible for ensuring that SSES is implemented in their city or region in compliance with the internationally agreed technical and administrative standards and procedures.

During the Field Test and Main Study administrations, the SPMs played a vital role in developing and validating the assessment instruments and ensured that project implementation was of high quality. Annex H lists the SPMs who participated in the 2019 administrations.

The OECD Secretariat is responsible for the overall management of SSES. It has continuously monitored development, worked closely with the participating cities to ensure

their priorities and interests are reflected in the design and implementation of the assessment, and facilitated collaboration among all parties.

Ohio State University won the original tender for SSES and, together with the OECD Secretariat, developed the SSES survey instruments. The overall management and implementation of the Field Test and Main Study was carried out by the Australian Council for Educational Research (ACER). The OECD Secretariat worked closely with the International Project Director, Dr Sue Thomson (ACER), to coordinate all aspects of implementation.

Chapter 2. Direct and Indirect Social and Emotional Skills Items Design and Development

This chapter describes the survey design for the Survey on Social and Emotional Skills (SESS), and the process that the OECD Secretariat, the international contractors[1], and the technical advisory group used to develop the direct (student) and indirect (parent and teacher) social and emotional skills items.

2.1. Survey design

The survey is designed to take a snapshot of the social and emotional skills of two student age groups: ages 10 (the younger cohort) and 15 years (the older cohort). The survey's assessment framework describes the design of the survey, which is based on a well-researched framework – the 'Big Five' – that distinguishes five dimensions of social and emotional skills: task performance, emotional regulation, open-mindedness, collaboration, and engaging with others (John and De Fruyt, 2015[1]; Abrahams et al., 2019[2]; Lipnevich, Preckel and Roberts, 2017[3]). Each of these five broad domains represents a cluster of related behaviours or thoughts and can be divided further into narrower subdomains of social and emotional skills, which are more descriptive, specific and accurate, and thus easier to measure. The survey's assessment framework (2019[4]) provides more information on the 'Big Five'.

Table 2.1. 'Big Five' domains and underlying social and emotional skills

Big Five domain	Skill	Example items
Task performance	self-control	I avoid mistakes by working carefully I stop to think before acting
	responsibility	I keep promises I avoid responsibilities
	persistence	I keep working on a task until it is finished I finish things despite difficulties in the way
Emotional regulation	stress resistance	I am relaxed and handle stress well I do not panic easily
	emotional control	I stay calm even in tense situations I am not easily upset
	optimism	I believe good things will happen to me I enjoy life
Engaging with others	energy	I maintain high energy throughout the day I show a lot of enthusiasm
	assertiveness	I like being a leader in my class I know how to convince others to do what I want
	sociability	I like to be with friends I like talking to a lot of different people
Collaboration	empathy	I understand what others want It is important that my friends are ok
	trust	I believe that most people are honest I trust others
	co-operation	I get along well with others I am always willing to help classmates
Open-mindedness	tolerance	I like hearing about other cultures and religions I learn a lot from people with differing beliefs
	curiosity	I love learning new things in school I like to know how things work
	creativity	I sometimes find a solution that other people don't see I have a good imagination

The survey measures students' self-reported perceptions of the extent to which they have these skills – 15 skills in total, as shown in Table 2.1. In addition, two indices, self-efficacy and achievement motivation are included, created from items that evaluate other skills in the survey.

The survey's conceptual framework (2018[5]) and assessment framework (2019[4]), available on the survey website, provide extensive discussion on why these skills were chosen for inclusion in the survey. In summary, the skills:

- Provide a broad and balanced coverage of the domains of social and emotional skills.
- Are appropriate for 10- and 15-year-olds.
- Are predictive of success in a wide range of important life outcomes and events. Numerous studies and meta-analyses have found the Big Five dimensions to be associated with academic achievement, health and well-being, job performance and occupational attainment.
- Are malleable and susceptible to possible policy interventions. Extensive research shows that children are not born with a fixed set of skills and little room for improvement, but instead have considerable potential to develop social and emotional skills which are influenced by their experiences and their environment throughout life. This is especially the case during adolescence.
- Are comparable and relevant across different cultures, languages, social and school contexts. There is extensive evidence that the Big Five dimensions and their subdomains are conceptually comparable across cultures; countries and economies for both adults and children from different cultural backgrounds. However, there is also evidence that a simple comparison of scale scores across cultures may not be possible due to cultural differences in the interpretation of questions or problems with translation of items into different languages. In order to improve cross-cultural comparability, the OECD has worked with leading experts in the field to develop collectively comprehensive methodological, translation and statistical procedures that minimize biases.

2.1.1. Triangulation

Another aspect of the survey's design is the use of three sources of information to measure students' social and emotional skills – the students themselves, and reports from parents and teachers. The survey assesses students' skills from different respondents for several reasons. First, the two indirect surveys from parents and teachers can improve the content validity of the instruments used to assess students' social and emotional skills by providing information on students' behaviours across different contexts. As students may behave differently in different settings, collecting information on students' behaviours from multiple sources and across multiple contexts may improve the representation and understanding of students' behaviours in the most important contexts for school-age students (at home and at school). In addition, information from parents and teachers may allow for controlling for measurement error in self-reports, such as social desirability and unrealistic self-perceptions.

Parents and teachers are valuable sources of information on students' behaviours. Parents have long-term and close relationships with children: they have first-hand knowledge of their children's life situation, preferences and practices, and have been able to observe their children across a wide variety of contexts. Teachers can provide a reasonably objective survey of students' social and emotional skills in a school context. Teachers' reports can be especially valuable because they have experience in assessing 10- and 15-year-old students.

2.1.2. Types of assessment instruments

In order for social and emotional skills items to be comparable across different types of respondents (students, parents and teachers), they must focus on observable behaviours. Item wording, syntax and semantics were kept as simple as possible to minimise the cognitive burden across the three responding groups, maximise the comparability of the assessments, and ease the burden of translation.

The survey items ask students to report on their own behaviour, thoughts and feelings. The items are in the form of simple statements such as "I like learning new things" (item assessing students' creativity) and "I stay calm even in tense situations" (item assessing stress resistance). While students provide a report on themselves and one parent reports on every student, teachers were often reporting on multiple students. Therefore, the number of items per skill was reduced for teachers in order to lower their response burden. The skill scales for students and parents contain 8 items, the skill scales for teachers contain 3 items.

The survey items have a 5-point Likert type agree/disagree response scale, with answers ranging from 1 – completely disagree to 5 – completely agree. All of the 15 skill scales use positively and negatively worded items.

2.1.3. Development of assessment scales

Developing valid, reliable and comparable assessment scales for each of the 15 skills was a long process, and included a multitude of rounds of empirical testing in various formats (both qualitative and quantitative). The goal was to develop scales for each of the skills that consisted of items that were not only simple, clear and at an appropriate reading level, but that measure the same construct.

The International Contractor and the OECD Secretariat first undertook a literature review, identifying 31 subdomains of the Big Five that were relevant to policy. These subdomains were then ranked, from low to moderate to high, based on their relevance for the following criteria: predictive validity for education, economic, health and quality of life; relevance for children; cross-cultural comparability; malleability; theoretical importance; and relevance for future trends (see Table 2.2).

Table 2.2. Evaluation of 'Big Five' subdomains

Big Five domains	Social and emotional skills	Predictive validity - education	Predictive validity - economic	Predictive validity - health	Predictive validity - quality of life	Relevance for children	Cross-cultural comparability	Malleability	Theoretical importance	Relevance for future trends
Task performance (conscientiousness)	**Achievement motivation**	moderate	moderate	low	high	moderate	moderate	high	high	moderate
	Orderliness	none	low	low	moderate	high	high	moderate	moderate	low
	Self-control/self-discipline	moderate	low	moderate	n/a	high	moderate	moderate	high	moderate
	Responsibility/ trustworthiness	low	high	moderate	high	moderate	moderate	moderate	high	moderate
	Persistence	moderate	n/a	n/a	high	moderate	moderate	moderate	moderate	moderate
	Discipline/rule following	n/a	n/a	moderate	n/a	moderate	moderate	moderate	moderate	moderate
	Honesty/virtue	n/a	n/a	moderate	n/a	low	low	high	moderate	moderate
Emotional regulation (Neuroticism)	**Stress resistance/anxiety**	moderate	low	moderate	high	high	high	moderate	high	high
	Emotional control/volatility	moderate	low	moderate	high	high	high	moderate	high	high
	Optimism/positive emotion	moderate	moderate	high	high	high	moderate	moderate	high	moderate
	Angry hostility/irritability	n/a	moderate	n/a	n/a	moderate	high	moderate	moderate	high
	Vulnerability	n/a	none	n/a	n/a	moderate	moderate	moderate	moderate	moderate
	Self-compassion	n/a	n/a	n/a	n/a	moderate	moderate	moderate	low	moderate
Engaging with others (Extraversion)	**Energy/enthusiasm**	none	low	moderate	high	high	high	low	high	moderate
	Assertiveness/dominance	low	low	low	moderate	moderate	moderate	moderate	high	moderate
	Sociability	none	low	low	moderate	high	high	moderate	high	moderate
	Attention seeking	none	none	n/a	n/a	moderate	low	moderate	moderate	low
Collaboration (Agreeableness)	**Empathy/compassion**	none	none	none	moderate	high	high	moderate	high	moderate
	Altruism/generosity	none	none	low	moderate	moderate	high	moderate	moderate	moderate
	Cooperation	n/a	low	n/a	n/a	high	moderate	moderate	moderate	moderate
	Trust	none	low	low	moderate	moderate	low	moderate	high	high
	Respectfulness	n/a	n/a	n/a	n/a	high	moderate	moderate	moderate	moderate
	Modesty	n/a	none	n/a	n/a	low	low	moderate	moderate	moderate

Open-mindedness (Openness to experience)	**Intellectual curiosity**	moderate	low	none	low	moderate	high	moderate	high	moderate
	Intellectual efficiency	n/a	n/a	n/a	n/a	moderate	high	moderate	moderate	high
	Creativity/imagination	moderate	low	low	moderate	moderate	moderate	moderate	high	high
	Aesthetic interests	none	none	none	low	low	low	moderate	moderate	moderate
	Tolerance/cultural flexibility	moderate	low	n/a	n/a	moderate	low	moderate	moderate	high
Compound skills	**Independence/critical thinking**	n/a	n/a	n/a	n/a	moderate	moderate	moderate	high	high
	Meta-cognition/self-reflection	low	none	n/a	n/a	moderate	moderate	moderate	high	high
	Self-efficacy	moderate	moderate	high	high	moderate	moderate	high	high	moderate

Note: The facets in bold were the social and emotional skills that went on to be tested during the item trials and field test.

Based on the ranking by the international contractor and the OECD, and feedback from the participating cities and the technical advisory group experts, 19 skills were chosen to be tested in the field test, 16 skills under each of the 'Big Five' domains, and 3 compound skills.

Reviewing item selection

For each of the 19 skills, the OECD Secretariat and the international contractor compiled an initial item pool of 20 items. A large number of items were selected from the International Personality Item Pool (IPIP) (Goldberg and Saucier, 2019[6]). The Hierarchical personality inventor for children (HiPIC) (Mervielde and De Fruyt, 1999[7]), the Inventory of Child Individual Differences (ICID) (Halverson et al., 2003[8]), and the Big Five Inventor-2 (BFI-2) (Soto and John, 2017[9]) databases were also reviewed as well as other existing social and emotional skills scales. Moreover, when necessary, new items were created using the same format as items from existing scales by the international contractor. The items went through six rounds of revisions and were given a ranking by the International Contractor, the technical advisory group and the OECD Secretariat based on reading level, idioms involved, cross-cultural comparability, sensitive topic/wording, and divergent and content validity issues. The technical advisory group reviewed the item bank, and based on their feedback, it was reduced to 15 items per skill for the 15-year-old student cohort and to 12 items for the 10-year-old student cohort. The items then went through a series of empirical rounds of testing in order to reduce the pool even further so that the most consistently reliable items remained.

Cognitive interviews

In December 2017, 37 interviews were conducted with children between ages 10 and 11 in the United States. The scales had 12 items and were in English. The interviews provided insights as to the difficulty of the items' wording, and students' different interpretations of the items. Based on results from these cognitive interviews and feedback from members of the technical advisory group, some items were reformulated, and two items per scale for the younger cohort were dropped to reduce their response burden.

Online survey

In February 2018, the OECD Secretariat and the first international contractor conducted an online survey with 1 000 parents from the United States, in order to assess the social and emotional skills survey items. This was a convenience sample as the first international contractor was based in the United States and the source version was in English. Ten items per skill were tested – 190 items in total. The online survey provided an opportunity for the OECD Secretariat to obtain initial information about indirect survey items, in particular the measurement properties of the 19 skill scales based on parent reports, and including their predictive validity with a range of outcomes.

The OECD Secretariat used an online service platform to connect to and invite participants to contribute to this research. The survey was available to parents who had a child between the ages of 9 and 16. The survey was also open to teachers. It took parents about 25 minutes to respond to 250 questions in all, 60 of which were background questionnaire items.

Item trials

The item trials moved ahead with all 19 skill domains and 15 items per skill for the older cohort, and 10 items per skill for the younger cohort.

In April and May 2018, six sites participated in the item trials:

- Daegu, South Korea
- Manizales, Colombia
- Moscow, Russia
- Ottawa, Canada
- Rome, Italy
- Turkey.

This empirical round of testing provided insights into the reliability, validity and cross-cultural comparability of the instruments in their early stages. The item trials also served as practice for the translation process, providing initial translations of the assessment items. It also allowed for the international contractor and the OECD to have an initial estimate of the response time and the testing burden. In addition, four of the sites had parents and teachers evaluate their assessment scales through cognitive interviews. The parents provided information on how they understood the questions, if the questions were well-formulated and appropriate, and on the time needed to complete these scales.

Field test

The focus of the field test was two-fold – to examine the performance of the survey instruments and the adequacy and effectiveness of the survey methodologies/procedures. The field test represented the last phase of the instrument development process. Based on data from the trials and feedback from the technical advisory group, the most reliable ten items for the older cohort and eight items for the younger cohort were selected to move into the field test.

The field test administration period began on 1 October and finished on 30 November 2018 with all ten participating cities that moved on to the main study the following year. The sample size included 500 students in each of the two cohorts, as well as their parents, teachers and school principals.

The field test results were reviewed to ascertain whether the survey instruments were able to capture the intended information in a reliable and valid manner. Some of the same criteria used to select the initial set of 19 skills, as outlined in Table 2.2 (e.g. predictive validity with life outcomes, scale reliability, unique contribution, cross-cultural comparability) were also used to reduce the number of items leading up to the main study (see Table 2.3 below).

Table 2.3. Final selection of social and emotional skills for the main study

Big Five dimension	Social and emotional skill subdomains	Predictive validity - education	Predictive validity - conduct	Predictive validity - health	Predictive validity – well-being	Scale reliability	Unique contribution	Cross-cultural comparability
Task performance (conscientiousness)	Achievement motivation	moderate	moderate	moderate	high	moderate	Low	moderate
	Self-control	low	moderate	low	moderate	moderate	moderate	high
	Responsibility	moderate	moderate	low	moderate	moderate	moderate	high
	Persistence	moderate	moderate	moderate	moderate	high	low	high
Emotional regulation (Neuroticism)	**Stress resistance**	low	low	high	high	moderate	moderate	high
	Emotional control	low	low	high	high	moderate	high	high
	Optimism	low	moderate	moderate	high	high	high	moderate
Engaging with others (Extraversion)	**Energy**	low	moderate	moderate	high	moderate	moderate	moderate
	Assertiveness	low	low	low	low	high	moderate	moderate
	Sociability	low	low	low	high	moderate	high	high
Collaboration (Agreeableness)	**Empathy**	low	moderate	low	moderate	moderate	low	high
	Cooperation	low	moderate	low	high	moderate	low	high
	Trust	low	moderate	low	high	high	high	high
Open-mindedness (Openness to experience)	**Curiosity**	moderate	high	low	moderate	high	low	high
	Creativity	moderate	moderate	low	moderate	high	moderate	high
	Tolerance	low	high	low	moderate	high	high	high
Compound skills	Critical thinking	low	moderate	low	low	low	moderate	high
	Meta-cognition/self-reflection	low	moderate	low	low	low	moderate	high
	Self-efficacy	low	moderate	low	high	high	low	high

Note: The social and emotional skill subdomains in bold were chosen to be included in the main study.

Based on the field trial results and feedback from the technical advisory group, the number of social and emotional skills was reduced from 19 to 15 for both cohorts. The scales that were removed did not meet the psychometric standards needed for inclusion. It was also decided that two additional indices would be created for achievement motivation and self-efficacy. Specific items belonging to these scales were removed, however, the skills were kept as additional indices created from items in other skill scales.

In addition, the technical advisory group recognised the need to reduce the overall response burden for the younger cohort. However, they thought it important to assess the same skills in both cohorts and therefore suggested reducing the contextual questionnaire for the younger cohort. As a result, the number of skill items was reduced leading up to the main study, from 10 to 8 items per skill for the older cohort, while it remained the same (8 items per skill) for the younger cohort. The skill scales for both cohorts for the main study included 120 items.

The instrument development process for the self-report scales, including the number of items in each of the phases is outlined in Table 2.4.

Table 2.4. Development of self-report scales

Study stage	Timeline	Items per skill		Number of skills		Total number of items	
		Older cohort	Younger cohort	Older cohort	Younger cohort	**Older cohort**	**Younger cohort**
Initial item pool compiled	Nov 2017	20	20	19	19	**380**	**380**
Item pool reduced after feedback from TAG	Dec 2017	15	12	19	19	**285**	**228**
Cognitive interviews	Dec 2017		12	19	19		**228**
Item Trials	Apr 2018	15	10	19	19	**285**	**190**
Field Test	Oct/Nov 2018	10	8	19	19	**190**	**152**
Main Study	**Oct/Nov 2019**	**8**	**8**	**15**	**15**	**120**	**120**

Parent and teacher skill scales were also administered during the field test. Eight items were included per scale for parents' reports and three items per scale for teachers' reports (for each of the students). Based on the findings from the field test, including results from the students' self-report scales and parents' and teachers' scales, skill items were then further revised for the main study. This revision was done with the goal of selecting the best eight items for the parent scale and best three items for the teachers scale based on the item properties in all of the scales these items were used. The instrument development process of parent and teacher skill scales, including the number of items in each of the phases, is presented in Table 2.5.

Table 2.5. Development of parent and teacher survey items

Study stage	Timeline	Items per skill		Number of skills		Total number of items	
		Parent	Teacher	Parent	Teacher	**Parent**	**Teacher**
Online study of parents	Feb 2018	10		19		**190**	
Item trials	Apr 2018	10	10	19	19	**190**	**190**
Field test	Oct/Nov 2018	8	3	19	19	**152**	**57**
Main study	**Oct/Nov 2019**	**8**	**3**	**15**	**15**	**120**	**45**

References

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. [2]

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*. [5]

Goldberg, L. and G. Saucier (2019), *International Personality Item Pool*, https://ipip.ori.org/. [6]

Halverson, C. et al. (2003), "Personality Structure as Derived From Parental Ratings of Free Descriptions of Children: The Inventory of Child Individual Differences", *Journal of Personality*, Vol. 71/6, pp. 995-1026, http://dx.doi.org/10.1111/1467-6494.7106005. [8]

John, O. and F. De Fruyt (2015), *Framework for the Longitudinal Study of Social and Emotional Skills in Cities*, OECD Publishing, Paris. [1]

Kankaraš, M. and J. Suarez-Alvarez (2019), *Assessment framework of the OECD study on social and emotional skills*. [4]

Lipnevich, A., F. Preckel and R. Roberts (2017), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice*, Springer International Publishing, Switzerland. [3]

Mervielde, I. et al. (eds.) (1999), *Construction of the hierarchical personality inventory for children (HiPIC)*, Tilburg Univiersity Press. [7]

Soto, C. and O. John (2017), "The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power.", *Journal of Personality and Social Psychology*, Vol. 113/1, pp. 117-143, http://dx.doi.org/10.1037/pspp0000096. [9]

Notes

[1] Ohio State University (OSU) was the first international contractor who helped develop the survey design and oversaw the first round of empirical testing in the United States from 2017-2018. The Australian Council for Educational Research (ACER) oversaw the implementation of the field test and the main study, and collecting the data.

Chapter 3. Development of contextual questionnaires

This chapter describes the development of the contextual questionnaire framework and the phases for developing the contextual questionnaire items.

3.1. Introduction: the SSES contextual framework

Apart from assessing students' social and emotional skills, the survey examines a wide range of contextual factors. These can be divided into five broad groups:

- socio-demographic background of students
- family environment
- school environment
- peer environment
- wider community environment.

The contextual questionnaires aim to capture the most relevant information that influences students' social and emotional skills development, paying close attention to characteristics that are more responsive to policy interventions and to adaptations of teaching methods.

Students learn in many different settings, including in their families, schools and communities, with each context playing an important role throughout childhood and adolescence. Contextual information gives us a better understanding of what helps and what hinders social and emotional skills development, including the policies and practices that support them.

Therefore, the survey collects contextual information from students, their parents, teachers and school principals in order to provide insights into each of these environments (see the assessment framework for more information).

3.2. Development of contextual questionnaire items

In developing the contextual questionnaires for the four respondents – students, parents, teachers and principals – the OECD aimed to collect background information as complete as possible, as shown in Table 3.1.

Table 3.1. Background information collected in the four contextual questionnaires

Student questionnaire	Parent questionnaire	Teacher questionnaire	Principal questionnaire
Demographics	Demographics	Demographics	Principal's demographics and structure of school
Well-being, attitudes and aspirations	About the child	Education and professional development	Student and teacher population
Relations with friends and parents	Well-being and skill profile of parents	Teaching practices	School resources
School life	Relationships with their child	Teacher's school	Pedagogical practices, curriculum and assessment
Information and Communication Technology (ICT)	Perceptions of social and emotional skills	Perceptions of social and emotional skills	School climate
Short cognitive ability measure			Policies and practices

Apart from understanding what factors influence the development of these skills, the survey also looks at the relations between social and emotional skills and various life outcomes. Six different indicators of student outcomes are measured in the survey through information from the contextual questionnaires.

- academic achievement: academic aspiration and grades (math, reading and arts)
- active citizenship and civic participation: global mindedness, extra-curricular activities
- social connectedness: closeness to family, closeness to others
- health: healthy behaviours, overall health and BMI
- quality of life: subjective well-being, life satisfaction and test anxiety
- behaviour/conduct: school absenteeism, disruptive behaviour in the classroom, sleep quality, class participation, etc.

3.2.1. Initial review of contextual factors

The first phase, in developing the contextual questionnaires for the survey, involved an in-depth literature review, mapping out the factors in students' family, school, peer and community environments that could potentially influence students' social and emotional skills development.

The second phase involved examining the empirical evidence for each of the contextual factors, their characteristics and relations with social and emotional skills and outcomes.

In the third phase, the OECD reviewed existing measures, such as contextual questionnaires from other OECD large-scale assessments[1] in order to see if there were questions that could be beneficial for the survey. Questions from other OECD studies have not only been well researched and tested, but they have also been translated into multiple languages.

3.2.2. Online Survey

In February 2018, some 1 000 parents of children between ages 9 and 16, living in the United States, completed a survey, which tested 190 items for 19 social and emotional skill scales (see Chapter 2 section on development of assessment scales). The survey also included some background questions in order to see how the skill items related with contextual information and with outcomes (predictive validity), such as:

- socio-demographic characteristics about the parent and their child: gender, age, country of residence/birth, marital status (parent), language, education, school, grade (student), preschool attendance (student)
- economic and well-being indicators: activity status, occupation, job and life satisfaction, health, subjective well-being
- parents' skills, attitudes and perceptions: perceptions of value of social and emotional skills, growth mind-set
- parenting styles: encouragement, academic expectations

3.2.3. Item trials

In April and May 2018, four sites[2] translated and tested contextual questionnaire items that were different from those already taken from OECD surveys or items that had not been tested repeatedly in other international studies. Each site conducted interviews either by phone or in person with participants.

Table 3.2. Number of contextual questionnaire interviews across four sites

Type of questionnaire	Number of interviews
Student	20
Teacher	20
Parent	30
Principal	2

Interviewers entered comments on how well the items worked with each respondent (e.g. easy to understand, how well they applied to the respondent, how well did the response scales work). Results were used to improve the questions and how they were formulated in preparation for administering the contextual questionnaires in the field test.

3.2.4. Field test

Eleven cities participated in the Field Trial. A total of 13 861 students from 515 schools from these sites completed a student questionnaire, which was conducted online and consisted of 51 questions (across cohorts). The parent questionnaire which consisted of 33 questions, was completed by 7,856 parents of participating children. The teacher questionnaire (32 questions) and the school principal questionnaire (52 questions) were completed by 4,697 teachers and by 448 school principals respectively. The teachers, principals and parent questionnaires were available online, although parents also had a paper option available to them.

The consortium staff undertook a full psychometric analysis of all questionnaire data. This included the following steps:

- Analysis of response frequencies and missing data
- Exploratory and Confirmatory Factor Analysis
- Correlations and triangulations between indirect and direct outcomes measures

The contextual questionnaires were also assessed for age appropriateness and length. Based on all of these factors, decisions were made about which questions should remain in the questionnaires, which need to be modified, and finally which questions needed to be removed. It was necessary to reduce the number of questions as much as possible while still capturing the most important contextual factors of the students' teaching and learning environment in order to limit the response burden of each of the four contextual questionnaires. Therefore, the contextual questionnaires were reduced by 23% for students, by 11% for parents, by 4% for teachers and by 44% for principals. This ensured that respondents' time was used more efficiently during the main study.

3.2.5. Main study

Ten sites went on to participate in the main study. On average, older students took 60 minutes and younger students 46 minutes to complete both the skill items survey and the contextual questionnaire. All respondents filled out the questionnaires online, while parents had an option to fill out the questionnaire on paper.

Notes

[1] Other OECD large-scale assessments include the Programme for International Student Assessment (PISA), the Programme for the International Assessment of Adult Competencies (PIAAC) and the Teaching and Learning International Survey (TALIS).

[2] The following four sites translated and tested contextual questionnaire items for the item trials: Daegu, South Korea; Moscow, Russia; Ottawa, Canada; Rome, Italy.

Chapter 4. Sample design

This chapter describes the processes used in selecting schools and students for the Field test and Main study stages of the OECD Survey on Social and Emotional Skills 2019 (SSES).

4.1. Overview of the sample design

The primary focus of the sample design was selecting enough students to participate in SSES with the purpose of maximising the quality of inferences to the student populations from which they were drawn. Background data were also collected from the schools attended by the sampled students, via a principal questionnaire; and from parents of the sampled students, and teachers who were nominated as most closely connected to the sampled students. As described, selecting these additional populations of principals, teachers and parents was directly connected to the students sampled to participate in the survey.

Prior to selection at the Field test and Main study stages, a sample design was negotiated with each Site, primarily through a series of sampling forms. Every Site was asked to complete four forms providing information about the following aspects:

1. A description of the Site
2. The target population, including testing periods and rules for age eligibility
3. Proposed exclusions and limitations of population coverage
4. Proposed stratification variables.

Figure F.1 in Annex F has the templates for each of these sampling forms.

The forms were submitted to the International Contractor and became the basis for discussions between the International Contractor and the Site. A major focus of the discussions was confirming that the proposal from the Site aligned with the Technical Standards established for SSES, with the goal of ensuring that the collected data for each participating Site was of high quality and comparable.

While the Technical Standards established for SSES were of most relevance at the Main study stage, Site proposals were carefully checked at the Field test stage to ensure that potential issues were identified early. As most Site Project Managers (SPMs) had not previously been involved in sample design work of this nature, the opportunity to consider these issues at the Field test provided valuable experience in the lead up to the Main study.

4.1.1. Site description

The sample design discussions began with participants clearly defining their Site. A Site could be a city, a region, or some other well-defined entity such as a school district. Following discussions between participants and the International Contractor, proposed definitions were submitted to OECD for final approval. Table F.1 in Annex F presents the agreed descriptions of each participating Site.

4.1.2. Target population

Within each participating Site there were two target student populations. These populations are described briefly in this document, and throughout communications related to SSES, as ‘10 year-olds’ and ‘15 year-olds’, and are also referred to respectively as the ‘younger’ and ‘older’ cohorts. The actual eligible age range varies by some months from these brief descriptions, as described below. In addition to the age component, there was a grade component to student eligibility, also described in the target population definitions below.

- **Older cohort:** Students aged between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing period, attending educational institutions located within the Site boundaries, and in grade 7 or higher. The age range of the population may vary up to one month, either older or younger,

but the age range must remain 12 months in length. That is, the population can be as young as between 15 years and 2 (completed) months and 16 years and 1 (completed) month at the beginning of the testing period; or as old as between 15 years and 4 (completed) months and 16 years and 3 (completed) months at the beginning of the testing period.

- **Younger cohort:** Students aged between 10 years and 3 (completed) months and 11 years and 2 (completed) months at the beginning of the testing period, attending educational institutions located within the Site boundaries, and in grade 2 or higher. The age range of the population may vary up to one month, either older or younger, but the age range must remain 12 months in length. That is, the population can be as young as between 10 years and 2 (completed) months and 11 years and 1 (completed) month at the beginning of the testing period; or as old as between 10 years and 4 (completed) months and 11 years and 3 (completed) months at the beginning of the testing period.

The adjustments described above were permitted so as to better align the timing of the testing period to the school year for the participating Site. Any proposed adjustments needed to maintain an age range that was exactly 12 months in length. With respect to the older cohort, these definitions also align student ages to those used for the PISA survey.[1]

4.1.3. Population exclusions

SSES aimed to be as inclusive as possible of students within the target populations. However, in some cases political or operational factors made complete coverage of the Site population difficult to obtain. To ensure maximal coverage of the Site population and comparability across Sites, the following limits were set on the degree of exclusions that could be implemented:

- school level exclusions that were due to geographical inaccessibility, or where administration of SSES would be not feasible within the school were limited to less than 0.5% of the SSES Desired Target Population
- school level exclusions that were due to a school containing only students that would be within-school exclusions were limited to within 2.0% of the SSES Desired Target Population
- within-school exclusions were limited to within 2.5% of the SSES Desired Target Population – these exclusions could include, for example, students not able to do the assessment because of a functional disability.

As part of the sample design preparations, the International Contractor consulted with each Site about exclusions that might be warranted based on their particular conditions.

School level exclusions

The types of school-level exclusions that were considered permissible within the SSES Sampling Framework included:

- geographical inaccessibility
- extremely small size
- administering the SSES assessment within the school would not be feasible
- all students in the school would be within-school exclusions, or
- other reasons as agreed upon.

Within-school student level exclusions

Across all Sites it was expected that certain students would not be able to access the assessment for various reasons. These include language barriers, or cognitive or functional disabilities. Students could be excluded from the assessment for one or more of the following reasons:

- They are functionally disabled in such a way that they cannot take the assessment. Functionally disabled students are those with a moderate to severe permanent physical disability.
- They have a cognitive, behavioural or emotional disability confirmed by qualified staff, meaning they cannot take the assessment. These are students who are cognitively, behaviourally or emotionally unable to follow even the general instructions of the assessment.
- They have insufficient assessment language experience to take the SSES assessment. Students who have insufficient assessment language experience are those who have received less than one year of instruction in the assessment language.
- They cannot be assessed for some other reason as agreed upon.

To avoid confusion with schools about whether students should be excluded, including the final option was only for a very specific reason to be negotiated between Sites and the International Contractor prior to sampling. For the Main study, no Site required an 'other' category for within-school exclusions.

Table F.2 in Annex F shows the school and within-school exclusion categories negotiated with each of the Sites.

4.1.4. Stratification

Stratification is a process of organising the sampling frame to improve the efficiency of the sample design. This process results in more reliable survey estimates, and ensures an appropriate representation of major population groups. Two types of stratification can be distinguished: explicit stratification and implicit stratification.

Explicit stratification involves partitioning a sampling frame into mutually exclusive parts according to certain variables, from which separate, independent samples are drawn. Examples of variables used in explicit stratification are those that defined important subpopulation categories, such as language and sector. Explicit stratification allows different sample designs to be applied to individual explicit strata.

Implicit stratification involves sorting the sampling frame by nominated variables and then systematically sampling from this sorted list. The purpose of implicit stratification is to ensure a proportional sample allocation across these variables. Where the variables are correlated with major outcomes – for example socio-economic background or geographic location variables – this process improves the reliability of survey estimates of those outcomes.

Table F.3 in Annex F contains the explicit and implicit stratification variables for each cohort and Site.

4.2. The sampling frame

Following the discussions and agreement on the sampling design, Sites submitted their list of schools – the sampling frame - to the International Contractor for the purposes of sample

selection. This school frame was checked by the International Contractor for consistency with the information provided on the sampling forms. The sampling frame contained the information needed to draw the school sample, in particular:

- A national unique school Identification number, commonly used in the education system of the Site and used for easy identification of the selected school.
- the number of eligible students in the younger and older cohorts enrolled in the school. This information was used to determine the school Measure of Size (MOS) for sampling purposes.
- the values of any nominated explicit and implicit stratification variables.

The sampling frame was delivered at both the Field test and Main study stages, ensuring that sample selection was based on the latest available data from the Site.

4.3. The Field test and Main study sample designs

For both the Field test and the Main study, two-stage stratified cluster samples were selected for each cohort. At the first stage, schools were sampled with probability proportional to size, and at the second stage, an equal probability sample of an agreed number of students (the Target Cluster Size, TCS), was selected from those sampled schools. The combination of probability proportional to the school measure of size (PPS) sampling of schools at the first stage, followed by an equal probability sample of a cluster of students at the second stage, ensured the desirable outcome that all students from the Site had the same chance of inclusion in the sample.

The TCS was negotiated with each Site during the sample preparation stage. For most Sites, the TCS was set to 50 students both for the Field test and Main study. However, this value was changed in some particular circumstances, particularly for Sites with a small number of schools where a larger TCS was needed to achieve the desired student yield. These cases were the census Sites of Manizales (Colombia) and Sintra (Portugal), as well as the older cohorts of Houston (United States) and Helsinki (Finland) (TCS=75) and Ottawa (Canada) (TCS=60).

For larger schools, the MOS was equal to the target population enrolment estimate for the school as recorded on the sampling frame. For schools with an enrolment estimate less than the TCS, the MOS was set to the TCS in most cases. The setting of the MOS to TCS for these smaller schools gave all students from these schools the same chance of inclusion into the sample.

In Sites where a substantial proportion of the enrolment was in small schools, an additional measure was taken in order to balance the objectives of preserving the overall sample yield, while minimising the costs and operational burden of having too many small schools sampled for the Study. This balance was achieved by the following steps, drawn directly from the processes used in PISA (OECD, 2017, pp. 77-79):

1. Classifying small schools into three categories: extremely small (0, 1 or 2 students); very small (3 students to half the TCS); and moderately small (between TCS/2 and TCS students).
2. If the percentage of students in the smallest schools (ENR < TCS/2) was 1% or more, then very small schools were under-sampled and the school sample size increased, sufficient to maintain the required overall yield.
3. If the percentage of students in the smallest schools was less than 1%, and the percentage of students in moderately small schools was 4% or more, then there was

no required under-sampling of very small schools but the school sample size was increased, sufficient to maintain the required overall yield.

Where the above conditions indicated that under-sampling was required, this was accomplished by assigning a MOS of TCS/2 to those very small schools with an ENR greater than two but less than TCS/2 and a MOS equal to the TCS/4 for the very small schools with an ENR of zero, one, or two. In effect, very small schools with a MOS equal to TCS/2 were under-sampled by a factor of two (school probability of selection reduced by half), and the very small schools with a MOS equal to TCS/4 were under-sampled by a factor of four (school probability of selection reduced by three-fourths).

Further details about the sampling approach at the Field test and Main study stages are provided below.

4.3.1. Field test design

The Field test was conducted across all Sites between October and December of 2018. The primary objectives of the Field test were:

- to collect sufficient response data across the target populations in order to assess the psychometric properties of the administered items; and
- to test field procedures, in preparation for the Main study. This included the process for selecting samples of students within selected schools.

The required sample size for the Field test was a minimum of 15 schools and 500 students for each cohort. For Sites having more than one language, an increased sample size was negotiated so as to ensure a good amount of response data across the different languages. Table F.4 in Annex F shows the school and student sample sizes across different language groups for the Field test.

The selection of schools to participate in the Field test was undertaken by the International Contractor following agreement between the Site, the International Contractor and OECD on the sampling forms and the school frame. A systematic, two-stage design was used. Prior to sampling, frames were sorted by the stratification variables identified in the design, and by enrolment size. Schools were then selected with PPS sampling. The school sampling approach is illustrated in Annex 11.A in Chapter 11.

For each school sampled to participate, up to two schools were assigned as potential substitutes to be used in the event that the sampled school could not participate in the Study. The assigned substitutes were the schools adjacent to the sampled schools on the frame used for sampling. Because the frame had been sorted by the stratification variables and by size, the substitutes generally matched the sampled school against these characteristics. The assignment of school substitutes and their (limited) use as a measure to address the non-response of sampled schools followed the practice used in all other major international education surveys for many years.

At the second stage of sampling, the agreed number of students to be sampled from each school (the TCS) was selected systematically and with equal probability, as described further below.

A software package was distributed to all Sites which facilitated the task of within-school sample selection. The software generated a form for each sampled school, with the school number and name. The student listing form included demographic variables to be recorded for all eligible students, including current school year level, date of birth, gender and any special educational needs. SPMs distributed these forms to each sampled school and collected the returned, completed sampling forms for importing into the sampling software.

All student records on the lists returned from schools were checked for eligibility to the target population, that is, that the recorded birthdate and grade level were within the eligible ranges for the cohort. Any ineligible records were removed from the lists prior to sampling. Eligible records were then sorted by grade level and gender. Students were systematically selected from this sorted list with equal probability using a 'random start, constant interval' procedure. The sampling interval was calculated as the ratio of the number of students on the list and the TCS. Within the first sampling interval, a student was randomly selected. The selection of all subsequent students was achieved by successfully adding the sampling interval to the initial random start. This process ensured that an equal probability sample of exactly the TCS number of students was selected from the school, and that the distribution of the selected sample mirrored the corresponding distributions of eligible students at the school for grade level and gender. If a school contained fewer eligible students than the TCS, then all students were automatically sampled into the Study.

4.3.2. Main study design

The Main study was conducted across all Sites between September and December of 2019. The primary objectives of the Main study design were to obtain high quality data from each Site in order make inferences to the target population, and also to maximise comparability of outcomes across Sites as well as over time.

A second round of sample design negotiations was undertaken with each Site, considerably enhanced with the experiences and lessons learned through the Field test stage.

The target sample sizes for the Main study were for a minimum of 75 schools and 3000 assessed students. If there were fewer than 75 schools within the Site, all schools were included. As for the Field test, Sites were expected to limit school-level and within-school exclusions to within 5% of the target population.

The process for student selection within the sampled schools followed the same patterns as those explained for the Field test.

Participation standards were established for the survey in order to minimise the potential of bias from the presence of non-responding schools and students. Sites were expected to achieve a school response rate of at least 85% of sampled schools and a student response rate of at least 80% of all sampled students from the responding schools.

Consistent with the practices used in the PISA survey, there were consequences with respect to response rate calculations and data inclusion for cases where within-school participation rates fell below certain benchmarks. If the within-school response rates fell below 50%, the school was considered a non-participant in response rate calculations. Furthermore, if the rate fell below 25%, the student data was removed from the international database. These measures were used because of the greater risk of non-response bias when response rates fall to these levels.

As for the Field test, up to two substitutes were matched to each sampled school, for use in the event that the sampled school did not participate in the survey. Only non-responding schools with eligible students were substituted; out-of-scope schools were not substituted. As with other major educational surveys, the limited use of school substitutes was permitted as a measure to reduce the risk of non-response bias.

In the event that data from a Site fell short of the sampling metrics – sample sizes, exclusion rates, participation rates, use of school substitutes – further information was sought from the Site to determine whether the shortfalls represented a risk to the quality and comparability of the data. Ultimately, these issues were discussed as part of the data adjudication process, described in Chapter 13. Sampling outcomes, including target and

achieved school and student sample sizes, exclusion rates and response rates are presented in Chapter 11.

References

OECD. (2017). *PISA 2015 Technical Report.* Paris: OECD Publishing. Retrieved from https://www.oecd.org/pisa/data/2015-technical-report/

Notes

[1] See (OECD, 2017) Chapter 4, p. 66 for details of the Population Definition for PISA.

Chapter 5. The online delivery platform

Online delivery was the primary delivery mode used during data collection for the OECD Survey on Social and Emotional Skills 2019 (SSES). This chapter provides detailed information about the platform used and main features deployed during the SSES implementation.

5.1. Introduction

The SSES online delivery platform was hosted by RM Assessment Master (AM), a web-based software for the delivery and online management of survey and assessment development processes, developed by SoNET Systems part of RM Results.

The AM software provided an integrated system for managing the translation verification process, the online delivery of the instruments, the real-time monitoring of participation during delivery, together with a platform for data management and reporting. The following AM software modules were used by the International Contractor:

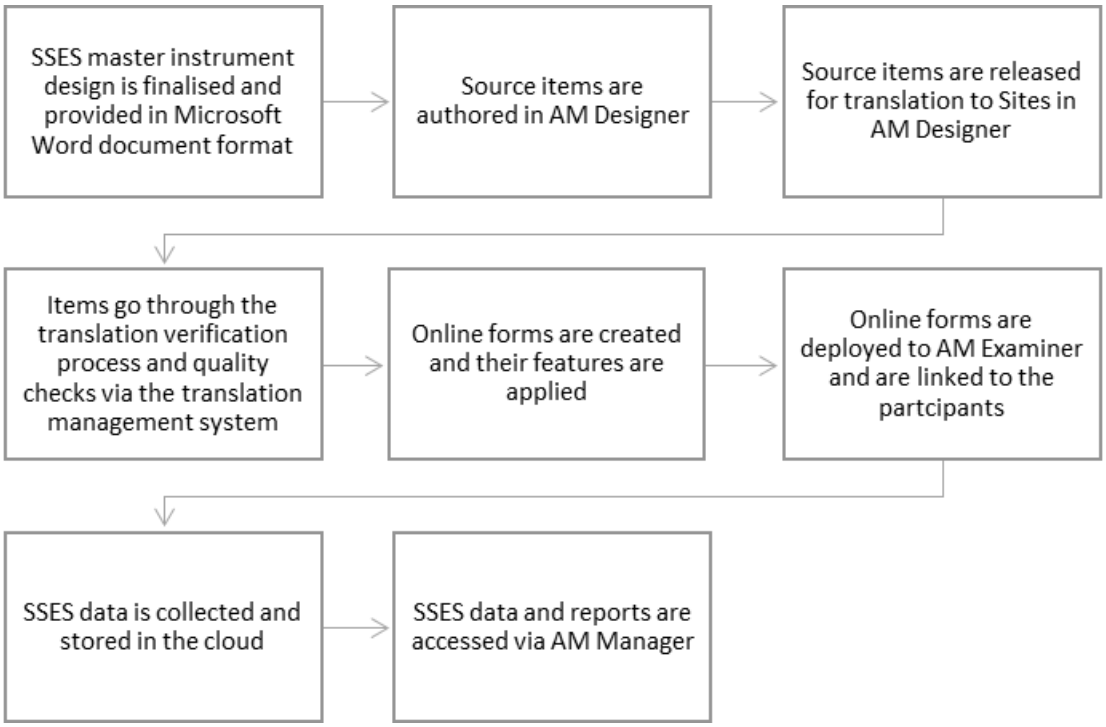
- AM Designer module for item authoring
- AM Designer module for managing the translation verification and finalisation process of the instruments
- AM Examiner module for managing online delivery
- AM Manager for data management, reporting and live monitoring during online delivery.

The following sections cover some of the software features and customised elements designed to support the delivery and management of SSES.

5.2. Overview of the online instrument production

Figure 5.1 shows the workflow within the AM modules from creation of the online forms to data collection and reporting. This workflow was implemented for both the Field Test and the Main Study.

Figure 5.1. Workflow for online instrument production



In preparation for the field test, the SSES master instruments were initially prepared in English in Microsoft Word document format. Once finalised, those files served as the source version from which the online instruments were created by the item authoring team in AM Designer.

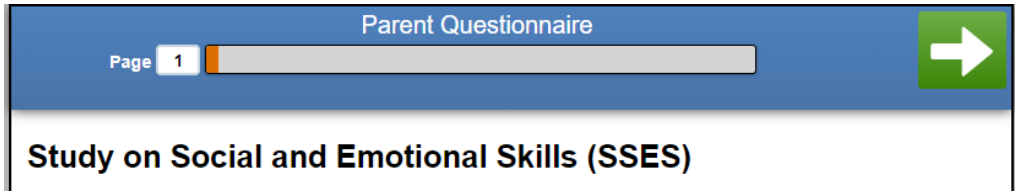
AM Designer allowed for a range of features to be implemented. However, as this was the first cycle of SSES, additional features had to be designed or customised to suit the particular requirements of the Study. These included customising the questionnaire interface, developing specific functions to assist participants while navigating through the forms and customising the reporting system.

Online delivery for the field test employed these customised features together with existing features of the system. After the field test data collection was completed, a number of new functionalities and reports were developed for the main study based on an extensive internal review as well as feedback from the participating Sites. This input helped enhance the usability and navigability of the online forms, and improve the monitoring ability of the system. These enhancements also assisted in improving the quality of the data and achieving higher completion rates.

5.3. Features of the online instruments

Below are further details regarding the technical and visual features that were developed and applied to the SSES instruments in the assessment delivery platform.

Figure 5.2. Online interface



5.3.1. Visual design features

- As shown in Figure 5.2, the SSES online interface design aimed to be clear and user-friendly. Each page included the title of the instrument, a progress bar, the page number, and 'next' and 'back' buttons to navigate through the questionnaires.
- High contrast colours were used for all interface elements enabling easier navigation for younger users and those with visual difficulties.
- The interface was designed to maximise the space allocated to the 'response area' so that as much content as possible was visible without the need to scroll.
- A 'remember to scroll' reminder message was incorporated in the interface and appeared automatically on detecting that scrolling was required on a page.
- A summary page and navigation pad were added to the online assessment forms for the main study. These features allowed users to identify the completion status of an item by colour and navigate to items by clicking on the corresponding button.
- Common online response formats such as Likert-like tables, multiple-choice options, drop-down menus and text boxes were used in each instrument. The radio buttons in the online forms were made large enough to accommodate touch-enabled

devices and consecutive rows in tables alternated in colour for easier differentiation.

- All respondents, except teachers, were allocated one online assessment form. Teachers were tasked with completing a teacher background questionnaire designed to capture information about themselves. Separately, they were asked to complete an assessment form for each student linked to them who was participating in the Study. Teachers were able to access their own personal interface that listed their background questionnaire and all the student assessment form/s allocated to them, identifiable by the respective student usernames. Teachers could track the completion status of each assessment form using the progress bar that appeared next to each form in the interface.

5.3.2. Technical features

- A customised short URL was created for all respondents to access the online assessment instruments. This link provided users with easier and faster access.
- Conditional question visibility rules and skip patterns were deployed, whereby the respondent would be shown certain items based on the responses they had given to previous items.
- Random allocation of items was applied to some item sets, so that each item from the specified set was randomly allocated to a fixed percentage of respondents in each cohort.
- Flexible completion of the forms was possible. Participants could leave the instrument and return to resume from where they left. This feature allowed the adults who were participating in the Study to complete the assigned forms in as many sittings as needed.
- Sites that were using instruments in more than one language were able to implement a language selection drop-down feature; Ottawa, Helsinki and Houston all implemented this feature in their online forms. This option allowed respondents to select the language with which they were more comfortable.
- All data were saved automatically once input by the respondent. This feature protected against possible loss of data in the event of internet disconnection, device failure and other issues such as electricity outage or low internet connection speeds that might impede progress.

5.4. The translation verification management system

The AM Designer module provided users with access to a multi-featured translation management software system. To access the software, users were advised to have a high-speed internet connection and use a recent version of popular web browsers; Google Chrome or Firefox were recommended.

5.4.1. Translation verification workflow management

After the field test assessment items were authored and checked in AM Designer, they were released to the translation management system, together with other texts that also required translation, such as system function texts, introduction and conclusion texts and assessment administration scripts. Sites were provided with access to the system where they could view the source text and create a localised translated version of the SSES items. The assessment items were arranged in folders by instrument type and cohort, such as instrument folders

for parents, principals, older students, younger students and teachers. System users could then select the item they required to edit from within these folders as needed.

As the items went through the various stages of translation, review, and verification, their translation status was updated to reflect their stage in the translation workflow. This status was visible at the item library level along with the item target language, and the unique item ID. Users were able to filter the items using these fields as needed. In addition, the system allowed for customisation of user access to the item folders, which ensured that only those responsible for that particular step in the translation workflow had access during that step. For example, while translation verification for the parent instrument was underway, access to the parent instrument was blocked to other Site Project Centre (SPC) staff so that changes could only be made by the verifier. This was a crucial step in ensuring the reliability of the translation workflow.

The main study followed the same instrument production steps as the field test but the folder hierarchy differed slightly. Within each instrument folder, a further distinction was made between items that were retained from the field test without changes and those that were new, changed or updated. The translations of retained items were set to view-only mode. This mode prevented them from being altered, as they did not need to go through the translation verification process once again, while all other items went through the entire translation verification process. This approach made it easier for users to search for items and organise the translation verification process.

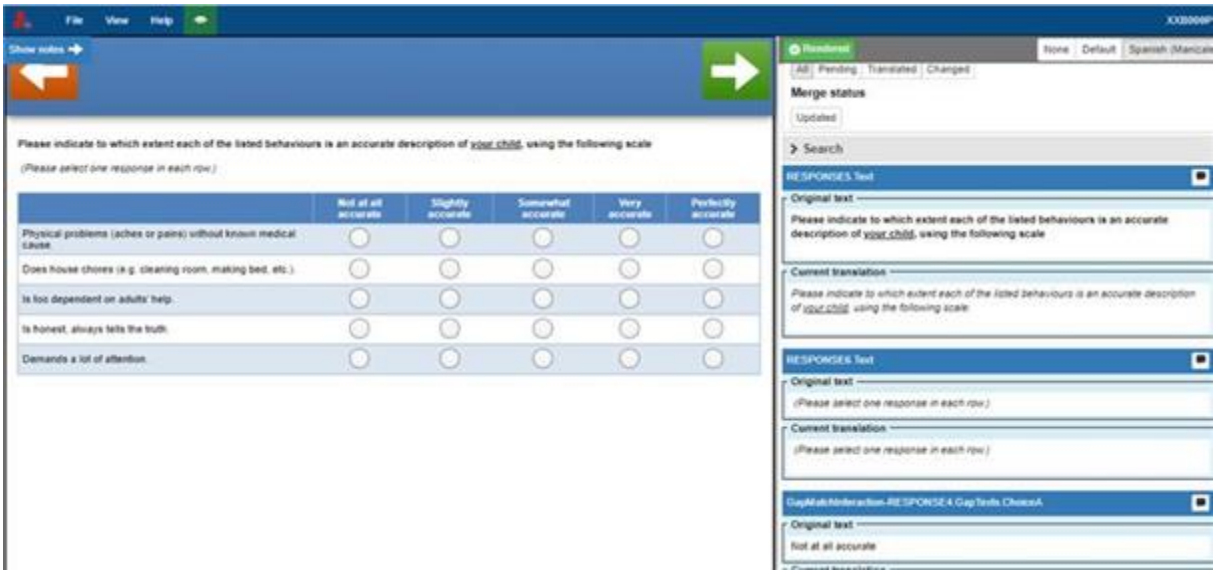
5.4.2. At the individual item level

Once an item was selected from the translation system main interface, the item editing window opened. The item window allowed translation system users to view each SSES item in the customised interface and update each segment of text with the target language translation. The item editing interface was split between a right panel for translation editing and a live preview window on the left. The live preview window provided users with a preview of the translation populated into the SSES item interface, allowing the user to review the changes made while each segment of text was being updated. It also enabled users to preview the overall layout of the page to see how it would appear in the live online form.

Other features of the item editor included:

- text editing and formatting tools
- the ability to track changes in the translation of text segments and view the translation history to compare versions
- translation memory and auto population: the translation system was able to detect identical segments of source text and automatically populate them with the translation entered previously
- dual window preview, where the source text preview could be reviewed alongside the target translated text
- preview mode allowed translators to interact with each page feature in near full function
- the ability to mark segments of text that had been reviewed as ‘ticked’ to distinguish them from those that had not yet been reviewed or had issues pending.

Figure 5.3. Translation editing panel



The platform also allowed the International Contractor's system administrators to download the translations for each Site in a specified spreadsheet format. The spreadsheet showed each segment of source text in a column, with the translated version in the target language in the adjacent column on the same row. This spreadsheet was provided to the SPCs and translation verifiers at the relevant stages in the translation process to assist with the review and verification of the translation.

5.5. Construction of online assessment forms

The draft forms for each language were created once the translation verification process had completed for that language. These Site-specific forms included all the localised adaptations for that language including the changes requested to the master version to ensure they were applicable to the local context. For each Site, any items/options agreed to be removed were excluded, while Site-specific items were added to the end of the relevant online forms.

5.5.1. Draft form quality assurance checks

A series of quality assurance checks were then carried out on each set of Site-specific online forms. First, each online form was verified against the agreed Site Adaptation Verification Form (SAVF) by the International Contractor and checked for overall functionality. Second, external language verifiers were given access to the online forms to check the accuracy and quality of the final translations post verification, and document in the SAVF any discrepancies or issues for the Site Project Manager (SPM) to review. Last, the SPM reviewed their online forms and accepted or rejected the changes suggested by the verifier and made final updates accordingly. By doing so, the SPM signed off on the contents of the draft online forms.

At that point, other features such as language selection, skip logic and random allocation were applied to the draft online forms. The International Contractor tested and verified the functionality and layout of the forms before giving access to the SPCs for the final User Acceptance Test (UAT). SPMs were provided with login credentials for the updated forms

and details of the special features applied and their expected functionality. They were required to navigate through each online form while testing the functionality of the assessment items and special features applied to the form to ensure they were operating as intended. This was a crucial step and a final check point before live testing and ensured that the online forms were free from error and functioned appropriately in the local setting.

After the SPM signed off on the final online forms that included all the functional features, their access to the translation system was removed. This ensured that no further changes occurred and that no accidental errors were introduced. It also marked the completion of the quality assurance checks of the draft forms.

5.5.2. Respondent management and online form assignment

A systematic SSES respondent ID system was developed for both within-school sampling and respondent management in the online delivery platform. Using this ID system, a pool of potential respondent IDs was created, and through the within-school sampling tool, unique IDs were allocated to the sampled or identified respondents. The SSES respondent ID comprised only digits and contained a Site ID. For each participating Site, the IDs for students, parents, teachers and principals (schools) were generated by cohorts following a numerical logic. Using this logic, student IDs were easily linked to their school and their parent IDs, and teacher IDs were also linked to their respective schools. Each unique ID in the pool had a corresponding password for each potential respondent. This numerical logic nested in the respondent IDs allowed for more flexible data manipulation and reporting.

Once the Site-specific final online forms were signed off by the SPM, all online forms, except the indirect student assessment form for teachers, were assigned to the corresponding respondent IDs generated for the Site. For instance, the Parent online forms were assigned to all parent IDs and the Younger Student online forms were assigned to all younger student IDs.

The linking between sampled student and their teacher was achieved using the Student Teacher Linkage Correspondence (STLC) form that was generated by the within-school sampling tool. SPMs were required to upload the STLC for all sampled students to the AM Manager platform. The upload process served two purposes: to link teachers to the student/s assigned to them as specified in the STLC, and to 'activate' all respondent IDs for their Site.

To 'activate' meant that:

- teachers were assigned one student assessment form for each student linked to them
- all the dynamic SSES participation reports would include the data of all the respondents, such as students and teacher, as specified in the STLC, as well as the students' respective parent and their school principal as associated by the ID logic
- online delivery could commence at the Site for the schools that had had their STLC uploaded.

5.5.3. Online form data collection

The data reports included data labels that were customised to report against each data point in the forms, whereby the reported contents depended on the item type. For radio button items and check boxes, a coded data value was recorded. Some drop-down menus were coded to report their values while others were coded to report a data value. Textual responses reported the values entered by the respondents. The nature of the SSES items entailed that only the cognitive set of items were scored either as '0' or '1'.

Every data point in each online form for each language version was tested to ensure it was reporting correctly in the final data output files. Test data runs were carried out to ensure that skip logic and random allocation were functioning correctly and reporting appropriately. Testing included 'dummy data' runs and spot-checking. Any inconsistency found was corrected and tested again to ensure the data collected was free from error.

Where paper forms were used for the parent questionnaire, SPC staff entered the data into the corresponding respondent online forms directly in the assessment delivery system.

The International Contractor's system administrators had access to reports detailing the language selected, responses and time of login and logoff for each respondent that could be downloaded from AM Manager, as well as to a visual record of each form with the participant responses.

5.6. Online delivery

SSES assessment materials were primarily delivered online via available local internet connection. To access the online delivery platform, all respondents were provided with a URL that linked to a secure portal where they could enter their unique login details before accessing their assigned materials. All cohorts of respondents were required to use a recent model of any popular web browser; Google Chrome and Firefox were recommended. They were also required to have access to a high-speed internet connection in order to be able to complete the online instruments successfully.

5.6.1. Preparation for assessment administration at schools

As outlined in the School Coordinator Manual, each participating school was provided with the required technical specifications of the devices that were to be used for SSES to ensure successful delivery of the online instruments. Laptops, tablets and desktops were allowed to be used if they met the following technical specifications:

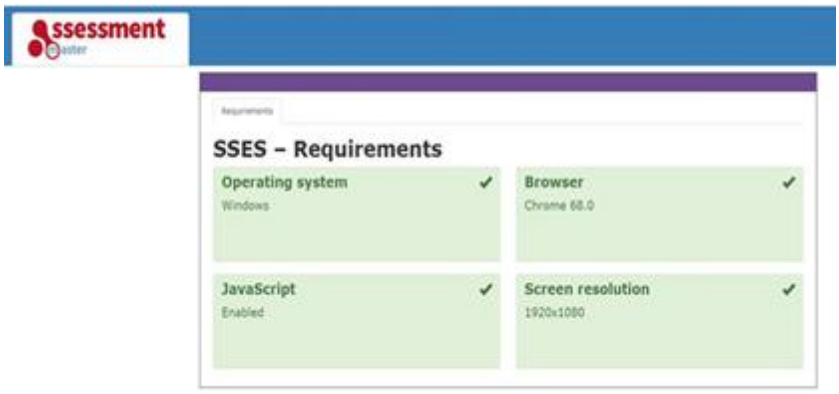
- adequate screen resolution, 1024x768 minimum
- JavaScript was enabled
- all up-to-date major internet browsers were supported – Google Chrome and Firefox were recommended for best performance
- able to access to a high-speed internet connection.

In addition, Sites were alerted about potential network-related issues which could impact on the performance of the online delivery, such as network filters imposed on foreign URLs, security firewall settings and limited bandwidth capacities of the local network providers for each sampled school. It was also recommended that the School Coordinator check the capacity of the school internet connection in terms of the number of students that could be accommodated in a single online assessment session and to schedule the session accordingly. If needed, concurrent use of the same online network in schools was to be limited during the scheduled assessment session.

Technical Readiness Test

In addition to the technical specifications provided, a Technical Readiness Test (TRT) was developed and provided in the School Coordinator Manual for participating schools to check the devices to be used in advance of the scheduled assessment day. The URL for the TRT directed the user to an automatic online test that was specifically designed to ensure that the device being used met all required technical specifications.

Figure 5.4. Technical Readiness Test onscreen report



In recognition that the TRT could require further technical intervention beyond the expertise of the School Coordinator, each school was required to nominate an IT Coordinator to assist with the TRT. The TRT instant onscreen report (see Figure 5.4) identified which (if any) of the technical criteria were not met on a device. This report allowed the IT Coordinator to carry out specific updates to the device and resolve any potential issues before the online assessment took place, such as updating the browser version or enabling JavaScript.

5.6.2. Data security

In compliance with the data security requirements of the Study, no identifiable personal data of the respondents, such as name, date of birth, or school name, were ever requested from participating Sites, and therefore were never stored in the AM platform or in the cloud.

All users of the SSES online delivery and management system were provided with login details individually via their email address. A default password was generated by the system, then the user was prompted to create their own unique password at first login. As a further measure to protect data security, different levels of user access were set up for AM Management users, which allowed for the customisation of access to the raw data reports as needed.

All SSES assessment data and test forms were stored in a centralised database repository with role-based user access control, hosted on the Amazon Web Services (AWS) cloud, secured in accordance with recommended best practices and adhering to ISO 27001 Standards and ASD ISM controls, by the RM Results team at their sites in Frankfurt and in Sydney.

5.7. Online assessment monitoring and reporting

A customised online management portal was developed for the Study to assist with the management and monitoring of assessment progress during the online delivery window. Each SPM was provided with login details to access the AM Management portal once the final forms for their Site had been created. This portal allowed SPMs to access various participation reports and customised downloads for occupation coding data. These reports could be downloaded as needed or scheduled to generate at a specified time interval. The portal also provided the STLC upload feature and a live dashboard to monitor assessment progress at the Site in real time.

5.7.1. Monitoring participation during the assessment period

The management portal enabled SPMs to track and monitor respondent participation progress for their Site in real time. Key features included:

- **Live dashboard**: A dynamic and interactive interface, which provided a high-level overview of the participation status in real time for each respondent group including students in each cohort, parents, teachers and principals. It also facilitated exports of this data in PDF and Excel format.
- **Participation Reports**: These downloadable reports provided detailed participation status data for each respondent, listed against each ID, the assigned form status of either 'logged in' and/or 'submitted'. The student participation report for each cohort was required as part of the data validation process at the end of the data collection period. These participation reports are available:
  - Student Participation Report
  - Parent Participation Report
  - Teacher Participation Report
  - Principal Participation Report
- **Summary Reports:** These downloadable reports provided a participation summary at the school level, including the total number and percentage of 'logged in' and 'submitted' respondents for each school. These summary reports could be generated for students, teachers and parents. They proved to be an effective tool that supported SPMs in monitoring the participation progress of each sampled school and identifying where follow up with schools was needed.

5.7.2. Other available reports

- **Occupational Data Reports:** Occupational data reports for students and parents were made available for Sites to download at the end of the data collection period. These specially designed reports exported the response fields that required coding for each respondent and presented them with column headers that were necessary for the occupation coding process.
- **Response Data files:** Response data files for each respondent group were downloadable by the International Contractor for data cleaning and analysis.

5.8. Summary

For the SSES main study, more than 160,000 online forms were completed by the respondents during the data collection period and all data were collected successfully without loss**.** Web-based online delivery presented many benefits over traditional paper-based delivery such as the ability to monitor progress in real-time, more efficient data capturing, improved data quality by eliminating human error from the data entry and coding process and a reduced SPC workload.

However, web-based delivery also presented a unique set of challenges in its own right. Factors such as limited access to high-speed internet connection in some areas, limited access to a device with suitable technical features and limited user familiarity with online interfaces were identified as issues that could affect parent cohort participation in some Sites. For the main study, a number of Sites prepared a paper version of the parent online form as a backup option. Only Bogota (Colombia) relied entirely on paper format for

delivering the parent questionnaires. Some Sites also prepared paper versions of the student forms as a backup option but did not need to use them.

Chapter 6. Site instrument quality assurance

This chapter describes the Site instrument quality assurance procedures undertaken during the preparation of national versions of survey instruments used in the OECD Survey on Social and Emotional Skills 2019 (SSES).

6.1. Introduction

The overall goal of Site instrument quality assurance in SSES (the Study) was to ensure the survey instruments used by all participants were of high quality, and that translations were appropriately adapted for the national context and internationally comparable. Rigorous procedures were implemented to achieve these goals.

These procedures included:

- adaptation and translation of the international source version of the SSES survey instruments into national languages
- the development and use of the monitoring tool and adaptation and translation guidelines
- the international verification procedures.

6.2. SSES instruments to be translated

The Study involved several instruments that were adapted and translated.

- **Student self-administered instrument** which comprises two components: an assessment of the student's social and emotional skills, and a contextual questionnaire.
- **Parent questionnaire** which includes questions about their child's social and emotional skills, and a contextual questionnaire.
- **Teacher questionnaire** which consists of two parts. The first part is a contextual questionnaire. The second part collects the teacher's assessment of a student's social and emotional skills. It was recommended that teachers complete the second part for up to fifteen students.
- **Principal questionnaire** which consists of contextual questions.

6.3. Languages used in SSES

For most participating Sites, identifying the target language – the language in which the survey instruments would be administered – was straightforward, because they have one dominant language. Some Sites, however, use more than one language of instruction in their education systems. For example, Ottawa (Canada) administered all survey instruments in both English and French. In other cases, while there may be one language of instruction there are other languages that are prominent in parts of the community. For example, Houston (United States) administered all survey instruments in English and the student and parent questionnaires also in Spanish. In total, 14 different language versions across 10 Sites were administered. Table 6.1 shows the languages used by each participant for the various survey instruments during the Main Study.

Table 6.1. Languages used for the SSES Main Study 2019 survey instruments

Site	Language	SSES MS instruments						
		Student		Parent		Teacher		Principal questionnaire
		Direct assessment	Contextual questionnaire	Direct assessment	Contextual questionnaire	Direct assessment	Contextual questionnaire	
Bogota	Spanish	●	●	●	●	●	●	●
Daegu	Korean	●	●	●	●	●	●	●
Helsinki	English	-	-	●	●	-	-	-
	Finish	●	●	●	●	●	●	●
	Swedish	●	●	●	●	●	●	●
Houston	English	●	●	●	●	●	●	●
	Spanish	●	●	●	●	-	-	-
Manizales	Spanish	●	●	●	●	●	●	●
Moscow	Russia	●	●	●	●	●	●	●
Ottawa	English	●	●	●	●	●	●	●
	French	●	●	●	●	●	●	●
Sintra	Portuguese	●	●	●	●	●	●	●
Suzhou	Chinese	●	●	●	●	●	●	●
Istanbul	Turkish	●	●	●	●	●	●	●

6.4. National translation process

High quality translations are essential to the success of the Study. Therefore, it was important that participating Sites established a sound translation process to ensure that nationally translated/adapted versions of the SSES survey instruments were equivalent to the international source version.

6.4.1. The team translation approach

As outlined in the Adaptation Guide and the Translation Guide, Sites were required to adopt a team translation approach in which three translators and a local subject matter expert (or one of the translators could assume this role) worked as a team to translate the survey items. Each of the translators would translate two-thirds of the items independently, resulting in all items being translated independently by two translators.

After the initial translations, the lead translator would merge a joint draft that displayed the independent translations together. All three translators and the local subject matter expert would then collaborate on a mutually agreed final translation, reviewing the entire translation item by item and suggesting improvements. Any item where the three translators could not reach unanimous agreement would then be reviewed once again together with the Site Project Manager (SPM) acting as an internal mediator as needed. The reasoning behind the strength of committee approach was that consensus among bilinguals would produce more accurate text than the subjective opinion of a single translator. Additionally, by striving for consensus, problems of idiosyncrasies, culture and uneven skill in either language would be overcome.

If Sites prepared translations in more than one language, the International Contractor suggested that professionals familiar with the various languages be involved in order to ensure that the translations were equivalent across the national languages.

Sites were encouraged to hire highly qualified translators and reviewers who were well suited to the task of working with the SSES materials. Essential qualifications for translators included:

- excellent knowledge of English
- excellent knowledge of the target language(s)
- experience in the Site's cultural context
- experience in translating texts in the subject areas related to social and emotional skills.

During the translation and adaptation of the national survey instruments, translators, adaptors and SPMs were asked to ensure the following:

- The translated/adapted text avoided complicating or simplifying vocabulary and syntax. This applied to the text materials used in the stimulus, and the wording of the items (both in the question stem and in proposed responses).
- The translated/adapted text used terminology equivalent to that of the international source version. Conversely, everyday terms used in the international source versions were not to be translated as more 'technical', 'scientific' or contain 'literary expressions'.
- The translated/adapted text had the same level of difficulty and register (language level and degree of formality) as the international source version.
- The translated/adapted text used correct grammar, punctuation, qualifiers and modifiers, as appropriate for the target language(s).
- The translated/adapted text did not clarify or remove text from the source text, did not add more information, was similar in length and was at an appropriate level for the target population.

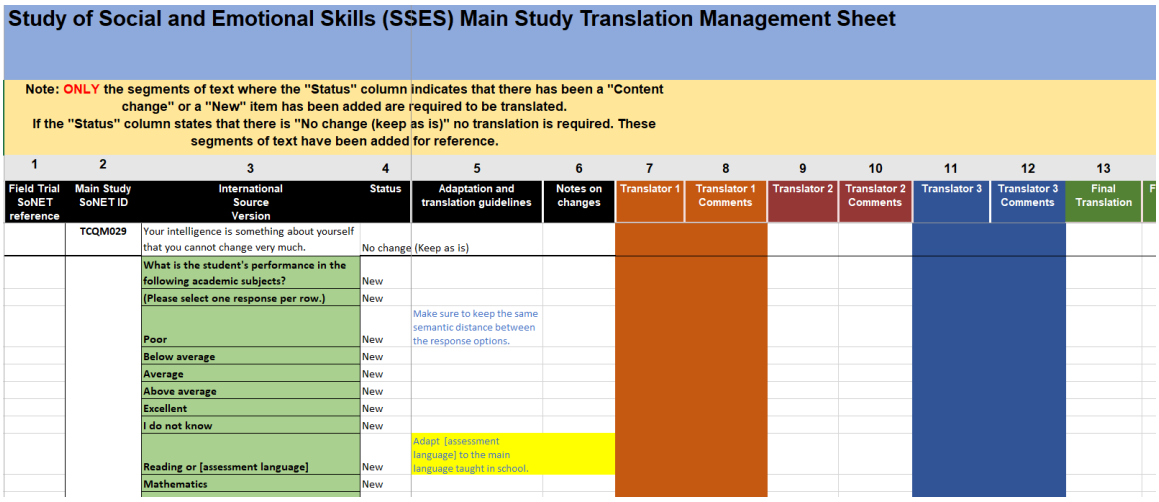
6.5. Development and use of the monitoring tool and the adaptation and translation guidelines

A number of instructional guidelines and monitoring tools were developed to assist Site Project Centres (SPCs) to streamline their processes and manage the translation of national survey instruments in a systematic and timely manner.

6.5.1. Translation Management Sheets

The Translation Management Sheet (TMS) was developed to support SPCs in managing and coordinating their team translation approach. The TMS had a built-in colour scheme which assigned a section for each translator. The TMS contained all the assessment items that required translation as well as adaptation and translation guidelines to assist SPCs in the translation process. Figure 6.1 is a screenshot of the TMS.

Figure 6.1. The Translation Management Sheet



Study of Social and Emotional Skills (SSES) Main Study Translation Management Sheet

Note: ONLY the segments of text where the "Status" column indicates that there has been a "Content change" or a "New" item has been added are required to be translated.
If the "Status" column states that there is "No change (keep as is)" no translation is required. These segments of text have been added for reference.

1	2	3	4	5	6	7	8	9	10	11	12	13
Field Trial SoNET reference	Main Study SoNET ID	International Source Version	Status	Adaptation and translation guidelines	Notes on changes	Translator 1	Translator 1 Comments	Translator 2	Translator 2 Comments	Translator 3	Translator 3 Comments	Final Translation
	TCQM029	Your intelligence is something about yourself that you cannot change very much.	No change (Keep as is)									
		What is the student's performance in the following academic subjects?	New									
		(Please select one response per row.)	New									
		Poor	New	Make sure to keep the same semantic distance between the response options.								
		Below average	New									
		Average	New									
		Above average	New									
		Excellent	New									
		I do not know	New									
		Reading or [assessment language]	New	Adapt [assessment language] to the main language taught in school.								
		Mathematics	New									

6.5.2. Adaptation and Translation Guides

For the Field Test (FT), two separate instrument preparation guides were produced, the Adaptation Guide and the Translation Guide. These guides provided comprehensive instructions about:

- how to implement the team translation approach to translating the survey instruments
- the role of the local subject experts
- how to avoid common translation and adaptation pitfalls
- how to adapt the survey instruments to the national context
- the international translation verification process
- how to use the online RM Assessment Master system.

For the Main Study, the two instrument preparation guides developed for the FT were merged to create the Instrument Adaptation, Translation and Translation Verification Guide. This single guide lessened the amount of repeated information and was further developed to support SPCs in their national translation process. The redeveloped guide included step-by-step instructions on how to complete the Site Adaptation Verification Form (SAVF) with respect to the adaptation, translation and international verification processes.

6.5.3. Site Adaptation Verification Form

Adaptations to the survey instruments and multiple levels of translation verification were carried out to ensure the best possible linguistic equivalence of the translated/adapted survey instruments across all participating Sites. Such modifications at various stages of the process were documented in the SAVF for each language and set of survey instruments. The SAVF was an Excel workbook with multiple worksheets containing the complete translation, adaptation and verification history of each set of national survey instruments. Each worksheet contained:

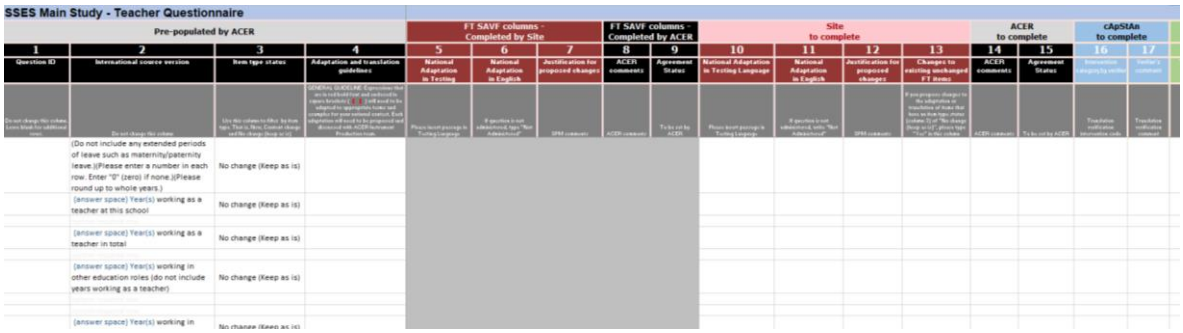
- the international source version of each item

- item-specific translation and adaptation guidelines (a set of notes that clarified the intended meaning of the item enabling translators to select the appropriate national term or expression necessary to convey the intended meaning)
- discussion and review notes of the adaptations used in the FT, which assisted the SPM and translators to assess suitability of the adaptation and, where an improvement to the adaptation was deemed necessary, proffered an alternative adaptation.

Figure 6.2. The Site Adaptation Verification Form

contains a screenshot of the SAVF.

Figure 6.2. The Site Adaptation Verification Form



During various stages of the instrument preparation process, sections of the SAVF were accessed and reviewed. While translating and adapting a set of national survey instruments, the first version of the SAVF was filled out by the SPM in collaboration with the translators. Throughout the process, the International Contractor, the international verifiers and SPMs updated and revised the SAVF after each round of international verifications.

Documenting an adaptation in the SAVF required entering the proposed adaptation in the target language, an English back-translation of the adaptation and a justification for any proposed changes to adaptations. For ease of use and documentation of the different stages of verification, the SAVF included designated areas for each item, respondent and instrument.

6.5.4. National interest questions

SPCs were permitted to add up to five national interest questions within each of the existing student, parent, teacher and principal questionnaires. All national items were placed at the end of the international questionnaires in order to avoid influencing responses to the international questions in any way. All submitted national interest questions needed to be reviewed and approved by the International Contractor before they could be included in the Site version of the respective questionnaires.

The SAVF provided an important record of each Site's final survey instruments, as it contained information used throughout the different stages of the translation and verification processes. The SAVF was referenced when adding national data to the international database and during data analysis.

6.5.5. RM Assessment Master

The RM Assessment Master (AM) translation management platform from RM Results (see Chapter 5 for further details) facilitated the integrated online translation verification process. Once the international source version of the items had been set up in AM, representatives from each SPC entered their final translations directly into AM.

AM enabled users to edit and format their translations while also keeping a full record of the editing history for the text content. It enabled all users to retrace changes made by other users thereby ensuring transparency of the translation process. In addition, AM displayed the differences between a previously saved translation and the current translation in the form of tracked changes. This feature supported SPCs with finalising translation processes.

6.6. International verification procedures

After the international instruments were adapted to the national context, translated and internally reviewed by the SPCs, the national versions of the instruments were submitted for external verification. This consisted of a rigorous four-part verification process: adaptation verification, translation verification, translation verification adjudication and layout verification.

6.6.1. Adaptation verification

SPMs were required to consult with the International Contractor to review all proposed national adaptations. In particular, they were strongly encouraged to discuss any adaptation that might result in a serious deviation from the international instruments. SPMs began completing the SAVF after reviewing the international source version of the survey instruments, and submitted the SAVF to the International Contractor for consultation. After reviewing the SAVF, the International Contractor provided the SPMs with feedback on their adaptations and, where appropriate, made alternative suggestions that aligned better with the source version. Some of the common issues identified during adaptation review were:

- inconsistent adaptations used within or across survey instruments
- difficulties in establishing country-appropriate adaptations for ISCED[1] levels
- deviations from adaptations used in the FT.

SPMs were requested to take the recommendations into account and update the SAVFs accordingly, so that these updated forms would be used during the translation verification process to evaluate the quality and accuracy of the translations.

6.6.2. International translation verification

International translation verification was carried out for all national versions in each target language. This process was managed by the International Contractor who enlisted the assistance of an external independent translation company, cApStAn LQC (Brussels).

The main criteria used by cApStAn LQC to recruit international translation verifiers were:

- native speakers with expert knowledge of the target language
- proficiency in English
- university-level education and (if possible) familiarity with the subject area

- residency in the target country, or close contact with the country and its culture.

Training of international translation verifiers was carried out by cApStAn LQC through web-based seminars, where verifiers received detailed instructions for reviewing the survey instruments and registering deviations from the international source version. International translation verifiers also received general information about the Study and design of the instruments, together with a description of the translation procedures used by the SPMs. The International Contractor supplied international translation verifiers with instructional materials to support their work.

6.6.3. International translation verifiers and their responsibilities

Each international translation verifier received the relevant manuals and instruments and a comprehensive set of directions, instructions and relevant examples for reviewing the national instruments and registering deviations from the international source version. cApStAn LQC also ensured all international translation verifiers received continuous training, and provided them with constructive ongoing feedback.

The international translation verifier's role was to:

- compare the target version in AM to the international source version, sentence by sentence and to correct any potential issues
- check whether the item-specific translation and adaptation guidelines listed in the SAVF were followed
- check whether the national adaptations agreed by the International Contractor were linguistically correct and accurately and consistently implemented.

Where applicable, international translation verifiers provided feedback in the SAVF for the SPM to follow up and make corrections to translations directly in AM. During international translation verification, some of the typical errors identified by the verifiers included mistranslations, omissions or additions of text, inconsistent translations and grammar. Any adaptations reported in the SAVF were also reviewed by the international translation verifiers, who were asked to comment on the adequacy of the adaptations. With the documented comments and suggestions from the international translation verifiers, SPMs were able to revise and improve their national versions. All comments that the international translation verifiers deemed to be deviations from the source (according to the criteria) were entered into the SAVF. Each deviation was allocated one of 14 codes that indicated to the SPMs the severity and type of deviation of the translated text from the international source version. Table 6.2 shows the 14 intervention codes used by the international translation verifiers.

6.6.4. Translation adjudication – post international translation verification review

At the completion of the international translation verification process, an external Translation Referee was engaged to review all international translation verifier feedback documented in the SAVF.

The Translation Referee's role was to:

- review all international translation verifier feedback as documented in the SAVF
- indicate (in the SAVF) any issues that may jeopardise equivalence and comparability, and that therefore required follow-up by the SPM

- assess whether potentially controversial linguistic adaptations needed to be endorsed by test developers.

The Translation Referee's feedback was documented and provided to SPMs in the SAVF. The SPM would then either make all the recommended changes in AM or initiate further discussion with the Translation Referee until a final agreement was made before making changes in AM.

Two rounds of test form content reviews were also carried out by the SPMs to ensure all issues recorded in the SAVF had been addressed and necessary changes were made in AM before signing off on the international translation verification process.

Table 6.2. cApStAn LQC's verifier intervention categories

Intervention category	Description
OK	**No intervention** is needed. The verifier has checked and confirms that the text element or segment is equivalent to source, linguistically correct, and – if applicable – that it conforms to an explicit translation/adaptation guideline. This category may also be used to report an appropriate but undocumented adaptation.
Added information	Information is present in the target version but not in the source version, e.g. an explanation between brackets of a preceding word.
Missing information	Information is present in the source version but omitted in the target version.
Matches and patterns	1. A ***literal match*** (repetition of the same word or phrase) or a ***synonymous match*** (use of a synonym or paraphrase) in the source version is not reflected in the target version. Most important: literal or synonymous matches **between stimulus and question or between question stem and response options**. 2. A ***pattern*** in multiple choice questions is not reflected in the target version (e.g. all but one option start with the same word, proportional length of response options).
Inconsistency	A recurring element across units (e.g. an instruction or prompt) is inconsistently translated.
Adaptation issues	An adaptation is an intentional deviation from the source version made for cultural reasons or to conform to local usage. An adaptation issue occurs when an adaptation would be needed but was not made, or when an inappropriate or unnecessary adaptation was made.
Register/wording issue	1. *Register*: difference in level of terminology (scientific term><familiar term) or level of language (formal><casual, standard><idiomatic) in target versus source. 2. *Wording*: inappropriate or less than optimal choice of vocabulary or wording in target version to fluently convey the same information as in the source. This category is used typically for vague or inaccurate or not quite fluent translations.
Grammar/syntax issue	1. *Grammar*: grammar mistake that could affect comprehension or equivalence, e.g. wrong subject-verb agreement, wrong case (inflected languages), wrong verb form. 2. *Syntax*: syntax-related deviation from the source, e.g. a long (source) sentence is split into two (target) sentences or two (source) sentences are merged into a single (target) one; or another syntactic problem due, e.g. to overly literal transition of the source.
Mistranslation	A wrong translation, which seriously alters the meaning. A mistranslation should always be reported with a back-translation. Note: a vague or inaccurate translation should rather be classified as a Register/wording issue (or sometimes Grammar/syntax) issue). This category covers cases where the source has been misunderstood, but also copy/paste errors that unintentionally result in a wrong text element or segment.
Guideline not followed	An explicit translation/adaptation guideline for a given text element or segment was overlooked or was not addressed in a satisfactory way.
Left in source language	A text element or segment that should have been translated was left in the source language.
Minor linguistic defect	Type or other linguistic defect (spelling, grammar, capitalisation, punctuation, etc.) that does not significantly affect comprehension or equivalence. Correcting such errors is usually not controversial and can be made in track changes without documenting them.
Erratum/update missed	An erratum or update notice has been overlooked.
Layout/format issue	A deviation or defect in layout or formatting: disposition of text and graphics, question labels, question numbering, styles (**boldface**, underlining, *italics*, UPPERCASE), legibility of captions, tables, number formatting (decimal separators, 'five versus '5'), etc. In computer-based materials, this includes truncated words in the preview, undesired scrolling, etc.

6.6.5. Layout verification of instruments

The final step in the verification process was layout verification. This commenced once all reviews of the translation of the items had been completed.

For SSES, layout verification mainly focused on:

- ensuring that formatting and spacing were adequately applied
- applying comparable text emphasis where needed

- assigning all data points correctly
- checking that adaptations were showing as intended
- confirming each Site's questionnaires appeared as agreed upon in the SAVF.

Other issues that were caused by irregularities in the translation system or that required specific intervention by the developer team were also adjusted during this phase and overall functionality of the instruments was checked.

All issues found during the layout verification of each Site's instruments were collated and provided to Sites in a layout verification report, including the actions taken by the International Contractor. Once SPMs responded to each issue listed, further layout review was implemented by the International Contractor to complete the layout verification process.

6.6.6. User Acceptance Testing

The final review step before live testing commenced was the User Acceptance Testing (UAT) of the test forms. SPMs were given access to their translated online test forms to review and check functionalities in their local online settings. UAT helped to ensure that all the required features, such as the language selection feature, navigation bars and item filtering, were appearing in the forms and functioning as intended. Once SPMs signed off on the UAT, all assessment respondents were assigned to the final test forms.

Notes

[1] *ISCED* is the reference classification for organising education programmes and related qualifications by education *levels* and fields. For further details, see http://uis.unesco.org/en/isced-mappings.

Chapter 7. Field operations

Successful administration of the assessments for the OECD Survey on Social and Emotional Skills 2019 (SSES) depended heavily on the contributions of the Site Project Managers (SPMs) and their Site staff. The administration of the assessment along with the overall coordination and logistical aspects of the Study, presented a set of significant challenges for each participating Site. These challenges were heightened by the demands of administering the SSES assessment online.

The SSES International Contractor therefore developed internationally standardised field operations procedures to assist the SPMs and to aid uniformity of their assessment administration activities. These procedures, outlined in this chapter, were designed to be flexible enough to meet both the needs of individual participants and the high quality expected by survey standards.

7.1. Overview of responsibilities

SSES was coordinated in each Site by a Site Project Manager (SPM) who implemented the procedures specified by the International Contractor at the Site level. Each SPM typically had several assistants, working from a base location that is referred to throughout this report as a Site Project Centre (SPC).

The number of staff members in the SPCs varied from one Site to another, depending on the Site size and how it chose to organise the national data collection work. Some SPMs worked with external survey organisations to conduct operational tasks. These organisations were appointed on a tendering and contracting basis.

For the school level operations, the SPM coordinated activities with school level staff, referred to as School Coordinators (SCs). Trained Study Administrators (SAs) administered the SSES assessment in schools.

7.1.1. Site Project Managers

SPMs were responsible for implementing SSES within their own Site. Detailed information about the roles and responsibilities is documented in the Site Project Manager's Manual. In summary, SPMs:

- established overall preparation and administration schedules in cooperation with the International Contractor
- attended SPM meetings and received training in all aspects of SSES operational procedures
- maintained clear and effective communication with the International Contractor and OECD
- established procedures for the security and confidentiality of materials during all phases of the implementation
- prepared a series of sampling forms documenting sampling-related aspects of their Site educational structure and system, as well as the school sampling frame for school sample selection
- recruited three experienced professional translators to produce the Site versions of the source instruments
- coordinated the translations and preparation of the Site versions of the assessment items, questionnaires, operational manuals and forms, Study Administrator Script, and coding guides
- reviewed and finalised the translations of all assessment instruments and online forms in the RM Assessment Master (AM) translation management platform
- contacted all sampled schools to confirm participation, then identified SCs from each of the sampled schools and worked with them on school preparation activities
- verified that the technical suitability of the devices and the network capacities for online assessment delivery in sampled schools were checked
- nominated suitable candidates to work with the International Contractor as Quality Monitors to observe the test administration in a selection of schools (for both the Field Test and the main study)

- selected the student sample using the specified tool from a list of eligible students in the Student Teacher Linkage Form (STLF) provided by the SCs
- completed User Acceptance Testing (UAT) of the localised and translated online test forms for each type of instrument
- uploaded the Student and Teacher linking information to AM to activate all the online forms and reports
- scheduled the assessment sessions
- recruited and trained SAs according to the Technical Standards for SSES, to administer the assessment within schools
- monitored participation of students, parents, teachers and principals during the administration period and organised follow-up sessions with schools where required and possible
- collected completed Student Grades Form (SGF) from schools
- coordinated coding of the occupation data
- submitted all required data to the project portal and notified the International Contractor
- completed the online Field Test (FT) and Main Study (MS) Feedback Forms reviewing project implementation processes and procedures.

7.1.2. School Coordinators

The SPMs identified and trained the SCs for all participating schools. The SCs acted as the main contact person within schools. They were responsible for preparing the schools for the successful implementation of the SSES assessment.

In this role, SCs:

- liaised with the SPC in regard to all aspects of their school's participation in the Study
- identified and prepared the eligible student list for their school with the required details and sent it to the SPM for within-school student sample selection
- identified eligible teachers for all sampled students and listed them in the STLF
- appointed an IT Coordinator within the school to provide technical support for the preparation and delivery of the online assessment session
- confirmed the assessment dates and time in consultation with the SPM and SAs
- ran, in coordination with the IT Coordinator, the Technical Readiness Tests on all devices to be used for the assessment, and created URL shortcuts to the SSES testing site on all devices being used for the assessment and completed the School Local IT Resources Form
- confirmed the login procedures with the IT Coordinator
- ensured that the assessment materials received were kept in secure places and confidential at all times (if applicable)
- informed the school community (principals, teachers, students, parents, and other school staff) about the purpose of SSES and the assessment schedule

- distributed consent forms to parents (if needed)
- liaised with all staff members who needed to respond to the teacher and principal questionnaires
- prepared assessment administration in the school (arranged suitable rooms, updated the Student Tracking Forms with students' attendance, etc.)
- distributed instructions and login details to parents, teachers and principal to complete their corresponding online questionnaires
- received and distributed the paper version of the parent questionnaires (if applicable)
- ensured that the paper parent questionnaires were returned in accordance with the schedule (if applicable)
- completed the SGF and the School Coordinator Report Forms after the assessment session.

SCs were expected to provide assistance to the SA on the assessment day to ensure that sampled students attended the online assessment session and that any technical issues were quickly resolved with the help of the designated IT Coordinator.

7.1.3. Study Administrators

The main responsibilities of the SAs were to administer the SSES online instruments under suitable testing conditions in sampled schools, in accordance with the procedures and standards outlined in the Site Project Manager Manual. To maintain fairness, it was preferred that an SA could not be a staff member at any participating school or, at least, could not be a teacher or instructor of any of the students to which the instruments were administered.

Prior to the date of the test administration, SAs were trained in person by the SPM or authorised staff. Training included a thorough review of the Study Administrator Manual and demonstration of the student online assessment system.

The SA training comprised, but was not limited to, the following topics:

- introduction to SSES, explaining what the results are used for in the local context
- discussion about the security of materials at all times of Study administration
- description of the role of SAs, emphasising the importance of uniform testing conditions in an international context
- review of SAs' activities before, during and after assessment
- review of the Student Tracking Forms and procedures for their completion, including some practice examples
- review of the Study Administrator Script, emphasising the importance of adherence to the wording of the script and to session timing
- review and discussion of issues which commonly arise
- review and discussion of any nationally specific issues such as protocols for entering schools or communicating with school staff
- demonstration of the online questionnaire and student login procedures.

Additional responsibilities of the SAs included:

- confirming administration dates with the SPM and SC
- confirming the final assessment plan for the administration day with the SC, especially for the login procedures of the devices to be used
- preparing all assessment materials, such as the Student Login Form, to hand out to the students on the assessment day
- reviewing and updating the Student Tracking Form
- completing the Study Administrator Report Form after each assessment session
- returning the required materials to the SPM
- organising a follow-up assessment session, if required by the SPM.

7.2. Survey operations manuals

During the SSES MS implementation, the SPCs were required to adhere to the standardised procedures prepared by the International Contractor. These procedures were outlined in the following documents released to the SPCs prior to the FT and MS:

- **Site Project Manager Manual:** provided an overview of SSES, description and details of the SPM's tasks, timeline of activities the SPCs had to carry out, phases of the project implementation and information about key milestones and deliverables.
- **School Coordinator Manual:** described in detail the role and responsibilities of the SC. The manual included instructions on how to prepare a student list and complete the STLF which were required for within-school sampling procedures. Another version of the School Coordinator Manual was developed specifically for the census Sites due to a slightly different approach in preparation of the STLF.
- **Study Administrator Manual:** described the role and responsibilities of the SAs in detail. Additionally, the manual included the Study Administrator Script with detailed instructions for the SAs on how to conduct the assessment sessions to ensure consistent delivery in all schools.
- **Working with School Guide:** paid special attention to the SPCs' relations with local partners and schools. This guide focused on ensuring that context-appropriate strategies were followed in order to promote the Study to sampled schools and staff. It described how to develop strategies for successful communication with schools, obtain school cooperation and prepare school principals, teachers, SCs, students and parents for their involvement in SSES, which are all key points in obtaining high participation rates in order to collect high quality data. Additionally, the guide included examples of letters to school principals and teachers, and the Study overview for distribution to key stakeholders.

The International Contractor provided the English version of all operational manuals updated for the MS. SPMs were responsible for translating the School Coordinator Manual and the Study Administrator Manual into the language(s) in which they administered the assessment unless all SCs/SAs were fluent in English, and for making national adaptations or adding national information where necessary. Since there was no international translation verification of the School Coordinator Manual and the Study Administrator Manual (except for the Study Administrator Script), it was the SPM's responsibility to produce a high-quality translation (where applicable) of the manuals with respect to administration procedures, timelines and contact details.

The Study Administrator Script was the only part of the Study Administrator Manual that underwent the translation verification process in AM.

7.3. Within-school sampling

Once the school sample was drawn by the International Contractor, SPMs contacted all the sampled schools requesting a list of students that were eligible according to the age definition stipulated in the Technical Standards, which was then used to select the within-school student sample.

The within-school sampling tool was developed by the International Contractor for SSES. SPMs were required to use this tool for the selection of their Site's student sample. The tool generated the list of sampled students for each school, titled the Student Teacher Linkage Form (STLF). The STLF was then given to schools to identify the teacher who was most familiar with each sampled student.

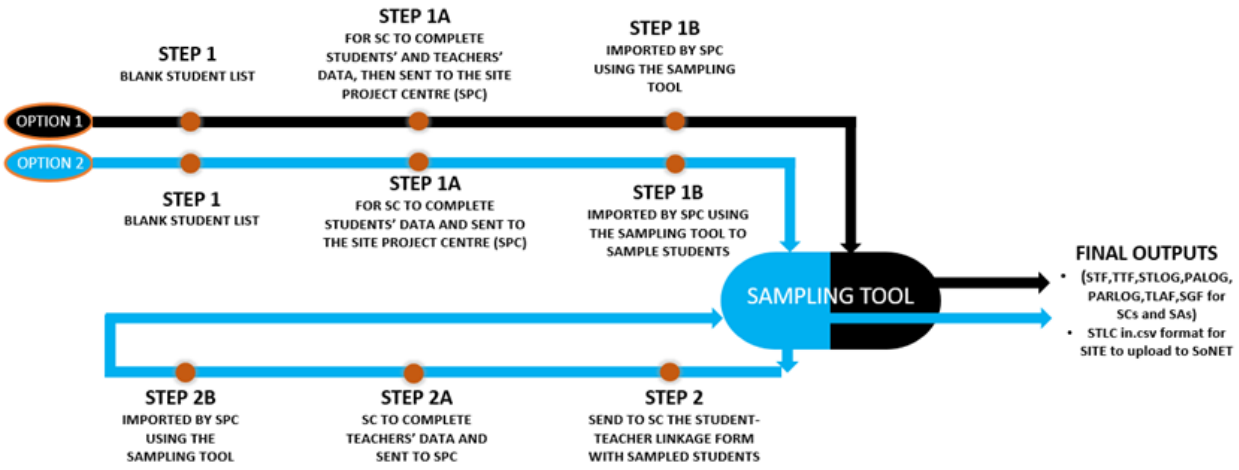
7.3.1. Student and teacher linkage

For the MS, the within-school sampling tool provided two options for SPMs to manage the student and teacher linking process. SPCs could either request sampled schools to identify eligible students and link teachers at the same time, or identify eligible students first, draw the student sample, then identify teachers for those sampled students only. The latter option required a two-step process:

- Step one:
  a) SC provided an eligible student list from their school
  b) SPM selected the student sample using the provided tool, then provided the sampled student list in the STLF to the school.
- Step two:
  a) SC identified teachers for each sampled student in the STLF, and returned the STLF to the SPC
  b) SPM imported the returned STLF into the Tool to generate tracking and login forms
  c) SPM uploaded the output file to AM.

The census Sites followed the one-step process (Option 1), and the non-census Sites followed the two-step operation (Option 2). Figure 7.1 shows the flowchart for how the STLF was generated.

Figure 7.1. Student-Teacher Linkage Form preparation flowchart



A series of checks were required to be undertaken as each within-school sample was drawn to ensure the correct outputs had been generated without error.

7.3.2. Materials dispatch to the School Coordinators

The SPC provided each SC with the following key documents separately for each cohort respectively:

- Student Tracking Form
- Teacher Tracking Form
- Teacher Login and Assignment Forms and cover letters
- Principal Login Form and cover letter
- Parents Login Forms and cover letters
- Parent questionnaires (if applicable)
- Student Grades Form
- School Coordinator Report Form.

The SCs were responsible for the security and safe storage of all materials prior to the SSES assessment. The cover letters, login forms and teacher assignment forms were distributed to the sampled parents/caregivers, teachers as indicated on the tracking forms and the school principal.

7.3.3. Assessment materials dispatch to the Study Administrators

The following key documents were sent to the SA:

- Study Administrator Script in language of SSES administration
- Student Login Forms
- Student Tracking Form for each allocated session
- Study Administrator Report Forms
- Study Administrator Manual.

7.4. Online Study administration

On the assessment day, the SA was required to arrive at the assigned school early to:

- review the Student Tracking Form with the SC and update the forms as necessary
- discuss all specific issues related to the assessment administration
- set up the assessment room and check the devices to be used, as described in the Study Administrator Manual.

7.4.1. Prepare for the Study administration

Before the start of the assessment session, the SA was expected to have:

- arranged an appropriate workspace and device for each participating student
- set up the devices for use:
  a) either have unlocked the device or have clear instructions for unlocking
  b) have created a URL shortcut to the SSES testing site (if possible), or
  c) have the login page open on the screen (if possible)
  d) have muted the device so stray audio files did not interrupt the session
- distributed student login slips in front of each device and ensured that the corresponding student would use their assigned login.

If neither option 'a' nor 'b' were possible, then the SA had to write down the URL to the SSES assessment site on a whiteboard, so it was clearly visible to all students.

7.5. Administering and monitoring the assessment

SAs were required to strictly follow the instructions described in the Study Administrator Manual to achieve comparable data and ensure standardised assessment procedures across participating Sites. The student assessment and student contextual questionnaire were administered in a single online session. Adjustments to the timing of the session were allowed. The guidelines for the session timing are shown in Table 7.1.

Table 7.1. Timing of the SSES administration

Activity	Expected Time (on average)
Logging in	5 - 10 minutes
Instructions	10 minutes
Administering the student assessment and student contextual questionnaire	60 minutes
Conclude the assessment session	5 -10 minutes
Total	1 hour 30 minutes

SAs were required to read out the Study Administrator Script verbatim to ensure that standardised assessment procedures were followed across all participating Sites. Students who arrived within 15 minutes after the assessment session had started were allowed to undertake the assessment. However, students were not allowed to leave the session unless it was absolutely necessary.

During the assessment session, SAs were not allowed to help the students with the assessment items by defining or explaining any of the words or items to the student. But SAs could help students resolve any technical issues affecting the online assessment.

At the end of the online administration, SAs closed the browsers for students who were still logged into the assessment form. Before dismissing the students, SAs collected all the login slips for secure destruction. SAs were required to complete the Student Tracking Form with the student attendance details, as well as the Study Administrator Report Form after each session, and return them to the SPM.

Once the assessment session had concluded, the SAs were required to:

- make sure the STF was completed correctly and students' attendance was recorded as required
- complete the Study Administrator Report Form and return it to the SPC.

7.5.1. Material receipt at the Site Project Centres

The major tasks for SPMs immediately after data collection included retrieving and collating the materials from schools and SAs and confirming the integrity of the materials by:

- verifying all identification data on all materials were accurate and legible
- checking that the participation status was recorded on the Student and Teacher Tracking Forms
- following up on any unreturned assessment materials.

7.6. Participation monitoring and reporting

During the administration period, SPMs were required to monitor the participation rates of all respondents on a regular basis to ensure that their Site would meet the requirements set out in the Technical Standards.

The International Contractor provided the SPM with access to AM Manager where SPMs could download 'participation reports' showing each participant's completion status. SPMs also had access to a Site information dashboard which showed progress updates in real time. This enabled SPMs to monitor the questionnaire completion rate and schedule follow-up administration sessions to improve their response rate, as needed.

At the end of the online assessment administration period, SPMs were required to download the final student participation reports from AM Manager and, together with the STF, to complete the data validation process. It was also anticipated that SPMs would download the occupational data for students and parents, then initiate the coding process.

7.7. The Main Study feedback

At the end of the MS data collection period, SPMs were required to complete an online feedback form to review their entire MS implementation processes and procedures. The Main Study Feedback Form provided an opportunity for SPMs to critically review various aspects of the project implementation of SSES, and to suggest areas for improvement. The Main Study Feedback Form was also part of the overall quality assurance mechanism that allowed the SPMs to formally document the project implementation details at the Site level, including details on communications, school sampling and within-school sampling, field operational procedures, translation verification processes, online test administration,

security arrangements and so on. The Main Study Feedback Form was submitted to the International Contractor six weeks after the submission of the Site data files.

Chapter 8. SSES quality monitoring

This chapter describes the key quality assurance procedures that were established and implemented throughout the OECD Survey on Social and Emotional Skills 2019 (SSES) to ensure that the SSES data were collected in accordance with SSES Technical Standards. The activities included and described in this chapter are: consultation and negotiation of key project implementation tasks; Field Test (FT) and Main Study (MS) feedback; and FT and MS Quality Monitor (QM) responsibilities.

8.1. Overview

To ensure that SSES data were collected in accordance with SSES Technical Standards across all participating cities, quality assurance procedures were implemented throughout the Study with a two-step approach: first, setting up quality control procedures to guide the data collection process; and second, monitoring and documenting how the procedures were implemented.

Quality control was set up through systematic and effective procedures to guide the various aspects of the implementation process. Quality monitoring activities were also built into the implementation process to monitor, rectify potential concerns early and document any deviations from the specified procedures during the Study implementation. Key quality monitoring activities included:

- consultation and negotiation of key project implementation tasks
- Field Test (FT) and Main Study (MS) feedback
- FT and MS Quality Monitor (QM) visits.

8.2. Consultation and negotiation of key project implementation tasks

The structured procedures and instructions in manuals described what was expected to occur during project implementation. Site Project Centres (SPC) were also required to document detailed information in a specified template or by email to record their plans for implementing the procedures and their progress at various stages and submit this documentation to the International Contractor for agreement. Through this review process, the International Contractor was able to clarify any misunderstandings about procedures with cities, negotiate and agree on a plan and, most importantly, address potential major issues.

These practices were implemented in all major areas of the SSES survey implementation, namely, school sampling, instrument adaptation, translation verification, within–school sampling, operational manual adaptations and online assessment form finalisation.

8.3. Field Test and Main Study feedback

Six weeks after the data collection period of the FT and the MS, participating cities were required to review all aspects of their project implementation activities and provide structured feedback to the International Contractor through an online feedback form. The FT and MS Feedback Forms covered the following aspects of the project implementation:

- *Key documents and processes* – Using a satisfaction rating system to review project key documents and provide feedback
- *Communication with the International Contractor* – Review of the project portal, management process, communication quality on key activities and training
- *School sampling* – Review of the school sampling process
- *Within-school sampling* – Review of the within-school sampling tool and process
- *Contacting schools* – Review of the processes for contacting schools and appoint a School Coordinator (SC)
- *Documenting and implementing national adaptations* – Review of the adaptation process

- *Translation and verification* – Review of the translation and verification process
- *MS administration* – Review of the Study Administrator (SA) recruitment and MS administration procedures
- *Online administration of instruments* – Review of the online delivery platform including finalising the online assessment forms, online delivery, monitoring reports, data downloads and data entry
- *Quality monitoring* – Review of QM nomination, communication with PQM coordinator regarding the QM visits
- *Coding occupational data* – Review of the occupational data coding process
- *Data validation and submission* – Review of the data validation and submission instructions and process
- *Security arrangements* – Review of the security requirements and confirmation of implementation.

After carefully reviewing the feedback data received after the FT, a number of revisions were carried out to improve the project implementation procedures and processes for the MS, including:

- further enhancements to the online assessment forms by adding more user-friendly features and functionalities
- further customised interface with more reports for Site Project Managers (SPMs) to monitor respondents' participation status during the testing period
- additional options and functions developed in the within-school sampling tool to accommodate both census and sampling approaches
- restructure and revision of the project manuals to provide more streamlined and concise guidelines and instruction.

Both FT and MS feedback data were collated and reviewed. A summary report with the International Contractor's responses to issues raised was provided to all participating cities.

8.4. Quality Monitor school visits

For the SSES data collection period, the International Contractor organised QM visits during both the FT and MS, to ensure that SAs administered the assessment sessions following standardised procedures across all participating cities.

The role of the International Contractor was to oversee and manage all aspects of the QM school visits in the participating cities, including developing manuals and data collection instruments, interviewing and hiring QMs in each of the cities, organising their training, selecting the schools to visit, collecting information from the QM visits and processing their payments.

8.5. Selection and training of Site Quality Monitors

All SSES QMs were nominated by the SPM through a formal process to the International Contractor. The final selection was based on the candidates' resumes and the interview questionnaire they had completed. The suitability of a candidate was determined taking into account whether they were independent from the SPC, were familiar with online

testing environments, had a background in education, assessment or research, were fluent in the testing language and able to communicate in English effectively.

For the FT, one QM was hired for each Site to visit five schools during the testing period. For the MS, one or two QMs were hired for each Site, and they visited a total of 10 schools.

The International Contractor provided QMs with online training materials. The Quality Monitor Manual, Quality Monitor self-training package, the Study Administrator Manual and Study Administrator Script, the School Coordinator Manual and Quality Monitor Data Collection Forms were made available to all QMs following their appointment. Independent self-training involved reading the materials provided and going over the QM step-by-step training PowerPoint presentation. Each QM was required to attend one SA training session conducted by the SPM to gain a more detailed understanding of the Study Administration procedures and also to observe the training session, then file a report back to the International Contractor.

Due to privacy regulations, the International Contractor randomly selected the schools and replacement schools using school ID only, then sent a list of selected schools to the SPM with instructions on how to prepare the final school list. The SPM and the QM collaborated to develop a schedule of school visits. Once the school list was finalised, the SPM provided contact details of the selected schools to the QM. The QM sent the final school list with school IDs and assessment dates to the QM coordinator for final sign off. Once approved, the QM made contact with the schools and organised their school visits.

Throughout this process, the QM coordinator from the International Contractor provided support and addressed any issues or concerns via email. QMs were contacted at each stage to ensure they were well prepared and ready for the school visits.

8.5.1. Quality Monitors' responsibilities

The role of the QM consisted of observing SA training and student assessment sessions and documenting their observations in structured data collection forms during those sessions. The QM provided reports on the following aspects:

- attending and observing the SA training session
- monitoring the student assessment session
- interviewing the SCs.

Attending and observing the Study Administrator training session

Attending the SA training session allowed the QMs to learn more about the Study administration procedures, and also provide feedback on the SA training session based on their observations. The report data from the QM evaluation specified important details about the training session, such as the degree to which the training materials and procedures were appropriate and in line with all Study specifications and standards.

Most of the SPMs organised the SA training event themselves and delivered the session in person. During the training, the SPM informed the SAs that a QM would observe the assessment sessions in a small number of schools.

Summary of findings

Overall, SA training was conducted effectively for all cities. Most of the cities provided high quality translations of the manuals and training materials. Some of the cities made the PowerPoint presentation available for the SAs and SCs after the training. All SA training

allowed time for the SAs and SCs to ask questions and the SA training providers were able to provide the answers.

Monitoring the student assessment session

The role of QMs was to observe assessment administration activities in a sample of schools. Their primary responsibility was to document the extent to which assessment procedures in schools were implemented in accordance with Study administration procedures.

The QMs used the Quality Monitor Data Collection Form to record their observations of each assessment administration session during their school visit. The form covered the following areas:

- assessment session date
- preparation for the assessment
- conducting the assessment
- managing the students
- deviations from the Study Administrator Manual and script of assessment procedures
- assessment environment
- general questions concerning the assessment.

Summary of findings

Based on the data collected during the MS assessment sessions visited by the QMs, it was reported that:

- the device set up time before the session took more than 15 minutes in more than half of the sessions visited, mainly due to the lack of advanced preparation and thorough checking of the devices by the SC; only a small number of cases were due to unexpected technical difficulties on the day
- the majority of the SAs read out the Study Administrator Script verbatim and the noted deviations were mostly minor; major deviations documented related to paraphrasing part of scripts and omission of the 'introduction to the Study' section
- in a small number of sessions, the student login process took longer due to slow internet connections
- most of the assessment areas were considered suitable for online assessment, and only in a couple of sessions were students tested in inadequate space where students were sitting too closely together
- the majority of the students observed complied with the session timing, followed instructions given, worked independently and appeared to be doing their best to complete the assessment
- the majority of the SAs observed were familiar with the Study Administrator Script, able to apply session timing correctly, competent in setting up the devices and supporting students, able to use the Student Tracking Form (STF) and appeared to have a positive attitude towards the Study and the school staff
- most of the SCs observed on the assessment day were familiar with the purpose of the Study and their role and responsibilities, able to organise and manage the online

assessment, confident in interpreting and using the STF and appeared to be enthusiastic about their school participating in SSES.

Interviewing School Coordinators

The purpose of the interview with the SC was to obtain information regarding their involvement in various aspects of the assessment, gain additional background information and record any difficulties they may have experienced.

Most QMs interviewed the SC after observing the assessment session, but in some cases the interview was carried out beforehand at the SC's request. The interview took approximately 10-15 minutes.

Summary of findings

- *Preparation prior to the Study administration*: a small number of the SCs did not receive the School Coordinator Manual and detailed information about SSES.
- *General impressions*: overall, most of the SCs were well organised and prepared for the assessment. They were able to get eligible students to participate in the assessment, organise a suitable room for the administration of the assessment, ensure the devices to be used passed the Technical Readiness Test in advance, arrange for the SA and QM to enter the school, assist the SA during the assessment session and get the parents to participate in the survey. In some instances, the SC still found it difficult to fulfil some of their responsibilities.

8.6. Data submission

The QMs were provided with access to an online platform to submit their observation reports for each school visit. The platform enabled the data to be delivered to the International Contractor immediately upon submission by the QM. The observation reports were collated and used to check assessment administration procedures were followed in each school in accordance with the Technical Standards, and also that they had been referred to during the data adjudication process as required.

Chapter 9. Survey weighting

This chapter describes the process used to develop and calculate the survey weights for students who participated in the OECD Survey on Social and Emotional Skills 2019 (SSES) Main Study. The method used is closely related to the sample method, as set out in Chapter 4. Weights for schools, teachers and parents were also derived from these student level weights, and will be explained later in this chapter.

9.1. Overview

The sampling process involved a two-stage stratified cluster sample. In the first stage, cities for each cohort were divided into explicit strata, based on some common characteristics. Schools were sampled independently by cohort and explicit stratum, with probability proportional to size (PPS), meaning that the probability of selection was higher for schools with higher student enrolments. In the second stage, a number of students indicated by the Target Cluster Size (TCS) were selected from each participating school. If a school had a number of eligible students smaller than the TCS, then all students in the school were chosen to participate. Selected students then participated in the SSES survey, and their teachers and parents were also involved in responding to the questionnaires. The survey weights represented the selection probabilities of the schools and students within the schools, as well as the adjustments for school and student non-response.

In analysing SSES data, it is important to use survey weights in order to calculate appropriate estimates of sampling error and to make valid estimates and inferences about the population. Survey weights permit users to make approximate unbiased estimates of standard errors, conduct significance tests and create confidence intervals that appropriately consider the complex sample design of SSES in each individual participating Site.

While the students included in the final SSES sample for a given Site were chosen randomly, selection probabilities of students vary. Survey weights must be incorporated into the analysis to ensure that each sampled student appropriately represents the correct number of students in the population of eligible 10-year- or 15-year-old students at each particular Site.

There are several reasons why survey weights vary among students at each Site:

- A school sample design may intentionally over- or under-sample certain segments of the school population. For example, a relatively small sub-population using a particular language of instruction might be over-sampled to obtain more reliable estimates for that sub-population to facilitate a comparative analysis. Or, very small or geographically remote schools might be under-sampled due to the relatively larger operational costs and burden associated with implementing the survey at these schools with relatively few students.
- Available information about school size at the time of school sampling may not have been completely accurate, or may be somewhat out of date, once the students from the school are actually selected to participate.
- School non-response may lead to an under-representation of students from certain parts of the population from which this school was sampled, which requires the application of weight adjustments.
- Student non-response within participating schools may occur to varying degrees across sampled schools. Here, a weight adjustment needs to be applied to participating students to adjust for the loss of sampled students who were eligible but did not actually participate in the assessment for different reasons (such as absences or refusals).
- Trimming the survey weights to prevent undue influence of a relatively small subset of the school or student sample is sometimes necessary if a small group of students would otherwise have much larger weights than average students from the same segment of the population from which they were sampled. Overly large survey weights may lead to the estimation of larger sampling errors and inappropriate

representations of sub-groups in national estimates. Trimming survey weights introduces a small bias into the estimation process but is efficient in reducing overall standard errors (Kish, 2004[1]).

The procedures used to derive the survey weights for SSES are drawn from the procedures used in other international studies of educational achievement, such as the *Programme for International Student Assessment* (PISA) (OECD, 2017, pp. 116-125[2]). The next section details the calculation of final weights for the reporting of SSES results.

9.2. Calculation of the weights

The statistics produced for SSES were derived from data obtained through samples of schools and students. For these statistics to be meaningful for each Site, they needed to reflect the population from which they were drawn and not merely the sample used to collect them. The process of drawing conclusions from the sample data to elicit information about the underlying population the sample represents is called inference.

The inference weight or final weight allows the generation of Site level estimates from the observed sample data. The final student weight indicates the relative contribution of that unit to the estimated outcomes of the survey.

The final student weight is the product of a design or base weight and of one or many adjustment factors. The former is the inverse of the selection probability and the latter compensates for random non-response and other random occurrences that could, if not accounted for, introduce bias in the estimates. These design weights and adjustment factors are specific to each stage of the sample design.

The following conventional notations are used in this chapter:

- The sample for each participating cohort in the Site was selected separately and independently from H explicit strata and the index $h = 1, \ldots, H$ points to the explicit stratum. If no explicit strata were defined, then $H = 1$.
- In each explicit stratum, a sample of size n_h schools was drawn from the N_h schools forming stratum h. The index $i = 1, \ldots, n_h$, therefore, points to the i^{th} sampled school in stratum h.
- Each school $i = 1, \ldots, n_h$ within the explicit stratum h has a measure of size (MOS)[1] noted as MOS_{hi}; the sum of the individual measures of size for a stratum is given as MOS_h.
- In each responding school, a sample of m_{hi} students was drawn from a list of eligible L_{hi} students from school i in stratum h. The within school sample size was denoted the TCS. If there had been no changes in the enrolment statistics for the school since the creation of the sampling frame, then $L_{hi} = MOS_{hi}$, but this was seldom the case.
- If the selected school was large enough, m_{hi} = TCS = 50 by design.[2] The index $j = 1, \ldots, m_{hi}$ points to the students and m_{hi} can, therefore, differ from 50 if local conditions dictated that the sample size should differ. For example, if the size of the listing was $L_{hi} = 35$, then all students were selected and $m_{hi} = 35$.

9.2.1. School base weight

The first stage of sampling in SSES consisted of drawing the sample of schools. Schools were sampled with PPS probability, $ProbP_{hi}$. A school base weight was calculated to

represent this first stage of sampling. If a census sample of schools was implemented in a Site or in an explicit stratum of a Site, then the school base weight is set to one.

Using the above notation, the school base weight, ***W1***, for each school ***i = 1, …, n_h*** and each explicit stratum ***h = 1, …, H***, can be defined as:

$$W1_{hi} = \frac{1}{ProbP_{hi}} = \begin{cases} \dfrac{MOS_h}{n_h \times MOS_{hi}} & If\ MOS_{hi} < \dfrac{MOS_h}{n_h} \\ 1 & Otherwise \end{cases} \qquad \text{Eq. 9:1}$$

In those cities and cohorts where all schools were selected (i.e. n = N), ***$W1_{hi}$*** = 1 for all ***i = 1, …, N_h***. These cases include the older cohorts of Helsinki, Houston and Ottawa and both cohorts in Manizales and Sintra.

In addition, for all the cases where a substitute school participated, the value of ***$W1_{hi}$*** corresponded to the base weight of the originally sampled school for which the substitute occurred.

9.2.2. School base weight trimming factor

The school trimming factor (***$T1_{hi}$***) is the ratio of the trimmed to the untrimmed school base weight, and for most schools (and therefore most students in the sample) this factor is equal to one. The school-level trimming adjustment was applied to schools that turned out to be much larger than had been assumed at the time of school sampling.

Schools were flagged in cases where the enrolment in the target population, as indicated by ***L_{hi}*** the number of listed eligible students used to draw the student sample, exceeded ***$3 \times MAX\ [TCS_h, MOS_{hi}]$***. For example, for a TCS of 50 students, a school flagged for trimming would have had more than 150 (= 3 x 50) SSES-eligible students, and more than three times as many students in the listing as was indicated on the school sampling frame.

Because the student sample size was set at TCS regardless of the actual enrolment, the student sampling rate was much lower than anticipated during the process of sampling these schools. As a consequence, the weights for the sampled students in these schools would have been more than three times greater than anticipated when the school sample was selected.

To adjust for this, these schools had their school base weights trimmed by having ***MOS_{hi}*** replaced by ***$3 \times MAX\ [TCS_h, MOS_{hi}]$*** in the school base weight formula. This means that if the sampled students in the school would have received a weight more than three times larger than expected at the time of school sampling (because their overall selection probability was less than one third of that expected), then the school base weight was trimmed so that such students received a weight that was exactly three times as large as the expected weight. The School Base Weight Trimming Factor, ***T1***, for each school was calculated as:

$$T1_{hi} = \begin{cases} \dfrac{MOS_{hi}}{3 \times MAX[TCS_h, MOS_{hi}]} & if\ L_{hi} > 3 \times MAX[TCS_h, MOS_{hi}] \\ 1 & Otherwise \end{cases}$$ Eq. 9:2

The choice of a value of three as the cut-off point for this procedure was based on prior experience where the need to avoid variance inflation due to weight variation not related to oversampling goals, had to be balanced against the necessity not to introduce any more substantial bias by altering student weights. There were only very few cases of trimming school weights in some of the cities while in most cities school weight trimming was not applied at all.

9.2.3. School non-response adjustment

Despite all efforts to secure the participation of all selected schools, some of these were unable or unwilling to participate. In these cases, population segments not represented due to non-participating schools had to be represented by other participating schools. To adjust for non-participation at the school level, a non-response adjustment factor is calculated for each explicit stratum.

It is important to note that a participating school is one for which at least 25% of the sampled students participated in the survey. For each explicit stratum ***h = 1, …, H***, the school non-response adjustment factor was:

$$SCNRA_h = \frac{\sum_{i \in \Omega(h)} W1_{hi} L_{hi}}{\sum_{i \in \Gamma(h)} W1_{hi} L_{hi}}$$ Eq. 9:3

The sum in the denominator relates to ***Γ(h)*** which is the group of all schools within the explicit stratum (main and substitutes) that participated. The sum in the numerator refers to ***Ω(h)***, which represents those same schools plus the original sampled schools that refused and were not substituted.

The numerator provides an estimate of the population of SSES-eligible students in the stratum, while the denominator reflects the size of the population of SSES-eligible students directly represented by participating schools. The school non-response adjustment factor ensures that participating schools are weighted to represent all students in the stratum. If a school did not participate because it had no SSES-eligible students enrolled, no adjustment was necessary since this was not considered to be non-response or part of the population of eligible students.

9.2.4. Student base weight

The term $W2_{hij}$ is referred to as the student base weight for student (***j***) within participating school (***i***) in stratum (***h***). Because the students from a school were sampled with equal probability (see Chapter 4), their base weights are the same within each sampled school. This weight is calculated as:

$$W2_{hij} = \frac{L_{hi}}{m_{hi}} \quad \text{Eq. 9:4}$$

Here, L_{hi} is the actual enrolment of SSES-eligible students in the corresponding cohort in the school on the day of the assessment (and therefore tends to be different from the school enrolment recorded on the sampling frame), and m_{hi} is the sample size within school ***i*** in stratum ***h***. It follows that if all SSES-eligible students from the school were selected, then $W2_{hij} = 1$ for all eligible students in the school. For all other cases, $W2_{hij} > 1$ as the selected student represents themselves as well as other students in the school.

9.2.5. Within-school non-response adjustment

The non-response adjustment of student base weights is necessary to ensure that these weights also represent those sampled students who did not participate, and it is applied to the weights of participating students at the school. In order to tailor adjustments to potential differences in the likelihood of participation across sub-groups, the adjustments were applied so that non-responding students were, as far as possible, represented by participating students with similar characteristics.

To this end, the following process was implemented:

- For each school, students were partitioned into four basic groups by combinations of gender (male / female) and grade level (low grade / high grade) categories. The *low grade / high grade* categories were determined for the Site as a whole and within cohorts based on the composition of the student sample by grade level, where the median grade was allocated to either the low grade or high grade group in order to derive two groups with approximately equal sizes.
- If any of these particular categories contained fewer than 10 responding students within the school, or the proportion of responding students was less than 50% of all eligible students, then the corresponding group was merged with the same grade category of the other gender at the school. If, after doing this, there were still categories with fewer than 10 students or 50% of responding students, all categories were merged into one adjustment category within the school.
- After completion of this process, the within-school non-response adjustment for every category ***k*** in school ***i*** from stratum ***h*** ($WSCNRA_{hik}$) was calculated as the ratio between the number of students who participated in the survey, plus those eligible who did not participate, divided by the number of students who participated in the survey:

$$WSCNRA_{hik} = \frac{STR_{hik} + STNR_{hik}}{STR_{hik}} \quad \text{Eq. 9:5}$$

9.2.6. Final weight

The overall student sampling weight is the product of the final weighting components for schools and students, and was estimated as follows for each student ***j*** in a subgroup ***k*** of school ***i*** in stratum ***h***:

$$TOTWGT_{hikj} = W1_{h1} \times T1_{h1} \times W1_{h1} \times SCNRA_h \times W2_{hij} \times WSCNRA_{hik}$$ Eq. 9:6

A final adjustment to the overall student weight is done to trim unusually high values for the final weight. For any final total weight that exceeds four times the median weight within the corresponding explicit stratum, its value was trimmed to be equal to that threshold value.

All student data reported in the SSES international reports were weighted by the overall student sampling weight, included as a variable named WT2019 in the SSES international databases.

9.2.7. Weights for other analysis units

In addition to collecting data from students, SSES also collected data from their school principals, teachers and parents. The selection of these participants was a function of the school and student selection for the survey.

The weights used for reporting the results for these additional analysis units were:

A. **School Principal questionnaire:** the final weight corresponds to the sum of final student weights in the school. An additional stratum non-response adjustment was applied, equal to the ratio of the sum of weights of all students in the corresponding explicit stratum ***h***, to the sum of weights of all students in schools with responding principals. If all schools had principals who responded, this factor is equal to 1.

B. **Teacher questionnaire**: each weight is equal to the sum of student final weights in the school, divided by the total number of participating teachers in the school. An additional stratum non-response adjustment was applied, equal to the ratio of the sum of weights of all students in the corresponding explicit stratum ***h***, to the sum of weights of all students in schools with responding teachers. If all schools had teachers who responded, this factor is equal to 1.

C. **Parent questionnaire**: each weight is equal to the sum of student final weights in the school, divided by the total number of participating parents in the school. An additional stratum non-response adjustment was applied, equal to the ratio of the sum of weights of all students in the corresponding explicit stratum ***h***, to the sum of weights of all students in schools with responding parents. If all schools had parents who responded, this factor is equal to 1.

9.2.8. Senate weights

Senate weights were used when analysing the pooled data so that each Site meeting sample participation requirements had the same contribution to the analyses. For this purpose, the final weights are adjusted so the sum of the weights for each Site has the same value for all participating countries. In this case, the value 1000 was used as the norm. Senate weights were calculated separately for student and parent final weights.

The senate weight for student ***i*** in Site ***S*** (***SENWGT_STA$_{Si}$***) was calculated as shown in Eq. 9:7 below:

$$SENWGT\_STA_{Si} = \frac{WT2019_{Si}}{\sum_{i \epsilon S} WT2019_{Si}} \times 1000$$ Eq. 9:7

Where $WT2019_{Si}$ is the student final weight for student **I** in Site ***S*** estimated as described in the previous section. The value in the denominator is the sum of all student final weights in the Site. This ratio is multiplied by 1000.

Similarly, the senate weight for parent ***j*** in Site ***S*** ($SENWGT\_STA_{Sj}$) from Site ***S*** was calculated as shown in in Eq. 9:8.

$$SENWGT\_PAA_{Sj} = \frac{WT2019\_PC_{Sj}}{\sum_{j \epsilon S} WT2019\_PC_{Sj}} \times 1000 \qquad \text{Eq. 9:8}$$

Where $WT2019\_PC_{Si}$ is the parent final weight for parent ***j*** in Site ***S*** estimated as described in the previous section. The value in the denominator is the sum of all parent final weights in the Site. This ratio is multiplied by 1000.

9.3. Replicate weights for variance inference

Cluster sampling techniques permit an efficient and economic data collection. However, because these samples are not simple random samples, it is not appropriate to apply standard formulae that assume a simple random sample for calculating standard errors to generate the sampling error for population estimates.

Replication techniques provide tools with which to derive appropriate estimates of sampling variance for population statistics. For SSES, the Jackknife repeated replication (JRR) technique was used to compute standard errors for population means, percentages, regression coefficients and any other population statistics (Foy and LaRoche, 2016[3]).

9.3.1. Calculation of Jackknife repeated replication weights

The basic feature of the JRR technique is a grouping of primary sampling units (PSUs) – schools in SSES – into zones based on similar sample design conditions (e.g. strata) and subsequent repeated draws of subsamples from these zones that match the original sample as closely as possible. This process is referred to as repeated replication. For SSES, the two main features of the sample design that JRR incorporated were repeated draws of subsamples reflecting the stratification of schools and the clustering of students within schools. This was done by defining Jackknife sampling zones according to the stratification scheme in each zone, and by pairing successive schools to model the clustering of each Site sample (see Chapter 4 for information on the sample design). The subsampling required for JRR was applied within each sampling zone. Sampling zones are always constructed within explicit strata. When an explicit stratum has an odd number of schools, either by design or because of school non-response, the students in the remaining school are randomly divided to make up two 'quasi' schools for the purposes of calculating Jackknife standard errors. Each sampling zone then consists of a pair of schools or 'quasi' schools.

Since most Site samples consisted of 75 schools and one explicit stratum, a total of 38 sampling zones were created, each containing a pair of schools, and the last one divided into two 'quasi' schools. In cases where more than 75 schools had been selected, additional zones were collapsed into the first 75 zones.

Within each of the sampling zones, the first school was assigned a value of two and the second school assigned a value of zero to form the first replicate subsample. To form a second replicate subsample from the same zone, these factors were reversed and a value of zero was assigned to the first school and a value of two to the second school. This means that one of the paired schools had a contribution of zero to the estimated outcome, the

second a double contribution and the contribution of all other schools remained the same (factor of one). After repeating this procedure for all 38 sample zones, 76 different subsamples were defined. The replicate weights procedure was achieved by multiplying student weights with the Jackknife indicators once only for each sampling zone.

This process resulted in a weight being added to the data file for each Jackknife replication. Thus, within one sampling zone at a time, each element of one PSU received a double weight and each element of the other PSU received a zero weight. This procedure can be illustrated by a simple example featuring 28 students from seven different schools (A−G) paired into four sampling zones (Table 9.1).

Table 9.1. Example of the computation of replication weights

ID	Student Weight	School	Sampling Zone	Replicate Weights							
				Zone 1.1	Zone 1.2	Zone 2.1	Zone 2.2	Zone 3.1	Zone 3.2	Zone 4.1	Zone 4.2
1	5.2	A	1	10.4	0.0	5.2	5.2	5.2	5.2	5.2	5.2
2	5.2	A	1	10.4	0.0	5.2	5.2	5.2	5.2	5.2	5.2
3	5.2	A	1	10.4	0.0	5.2	5.2	5.2	5.2	5.2	5.2
4	5.2	A	1	10.4	0.0	5.2	5.2	5.2	5.2	5.2	5.2
5	9.8	B	1	0.0	19.6	9.8	9.8	9.8	9.8	9.8	9.8
6	9.8	B	1	0.0	19.6	9.8	9.8	9.8	9.8	9.8	9.8
7	9.8	B	1	0.0	19.6	9.8	9.8	9.8	9.8	9.8	9.8
8	9.8	B	1	0.0	19.6	9.8	9.8	9.8	9.8	9.8	9.8
9	6.6	C	2	6.6	6.6	13.2	0.0	6.6	6.6	6.6	6.6
10	6.6	C	2	6.6	6.6	13.2	0.0	6.6	6.6	6.6	6.6
11	6.6	C	2	6.6	6.6	13.2	0.0	6.6	6.6	6.6	6.6
12	6.6	C	2	6.6	6.6	13.2	0.0	6.6	6.6	6.6	6.6
13	7.2	D	2	7.2	7.2	0.0	14.4	7.2	7.2	7.2	7.2
14	7.2	D	2	7.2	7.2	0.0	14.4	7.2	7.2	7.2	7.2
15	7.2	D	2	7.2	7.2	0.0	14.4	7.2	7.2	7.2	7.2
16	7.2	D	2	7.2	7.2	0.0	14.4	7.2	7.2	7.2	7.2
17	4.9	E	3	4.9	4.9	4.9	4.9	9.8	0.0	4.9	4.9
18	4.9	E	3	4.9	4.9	4.9	4.9	9.8	0.0	4.9	4.9
19	4.9	E	3	4.9	4.9	4.9	4.9	9.8	0.0	4.9	4.9
20	4.9	E	3	4.9	4.9	4.9	4.9	9.8	0.0	4.9	4.9
21	8.2	F	3	8.2	8.2	8.2	8.2	0.0	16.4	0.0	0.0
22	8.2	F	3	8.2	8.2	8.2	8.2	0.0	16.4	0.0	0.0
23	8.2	F	3	8.2	8.2	8.2	8.2	0.0	16.4	0.0	0.0
24	8.2	F	3	8.2	8.2	8.2	8.2	0.0	16.4	0.0	0.0
25	9.5	G	4	9.5	9.5	9.5	9.5	9.5	9.5	19.0	0.0
26	9.5	G	4	9.5	9.5	9.5	9.5	9.5	9.5	19.0	0.0
27	9.5	G	4	9.5	9.5	9.5	9.5	9.5	9.5	0.0	19.0
28	9.5	G	4	9.5	9.5	9.5	9.5	9.5	9.5	0.0	19.0

As shown in Table 9.1, since there were seven schools in total, the last school had to be split to have an additional pair, resulting in a total of four paired schools. For the first pair of schools (A – B), the first replicate consists of doubling the student weight for all students in School A while setting the student weight to zero for all students in School B. In the second step, the factor was reversed resulting in students in School A having a weight of zero, while those in B having their weights doubled. In both cases, the student weight for the remaining schools stayed the same. The process continued for the remaining pairs until a total of eight replicate weights was achieved.

For each Site and cohort, 76 replicate weights were computed (38 sampling zones of two schools or pseudo-schools each) regardless of the number of sampling zones. In cities with fewer sampling zones, the remaining replicate weights were equal to the original sampling weight and therefore did not contribute to the sampling variance.

Estimating the sampling variance for a statistic, $\boldsymbol{\mu}$, involves computing it once with the sampling weights for the original sample and then again with each of the 76 replication weights separately. The sampling variance $\widehat{\boldsymbol{SV}}_{\boldsymbol{\mu}}$ estimate is computed using the formula:

$$\widehat{SV}_{\mu} = \frac{1}{2}\sum_{i=1}^{76}[m_i - m_s]^2 \qquad \text{Eq. 9:9}$$

Where $\boldsymbol{m_s}$ is the statistic $\boldsymbol{\mu}$ estimated for the population through use of the original sampling weights and m_i is the same statistic estimated by using the weights for the $\boldsymbol{i^{th}}$ of 76 Jackknife replicates. The standard error $\widehat{\boldsymbol{SE}}_{\boldsymbol{\mu}}$ for statistic $\boldsymbol{\mu}$, which reflects the uncertainty of the estimate due to sampling, is computed as:

$$\widehat{SE}_{\mu} = \sqrt{\widehat{SV}_{\mu}} \qquad \text{Eq. 9:10}$$

The JRR method appropriately estimates the variation arising from the selection of students under a multi-stage stratified cluster sample design. The sampling variance and standard error can be estimated for any statistic derived from the sample, including means, percentages, standard deviations, correlations, regression coefficients and mean differences.

References

Kish, L. (2004), *Statistical Design for Research*, John Wiley & Sons. [1]

Martin, M., V. Mullis and M. Hooper (eds.) (2016), *Estimating Standard Errors in the TIMSS 2015 Results*, Boston College, TIMSS & PIRLS International Study Center, http://timss.bc.edu/publications/timss/2015-methods/chapter-4.html. [3]

OECD (2017), *PISA 2015 Technical Report*, OECD Publishing, https://www.oecd.org/pisa/data/2015-technical-report/. [2]

Notes

[1] See Chapter 4 (Sample Design) for an explanation of how MOS is calculated.

[2] The value of TCS = 50 was the default value set according to the Technical Standards for SSES. However, this value was changed in some particular circumstances, at the request of the cities, and an insufficient number of schools to draw a proper number of students. These cases were the census cities of Manizales and Sintra; as well as the older cohorts of Houston and Helsinki (TCS = 75) and Ottawa (TCS = 60).

Chapter 10. Data management processes

This chapter describes the procedures implemented for the online data collection, creation of the database and cleaning and validation of the data collected during the Main Study implementation for the OECD Survey on Social and Emotional Skills 2019 (SSES) by the International Contractor and the participating cities collectively.

10.1. Introduction

The SSES Technical Standards stipulate that the data must be consistent, precise, able to be generalised and provided in a timely manner. During the data collection preparation and data management process, extensive efforts were made to ensure the data collected was of the highest quality and as free from error as possible. When setting up the online data collection system, multiple levels of quality assurance procedures were put in place to check each data collection point (i.e. every possible option respondents may select). Once participant response data were exported from the online platform and cities had submitted their coded occupational data files and data validation reports, data cleaning was carried out Site by Site to check the data for inconsistencies and to create a standardised dataset for each Site.

10.2. SSES online data collection

Online data collection has many advantages compared to traditional paper-based data collection. Assessment items can be interactive and individualised using advanced online testing features. Respondents' responses are collected, and coded in some cases, simultaneously as the assessment sessions progress. Completion rates can be reported immediately on demand, and response data is available for download as assessment sessions complete. In addition, printing and logistics-related management and cost are no longer involved, and data entry related errors, such as illegible handwriting, invalid responses and clerical mistakes are no longer a concern. These key features enable online data collection to be more accurate and efficient.

For SSES, all instruments were prepared and delivered using an online platform for students, parents, teachers and principals. In addition, a paper option was also provided for the parent questionnaire.

Various types of data collection techniques were used in the online assessment forms including check boxes, radio buttons, drop-down menus and text input fields. When developing the instruments, measures were incorporated at each step of the creation and production workflows to ensure the accuracy of the final data collection.

10.2.1. At the source item creation stage:

Each data collection point had certain output specifications coded within the source item. For example, each text data collection point was configured separately to allow for only a defined range of numbers and type of characters to appear that were relevant to the question. Fields that required numerical responses were coded to only accept numbers in a valid range of characters that would be relevant to responding to the question. The cognitive items were coded to automatically score each response.

10.2.2. At the translation stage:

For SSES, most data collection points (i.e. each available response option) that appeared in the source version of the assessment item were linked to a segment of text. The RM Assessment Master (AM) translation system was designed to ensure that the translations of each of these segments of text also linked to the same data point. This configuration ensured data collected from various language versions for the same data point were linked to and reported against the same data point.

During the translation verification and layout verification process, in addition to verifying the translation, each segment of text was checked and verified to ensure that it was linked to the correct data point.

10.2.3. Data report design and verification of the raw data outputs

Once the test forms were created, the format of the raw data output file was designed, coded and tested for each form. Each data report included the test language (and country), the login time, logout time, and a field for every data point in the forms together with additional data points for Site-specific variables and additional items that were added by cities. All data points in the data output report were checked for each testing language and each online assessment form, using dummy data runs and spot checking. The AM system captured a screenshot of each page from each respondent's form. This image was used for reference while verifying the spot checking of the data outputs.

Some of the more specific aspects verified were:

- Language reported in output files matches the language selected in the drop-down option in the online platform.
- Reporting of items that had been configured to appear as part of rotated item sets.
- Time of login and logout.
- Site-specific adaptations, national questions and exclusions were all mapped out and cross-checked against the data output reports.
- The mapping of data response points that were unique to a Site despite being a shared data point. For some contextual questionnaire items, the same data collection point was linked to different content for different languages/cities. For example, one question asked about the language spoken at home and offered six options. The choices of languages and the order of the chosen languages were different for each Site, therefore the same data collection point (each option) was linked to different languages for each Site. The mapping of the data points assisted with the data cleaning process.
- Text inputs and outputs were consistent with the specifically coded requirements.
- Cognitive items were reporting scores as well as responses.

During these verification processes, tens of thousands of data points were checked to verify that their raw data output was consistent with the input data and had been reported accurately. All issues found were corrected and recoded where necessary in the data output (.csv) files. The output files were then checked again to ensure there were no further inconsistencies. Spot checking was also carried out on the final data reports for the cities to further check the accuracy of the outputs.

10.2.4. During live data collection

It was anticipated that some online assessment sessions could be affected or interrupted by external factors such as unreliable internet connections, power failure and computer errors which may result in data loss. To minimise the impact of these factors, all data collected from each assessment was stored in the online delivery cloud system as soon as it had been entered. Respondents were able to re-login to the assessment and continue from where the interruption had occurred without loss of data from the previous session.

A navigation tool was developed and deployed in the forms to assist respondents to easily identify which responses were incomplete and allow for navigation back to items that required further response. Combined, these features minimised the loss of data due to external factors and encouraged completion of the assessments.

10.3. Occupational data

Five items required occupational coding by the cities; three items from the student questionnaire and two items from the parent questionnaire. To assist with this process, a specific occupational data export file was designed which included the respondent ID, their response to items on the form, and headers on the form that were relevant to the occupation coding process. The occupation data reports were also checked and verified prior to live testing to ensure that they were reporting correctly for each Site and each assessment form and that all text was appearing correctly in the download file for each language.

Once the coding of these items was completed, coded data files were submitted to the International Contractor through the project portal by each Site together with their data validation report.

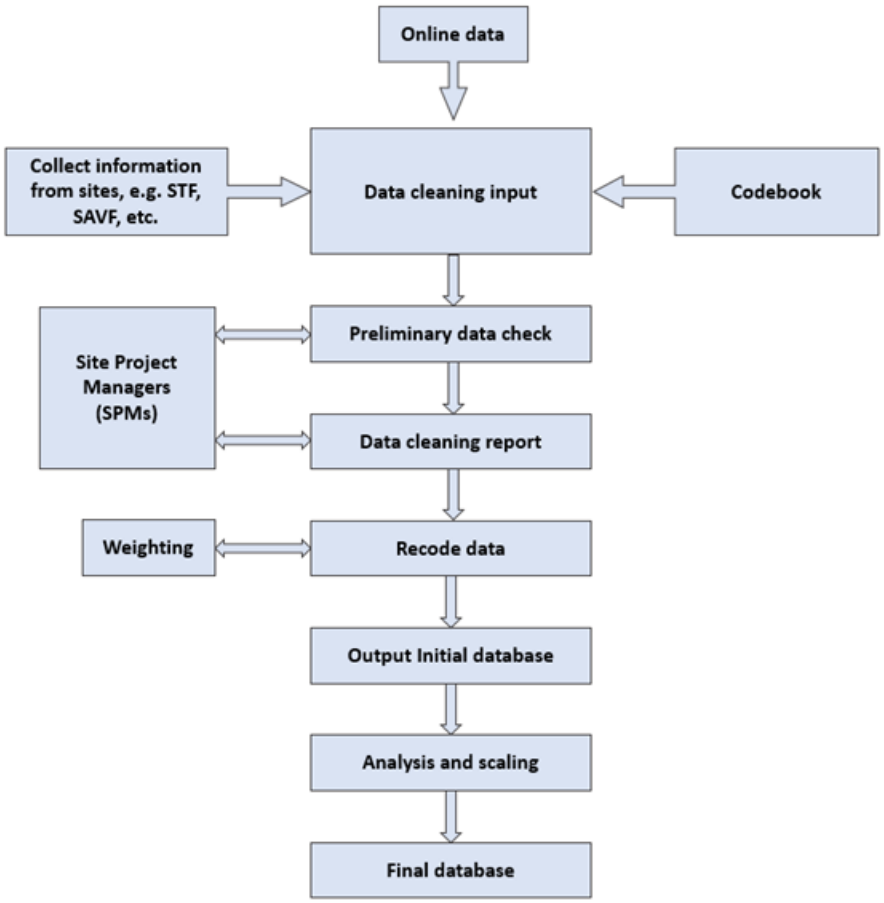
10.4. Confirming the integrity of the international and national databases

Data processing is an essential step in the implementation of any large-scale survey. For SSES, the objective of this task was to ensure the data were accurate and consistently reflected the information collected from participating cities. Any errors and inconsistencies across different data sets needed to be detected and resolved in collaboration with Site Project Managers (SPMs). Data processing is vital to enable subsequent analysis, reporting and an appropriate interpretation of Study results. The procedures used for data processing in SSES are outlined in the following sections. Figure 10.1 provides an overview of the data cleaning process.

After the completion of the scheduled data collection period at each Site, a complete set of raw data files by instrument in text file format (.csv) was downloaded via the RM Assessment Master (AM) portal by the International Contractor's data processing team. Upon receiving these data files, the team conducted additional checks and data cleaning. Where required for the implementation of data cleaning, data queries were sent to Site Project Centres (SPCs) for clarification. Where necessary, the International Contractor recoded data to ensure that submitted data files met the requirements of the internationally defined data structure. Any additional variables that were not international variables were supplied to SPCs as in the received form, without undergoing data cleaning procedures.

A complex survey such as SSES, which involved multiple surveys from multiple populations and across multiple cities, requires a well-designed data cleaning tool and procedure. For SSES, the International Contractor developed a set of standardised data cleaning procedures using the *Statistical Package for the Social Sciences* (SPSS®) as a processing tool. This ensured that the same procedures were applied to process the data set from each Site and thereby maintain the highest possible levels of data quality and integrity.

Figure 10.1. Flowchart of data cleaning process overview



In order to ensure data processing steps were carried out correctly and consistently, the following procedures were implemented:

- A master data cleaning tool and procedures were developed and tested before the data processing started. This ensured that the data cleaning procedures would be carried out consistently across all cities.
- Data cleaning reports were developed to assist communication between the International Contractor's data processing team and SPCs. The reports were based on a template that listed all observations and possible corrections resulting from the data cleaning procedures.
- Initial training and ongoing supervision were provided to all International Contractor staff who would be involved in the data processing. This ensured that the data cleaning procedures would be carried out consistently by all team members.
- The International Contractor developed codebooks for each survey instrument to accompany the final data sets.
- Additional national variables were copied at the end of final datasets without undergoing data cleaning procedures.

- Any potential errors and inconsistencies detected during the process of data cleaning were recorded in the data cleaning report and sent to SPMs for review and rectification.
- Upon receiving feedback from SPCs, any required corrections were implemented in the database. If SPCs did not provide any feedback within the required timeframe, the International Contractor's data cleaning team processed data according to pre-agreed rules for resolving inconsistencies, and decisions about final data treatment were sent to the SPC. Data cleaning procedures were applied repeatedly until no further issues were identified.
- The International Contractor implemented a centralised monitoring system to assist with the data processing procedures for SSES.

Please refer to the OECD Technical Standards for details of the SSES standards of data collection (see Annex A).

10.4.1. Data cleaning quality control

Multiple data files and other necessary information from each Site were forwarded to a designated member of the International Contractor's data cleaning team responsible for the corresponding Site data. The files included the raw Site data files for each instrument (from students, parents, teachers and schools), Student Tracking Forms (STF), Site Adaptation Verification Forms (SAVF) and any other corresponding variable adaptation tables. Email notifications from cities (about any records that were flagged for corrections, updates or follow-up activities) were also provided to the responsible member of the International Contractor's data cleaning team. If required, the team member first checked for possible discrepancies and where necessary, additional data clarifications were sought from SPMs. The Site data would then go through the data cleaning steps as described below.

Setting parameters

As a first step, the International Contractor's data cleaning team member set the Site-specific data parameter variables such as a 3-letter code indicating the Site, a 3-letter code for language, a cohort indicator variable, international variables and national adaptation variables where applicable. These data parameter variables were later incorporated into the final dataset for analysis and identification purposes.

Reading in raw data files

Table 10.1 includes a list of raw data files extracted from the RM portal and read into SPSS® for data cleaning. Unique record identifiers were created in each SPSS data file and preliminary checks as described below were then carried out unless specified otherwise.

Table 10.1. List of Site raw data files

Raw data files extracted from RM portal	SPSS data files
Student Questionnaire – Older students Student Questionnaire – Younger students	ST
Parent Questionnaire	PA
Teacher Assessment Questionnaire – Older students Teacher Assessment Questionnaire – Younger students	TC
Teacher Contextual Questionnaire	TcQ
Principal Questionnaire	PrQ
Student Tracking Forms	STF

Verification of records in the SSES databases

SSES surveys were conducted online and data were captured by the RM system. Duplicate records could be caused when different user IDs were assigned to the same respondent during the assessment administration period. If the International Contractor found duplicate records, these would be verified by the SPM and recommended for updating (including deletion of the duplicate). Similar internal checks were also conducted for other data files. All identified duplicate records had to be resolved before any further data cleaning steps could be conducted. Sometimes SPCs also requested the removal of a student record due to withdrawal from participation by a student or parent. Information about the final records in each Site database were reported by SPCs.

Site-specific adaptations

Some variables were Site-specific in various data files and required Site-specific coding and cleaning. Site-specific variables that were verified included:

- Country of birth (COB)
- Language
- Site adaptation variables as documented in the SAVF.

Where necessary, recodings were performed for any of the following reasons:

- Categories in a Site-adaptation variable required translation into international categories to meet the requirements of the international variable structure
- Data errors were identified by cities and reported to the International Contractor
- Items were identified as having issues in the online version or printing errors in the paper version, and these issues had been reported by SPCs.

10.4.2. Handling of missing data

There were three types of missing data in SSES: (a) omitted missing data, (b) missing-by-design data and (c) Site-specific missing data and missing due to a technical issue during an administration session.

- Omitted missing data is a response to an item that was administered to a respondent who did not provide a response. This type of missing data was coded as ‘9’.
- If an item was designed not to be administered to a participant, who thereby had no opportunity to respond, the item response would be considered as value missing by design and coded as ‘7’ (not applicable).
- Where a Site requested an item be omitted from an instrument, the corresponding values were coded as ‘8’. The same code was also applied in cases where an

unexpected system error occurred while a respondent completed the survey, for example, when an item failed to display properly.

10.4.3. Preparing national and international data files for analysis

To enable the data files from all cities to be analysed collectively, they had to have the same final file structure. Therefore, the original variable names in the raw Site online data files were replaced with corresponding international variable names. The variables were then rearranged within the data file. This procedure involved formatting, assigning auxiliary codes, establishing the correct order and checking and validating all codes. The International Contractor also computed the additional variables listed in Table 10.2.

Table 10.2. Variables created for SSES instruments

Variable type	Variable created in the SPSS data files *				
	ST	**PA**	**TC**	**TcQ**	**PrQ**
Count of valid responses	Resp_STA	Resp_PAA	Resp_TCA	Resp_TCV	
Indicator of at least 50% valid responses	Pct50_STA	Pct50_PAA	Pct50_TCA	Itm3_TCV	Pct50_PRQ
Indicator of missing responses to certain sections or the entire instrument	ST_S1_Empty STQ_Empty	PA_S1_Empty PAQ_Empty	TC_S2_Empty	TCV_Empty TCQ_Empty	PRQ_Empty
Percentage of completion with valid responses	Pct_STQ_complete	Pct_PAQ_complete	Pct_TCA_overall	Pct_TCQ_complete	Pct_PRQ_complete

* Please refer to codebooks for further variable details.

Where applicable, the following additional descriptive information was added to each Site data file:

- The 3-letter code indicating the language of the survey instrument (example: LANG, STQM01601_Alpha).
- The 3-letter and 3-digit codes indicating country of birth (example: STQM01501_Alpha).
- The 3-letter and 3-digit codes indicating the language spoken at home.

Background variables

The International Contractor used data from Student Tracking Forms (STF) to validate the status of student participation and capture student background information, such as the students' grade (TF_Grade) and gender (TF_Gender). A student gender variable (Gender_Std) was computed by using valid codes (i.e. not missing) from the variable TF_Gender. When TF_Gender had a missing code, a valid code from the variable STQM00401 reflecting student gender information from the student questionnaire was inserted. The students' date of birth (DOB_std) and age (Age_Std) were also computed using the same procedure. Student age was derived from information about the student's date of birth and the actual start date of the administration of the student questionnaire. Generally, data from the STF were given priority over information provided by students' when responding to the questionnaire.

Student participation status

cities provided information about students who were excluded from SSES due to special education needs (SEN), having left school or being ineligible. Criteria based on age and on the international grade definition for each cohort were applied to determine whether the remaining students were eligible to be part of the SSES target population (see Technical Standards in Annex A). Subsequently, a final student participation indicator was computed based on student participation status, SSES target population eligibility and the percentage of their completion of the Student Direct Assessment (Item labels start with STA).

School participation

The final student participation indicator was aggregated at school level to create a school participation indicator. Schools were flagged for participating on the basic requirement of having a minimum of 25% eligible student participation. Table 10.3 below contains a list of indicators that were computed to the different data files.

Table 10.3. Indicators to identify student eligibility and school participation

Variables	Variable available in the SPSS data files *				
	ST	PA	TC	TcQ	PrQ
Student participation code (PartCode)	✓				
Student age (Age_Std)	✓				
Student grade (TF_Grade)	✓				
STF eligible student (TF_Eligible)	✓				
SSES Student participation (InSSES_Std)	✓	✓	✓		
Percentage of eligible students with at least 50% of STA valid responses at the school level (pct_SchPart)	✓				✓
School participation (SchPart)	✓	✓	✓	✓	✓

* Please refer to codebook for variable details.

The International Contractor generated various cleaning reports reflecting the results of the checking procedures during different stages of Site data management. If Site data required additional cleaning, the corresponding reports with recommended corrections were collated and sent to the SPM for further checks of data inconsistencies found in the data files.

Upon receiving all cleaning reports with their corresponding data clarifications from the SPMs, the agreed corrections recommended in these reports were implemented. For unresolved data inconsistencies, the following general rules were applied:

- Unresolved student, parent, teacher or school identification issues led to record deletions in the data file.
- Unresolved data or coding issues in variables led to applying standard data treatments as stated in the cleaning reports provided to the SPM for data clarification.
- Unresolved systematic errors in the data file such as coding errors led to replacements with auxiliary codes such as ‘not applicable’ (N/A).

10.4.4. Final data yield

Generally, all records from received data sets were retained in the yield data set with the exception of those that were identified in collaboration with SPMs as duplicate records, parent refusals or student withdrawals. The variables Sch_Part and InSSES_Std in the

student data set were created to identify valid records (indicated by both variables having a code of '1'). Only valid records were included in the data files for scaling and analysis of SSES data. For parent and teacher data, the indicators SchPart and Pct50_PAA for parent (or Pct50_TCA for teacher) were used to determine which records could be regarded as valid. Table 10.4 shows the numbers of data records from received data files and the final valid records for student, parent and teacher surveys by cohorts following the data cleaning procedures.

Table 10.4. Counts of the data records as received and the valid records of ST, PA and TC by cohort

Site	Younger cohort						Older cohort					
	Data records as received			Valid data records			Data records as received			Valid data records		
	ST	PA	TC	ST	PA	TC	ST	PA	TC	ST	PA	TC
Bogota	3430	3376	3466	3415	3222	3438	3380	3252	3465	3356	3036	3440
Daegu	3021	1400	2587	3008	1296	2548	3326	901	2868	3326	803	2837
Helsinki	3071	1042	2860	3034	932	2830	2466	691	2802	2448	607	2749
Houston	3355	672	2956	3333	577	2931	3117	450	2715	3101	385	2681
Manizales	3243	2705	3093	3226	2657	3048	3531	2811	3298	3531	2748	3265
Moscow	3384	3227	3407	3363	3108	3354	3441	3041	3526	3429	2922	3477
Ottawa	3325	820	1884	3250	582	1860	2349	485	1037	2190	372	985
Sintra	2242	831	2080	2224	725	2042	1643	440	1532	1636	390	1499
Suzhou	3647	3606	3662	3633	3589	3648	3621	3530	3570	3613	3510	3547
Turkey	2796	2095	2903	2701	1974	2841	3184	2155	3416	3168	2033	3373
International	31514	19774	28898	31187	18662	28540	30058	17756	28229	29798	16806	27853

10.4.5. Data Delivery

The International Contractor built datasets iteratively at multiple stages and delivered them at two points in time – once after completing the data cleaning procedures (as an initial dataset) and then again after completing the scaling procedures (as a final dataset). After completing all data cleaning procedures and resolving identified data issues, the International Contractor built student datasets for each Site. These datasets were submitted to OECD for approval and were also used for computing student weights.

Once student weights had been computed for all cities, Site datasets were combined at the international level for data from students, teachers, parents and school principals. The International Contractor then assigned variable labels and value labels to data variables in all data files. Two separate datasets were built for the teacher survey, one of which (TC) contained teacher responses to the indirect assessment of students and consisted of one record per student. For most teachers there was more than one record. Another dataset (TCQ) contained responses to the teacher questionnaires, which had one record per teacher. The International Contractor also compiled detailed codebooks for all datasets.

Initial data files

The first versions of cleaned and weighted international datasets were submitted to the OECD. The student dataset contained the students' final weights and replicate weight variables. These datasets were then used for further scaling and initial analyses. Subsequently, national datasets were constructed for each Site. Complete sets of five data files with their corresponding codebooks were uploaded to the designated project portal for cities to download.

Final data files including scale scores

Once the scaling procedures (see Chapter 12) had been finalised, two sets of scale scores for each of the 17 scales were generated for students and parents. These scores were merged to the respective data files. Further, the International Contractor computed raw scores per scale (as the average of item response scores) for each student for whom a teacher had provided an indirect assessment and merged these to the teacher response (TC) data file. These updated datasets were then submitted to the OECD as final data files while final national data files were made available at the designated project portal for Site teams for download.

Chapter 11. Sampling outcomes

Rigorous sampling of schools and students was a key component of the OECD Survey on Social and Emotional Skills 2019 (SSES). Implementing the sampling plan was the responsibility of the Site Project Manager (SPM) in each participating city. SPMs were supported in this endeavour by the sampling consultants from the International Contractor.

Sampling consultants conducted the school sampling for all cities and trained Site Administrators using the within-school sampling software provided by the International Contractor to implement within-school sampling. As an essential part of their sampling activities, SPMs were responsible for providing detailed documentation describing their national sampling plans (sampling data, school sampling frames and school sample selections). The documentation for each SSES participant was reviewed by the sampling consultants, including the target population definitions, detailed information on estimated coverage and exclusion levels, and stratification variables.

This chapter summarises the major characteristics of the national samples for SSES.

11.1. Target population

As described in Chapter 4 Sample Design, the international target populations for the SSES older cohort and younger cohort assessments were defined as those students aged between 15 years 3 months and 16 years 2 months with at least 7 years of formal schooling; and those aged between 10 years 3 months to 11 years 2 months with at least 2 years of formal school education. Minor adjustments could be made to these age groups, to better align the timing of the testing period to the school year for the participating Site. In any case, the age definitions comprised no more than a period of 12 months with regard to the students' date of birth.

Table 11.1 and Table 11.2 present the ranges of birthdates of the sampled students eligible to participate in the SSES survey for each Site, as well as the range of student grades of eligible students observed for the Site. In all cities, the birthdates were within the parameters set out in the SSES Sampling Framework distributed to all cities as part of the sample preparations.

Table 11.1. National age and grade definition – SSES – younger cohort

Participating Site	Student date of birth		Average age at administration time	Student Grade	
	From	To		Minimum	Maximum
Bogota	01-JUL-2008	30-JUN-2009	10.7	2	6
Daegu	01-JUN-2008	31-MAY-2009	10.9	4	5
Helsinki	01-SEP-2008	31-AUG-2009	10.7	2	6
Houston	01-AUG-2008	31-JUL-2009	10.7	3	6
Manizales	01-JUL-2008	30-JUN-2009	10.7	2	6
Moscow	01-JUL-2008	30-JUN-2009	10.8	2	6
Ottawa	01-JUL-2008	30-JUN-2009	10.8	4	6
Sintra	01-JUL-2008	30-JUN-2009	10.9	2	6
Suzhou	01-SEP-2008	31-AUG-2009	10.7	4	6
Istanbul	01-JUL-2008	30-JUN-2009	10.8	2	6

Table 11.2. National age and grade definition – SSES – older cohort

Participating Site	Student date of birth		Average age at administration time	Student Grade	
	From	To		Minimum	Maximum
Bogota	01-JUL-2003	30-JUN-2004	15.7	7	11
Daegu	01-JUN-2003	31-MAY-2004	16.0	8	11
Helsinki	01-SEP-2003	31-AUG-2004	15.7	7	11
Houston	01-AUG-2003	31-JUL-2004	15.7	7	12
Manizales	01-JUL-2003	30-JUN-2004	15.7	7	11
Moscow	01-JUL-2003	30-JUN-2004	15.7	7	11
Ottawa	01-JUL-2003	30-JUN-2004	15.8	9	12
Sintra	01-JUL-2003	30-JUN-2004	15.9	7	12
Suzhou	01-SEP-2003	31-AUG-2004	15.6	9	12
Istanbul	01-JUL-2003	30-JUN-2004	15.8	9	12

11.2. National coverage and exclusions

Table 11.3 and Table 11.4 summarise population coverage and exclusion rates for the SESS target populations. Site coverage of the international target population was generally comprehensive, according to the Site definitions negotiated with each of the participants. A comprehensive description of the cities can be viewed in Table F.1 in Annex F. The Site definitions correspond to the political and/or geographical area usually associated with the Site name, with the exception of: Helsinki, which includes only public schools; Houston, which corresponds to the public schools under the authority of the Houston Independent School District; and Ottawa, which corresponds to public schools within this city.

Regarding exclusions, most cities adhered to the conditions set out in the Technical Standards, to allow no more than 2.5% of school level and 2.5% of within-school level exclusions, for an overall exclusion rate of 5% or less of the target population.

Only Helsinki in the younger cohort and Helsinki and Houston in the older cohort reached exclusion levels higher than those stipulated by the Technical Standards. In the case of Helsinki, the main reason for the breach in the upper limit of the exclusion rates corresponds to the within-school exclusions. These exclusion levels were higher than expected and were due to a large number of students with foreign language backgrounds being unable to access the survey instruments. In the case of the older cohort of Houston, the main reason was the exclusion of schools with fewer than 20 eligible students, due to concerns from the SPM regarding possible identification of individual students if these schools were to participate.

Table 11.3. Population coverage and exclusions – SSES – younger cohort

Participating Site	Population from School Frame	School-Level Exclusions	Target minus school-level exclusions	School Level Exclusion Rate	Weighted Number of Excluded Students from student sample	Within Sample Exclusions	Overall Exclusion Rate
Bogota	98,262	118	98,144	0.12%	146	0.15%	0.27%
Daegu	20,497	76	20,421	0.37%	12	0.06%	0.43%
Helsinki	5,355	7	5,348	0.13%	422	7.89%	8.02%
Houston	16,063	293	15,770	1.82%	78	0.48%	2.31%
Manizales	4,052	154	3,898	3.80%	5	0.12%	3.92%
Moscow	105,009	643	104,366	0.61%	308	0.29%	0.91%
Ottawa	11,050	10	11,040	0.09%	98	0.89%	0.98%
Sintra	3,473	0	3,473	0.00%	45	1.30%	1.30%
Suzhou	103,212	163	103,049	0.16%	275	0.27%	0.42%
Istanbul	220,861	361	220,500	0.16%	202	0.09%	0.26%

Table 11.4. Population coverage and exclusions – SSES – older cohort

Participating Site	Population from School Frame	School-Level Exclusions	Target minus school-level exclusions	School Level Exclusion Rate	Weighted Number of Excluded Students from student sample	Within Sample Exclusions	Overall Exclusion Rate
Bogota	89,071	29	89,042	0.03%	193	0.22%	0.25%
Daegu	22,870	132	22,738	0.58%	12	0.05%	0.63%
Helsinki	3,230	7	3,223	0.22%	285	8.82%	9.04%

Houston	12,781	654	12,127	5.12%	139	1.09%	6.20%
Manizales	4,446	156	4,290	3.51%	12	0.27%	3.78%
Moscow	81,462	640	80,822	0.79%	153	0.19%	0.97%
Ottawa	10,848	0	10,848	0.00%	220	2.03%	2.03%
Sintra	3,277	0	3,277	0.00%	55	1.68%	1.68%
Suzhou	52,305	0	52,305	0.00%	370	0.71%	0.71%
Istanbul	173,049	0	173,049	0.00%	429	0.25%	0.25%

Table 11.5 shows the detail of the within-school exclusions, according to the categories, to explain such exclusion of students already sampled for the survey. The main categories were Special Education Needs (SEN) students who were identified by the school as unable to access the survey; students who had left school in the period between when the lists were prepared and the day of the assessment; and ineligible students.

Table 11.5. Detail of within-school exclusions – SSES

Participating Site	Younger Cohort					Older Cohort				
	SEN Students	Student Left School	Ineligible Student	Total	Exclusion Rate	SEN Students	Student Left School	Ineligible Student	Total	Exclusion Rate
Bogota	15	130	1	146	0.15%	5	186	3	193	0.22%
Daegu	3		9	12	0.06%	1	10		12	0.05%
Helsinki	403	20		422	7.89%	247	38		285	8.82%
Houston	6	71		78	0.48%	16	123		139	1.09%
Manizales		5		5	0.12%		12		12	0.27%
Moscow	231	63	14	308	0.29%	79	71	3	153	0.19%
Ottawa	40	17	41	98	0.89%	103	97	20	220	2.03%
Sintra	12	33		45	1.30%	8	47	0	55	1.68%
Suzhou	94	154	26	275	0.27%	57	294	19	370	0.71%
Istanbul	31	117	54	202	0.09%	11	302	116	429	0.25%

11.3. Target population and sample size

Table 11.6 and Table 11.7 show the number of schools and students in each participant's sampling frame, the number of schools and students sampled, and an estimate of the student population size based on the sample data. The estimates of the population were computed using sampling weights, which are explained in more detail in Chapter 9. The student population estimate based on the sampling frame does not take into account the portion of the population excluded within sampled schools, nor changes in the number of enrolled students in the school from the moment the sampling frame was constructed to the time the students were selected to participate. In most cases, the population size estimated from the sample closely matched the population size from the sampling frame. Differences were mainly the result of discrepancies between the number of students reported to be enrolled in the school in the sampling frame, and the number actually reported by schools after they were selected to participate in the survey.

Table 11.6. Population and sample sizes – SSES younger cohort

Participating Site	Sampling Frame		Sample		
	Schools	Students	Schools	Students	Student Population Site Estimated from Sample
Bogota	1,685	98,144	87	3,415	78,943
Daegu	229	20,421	77	3,008	19,240
Helsinki	94	5,348	83	3,034	5,031
Houston	207	15,770	74	3,333	15,292
Manizales	85	3,898	83	3,226	3,623
Moscow	644	104,366	77	3,363	100,983
Ottawa	229	11,040	89	3,250	10,735
Sintra	52	3,473	48	2,224	3,405
Suzhou	387	103,049	76	3,633	100,857
Istanbul	3,074	220,500	91	2,701	210,317

Table 11.7. Population and sample sizes - SSES older cohort

Participating Site	Sampling Frame		Sample		
	Schools	Students	Schools	Students	Student Population Site Estimated from Sample
Bogota	1,150	89,042	82	3,356	67,340
Daegu	217	22,738	78	3,326	20,521
Helsinki	57	3,223	55	2,448	3,207
Houston	45	12,127	45	3,101	11,497
Manizales	72	4,290	70	3,531	4,001
Moscow	653	80,822	77	3,429	78,510
Ottawa	62	10,848	58	2,190	10,672
Sintra	33	3,277	29	1,636	3,161
Suzhou	88	52,305	75	3,613	50,107
Istanbul	1,505	173,049	80	3,168	161,275

11.4. Participation statistics

During data collection, some sampled schools and students did not participate in the survey. In some instances, selected schools were approached but turned out to not have eligible students in the corresponding cohort; or were closed. In those cases, schools were excluded and not substituted. In other cases, the school had eligible students but refused or was unable to participate. In those cases, Site Administrators approached one of the substitute schools assigned to that school. If no substitute school could be approached or was willing to participate, these were counted as a school-level non-respondent.

As is the practice with the *Programme for International Student Assessment* (PISA) (OECD, 2017[1]), participation rates within schools were measured against certain benchmarks, with consequences when these rates fell to low levels, because of the increased risk of non-response bias in these situations. If fewer than 50% of sampled eligible students responded to the survey, the school was considered to be non-responding

for the purpose of estimating participation rates. Furthermore, schools with less than 25% student participation were removed from the database and weights for the remaining schools from the stratum were adjusted to take into account their non-response.

Within schools, all eligible students in the school were considered to be non-responding if they were absent from the survey session, or were present but did not manage to complete at least 50% of the questions asked in the survey. Table 11.8 and Table 11.9 show the participation statistics for both cohorts in each Site.

Table 11.8. School and student participation statistics – younger cohort

Participating Site	Sampled Schools	Excluded Schools	Eligible Schools	Non Participating Schools		Participating Schools		Eligible Students	
				Low student participation	Refusals	Sampled	Substitute	Respondent	Non respondent
Bogota	90	0	90	0	3	77	10	3,415	44
Daegu	80	0	80	6	2	70	2	2,895	832
Helsinki	84	0	84	0	1	83	0	3,034	325
Houston	76	0	76	0	2	74	0	3,333	304
Manizales	85	0	85	0	2	83	0	3,226	357
Moscow	78	1	77	1	0	76	0	3,339	341
Ottawa	89	0	89	0	0	87	2	3,250	473
Sintra	52	2	50	5	2	43	0	1,952	1,401
Suzhou	76	0	76	0	0	76	0	3,633	131
Istanbul	101	2	99	3	7	89	0	2,693	509

Table 11.9. School and student participation statistics – older cohort

Participating Site	Sampled Schools	Excluded Schools	Eligible Schools	Non Participating Schools		Participating Schools		Eligible Students	
				Low student participation	Refusals	Sampled	Substitute	Respondent	Non respondent
Bogota	84	0	84	0	2	74	8	3,356	48
Daegu	78	0	78	2	0	73	3	3,290	517
Helsinki	57	0	57	5	1	51	0	2,338	869
Houston	45	0	45	0	0	45	0	3,101	540
Manizales	72	1	71	0	1	70	0	3,531	435
Moscow	78	1	77	1	0	75	1	3,412	288
Ottawa	62	0	62	3	2	57	0	2,176	921
Sintra	33	1	32	7	2	23	0	1,383	1,708
Suzhou	75	0	75	0	0	75	0	3,613	111
Istanbul	81	0	81	3	1	77	0	3,108	610

11.5. Participation rates

SSES reports weighted and unweighted participation rates for schools and students, as well as overall participation rates that are a combination of both. To distinguish between participation based solely on originally sampled schools and participation that also relies on substitute schools, school and overall participation rates are computed separately for

originally sampled schools only and for originally sampled together with substitute schools.

11.5.1. Unweighted school participation rate

The unweighted school participation rate is the ratio of the number of participating schools to the number of originally sampled schools, excluding any sampled schools found to be ineligible. A school is considered to be a participating school if it has a student participation rate of at least 50%. The two unweighted school participation rates are calculated as follows:

$\boldsymbol{R_{unw}^{sc-s}}$ unweighted school participation rate for originally sampled schools only

$\boldsymbol{R_{unw}^{sc-r}}$ unweighted school participation rate, including originally sampled and first and second substitute schools

$$R_{unw}^{sc-s} = \frac{n_s}{n_s + n_{sub1} + n_{sub2} + n_{nr}} \qquad \text{Eq. 11:1}$$

$$R_{unw}^{sc-r} = \frac{n_s + n_{sub1} + n_{sub2}}{n_s + n_{sub1} + n_{sub2} + n_{nr}} \qquad \text{Eq. 11:2}$$

Where $\boldsymbol{n_s}$ corresponds to the number of originally sampled schools who participated; $\boldsymbol{n_{sub1}}$ the number of participating first substitute schools; $\boldsymbol{n_{sub2}}$ the number of participating second substitute schools; and $\boldsymbol{n_{nr}}$ the number of sampled schools who refused to participate and were not replaced by any substitute schools or schools who participated but student participation within the school was less than 50%.

11.5.2. Unweighted student participation rate

The unweighted student participation rate is the ratio of the number of selected students that participated in SSES to the total number of eligible, sampled students from the participating Site. The unweighted student participation rate $\boldsymbol{R_{unw}^{st}}$ is computed as follows:

$$R_{unw}^{st} = \frac{\sum_{i\in(Y\cup S)} s_{rs}^{i}}{\sum_{i\in(Y\cup S)} s_{rs}^{i} + \sum_{i\in(Y\cup S)} s_{nr}^{i}} \qquad \text{Eq. 11:3}$$

Where **Y** denotes the set of responding original sample schools with eligible students, **S** denotes the set of responding substitute schools for which the corresponding original sample school was eligible but was non-responding; $\boldsymbol{s_{rs}^{i}}$ is the number of participating students in school $\boldsymbol{i}$; $\boldsymbol{s_{nr}^{i}}$ the number of eligible sampled students who did not participate in the survey.

11.5.3. Overall unweighted participation rate

The overall unweighted participation rate is the product of the unweighted school and student participation rates. Because SSES computes two versions of the unweighted school participation rate, one based on originally sampled schools only and the other including substitutes as well as originally sampled schools, there also are two overall unweighted

participation rates:

$\boldsymbol{R_{unw}^{ov-s}}$ unweighted overall participation rate for originally sampled schools only

$\boldsymbol{R_{unw}^{ov-r}}$ unweighted overall participation rate, including originally sampled and first and second substitute schools

$$R_{unw}^{ov-s} = R_{unw}^{sc-s} \times R_{unw}^{st} \quad \text{Eq. 11:4}$$

$$R_{unw}^{ov-r} = R_{unw}^{sc-r} \times R_{unw}^{st} \quad \text{Eq. 11:5}$$

11.5.4. Weighted school participation rate

The weighted school participation rate is the ratio of the estimated population represented by participating schools to the estimated population represented by the eligible sampled schools. These population estimates are obtained through application of the school base weight.

Weighted school participation rates are computed for originally sampled schools and for originally sampled and substitute schools combined, as follows:

$\boldsymbol{R_{wtd}^{sc-s}}$ weighted school participation rate for originally sampled schools only

$\boldsymbol{R_{wtd}^{sc-r}}$ weighted school participation rate, including originally sampled and first and second substitute schools

$$R_{wtd}^{sc-s} = \frac{\sum_{i \in Y} W1_i \times ENR_i}{\sum_{i \in (Y \cup S \cup N)} W1_i \times ENR_i} \quad \text{Eq. 11:6}$$

Where $\boldsymbol{Y}$ denotes the set of responding original sample schools with eligible students; $\boldsymbol{S}$ denotes the set of responding substitute schools for which the corresponding original sample school was eligible but was non-responding; $\boldsymbol{N}$ denotes the set of non-responding original sample schools that were not substituted; $\boldsymbol{W1_i}$ denotes the base weight for school $\boldsymbol{i}$, and $\boldsymbol{ENR_i}$ denotes the enrolment size of age-eligible students, as indicated on the sampling frame.

The weighted school response rate, after substitution, is similar to Eq. 11:7, except the set of substitute schools $\boldsymbol{S}$ is added to the numerator, and is given by the formula:

$$R_{wtd}^{sc-r} = \frac{\sum_{i \in (Y \cup S)} W1_i \times ENR_i}{\sum_{i \in (Y \cup S \cup N)} W1_i \times ENR_i} \quad \text{Eq. 11:7}$$

11.5.5. Weighted student participation rate

For weighted student participation rates, the same number of students appears in the numerator and denominator as for the unweighted rates, but each student is weighted by its student base weight. This is given as the product of the school base weight ($W1_i$) – for the school in which the student was enrolled – and $W2_{ij}$, the reciprocal of the selection probability of student ***j*** within school ***i***.

The weighted student participation rate is computed as follows:

$$R_{wtd}^{st} = \frac{\sum_{i,j}^{s+sub1+sub2} W1_i \times W2_{ij} \times s_{rs}^{i,j}}{\sum_{i,j}^{s+sub1+sub2} W1_i \times W2_{ij} \times s_{rs}^{i,j} + \sum_{i,j}^{s+sub1+sub2} W1_i \times W2_{ij} \times s_{nr}^{i,j}} \quad \text{Eq. 11:8}$$

where both the numerator and denominator are summations across all participating schools in the Site.

11.5.6. Overall weighted participation rate

The overall weighted participation rate is the product of the weighted school and student participation rates. Because there are two versions of the weighted school participation rate, one based on originally sampled schools only and the other including substitute as well as originally sampled schools, there also are two overall weighted participation rates:

$\boldsymbol{R_{wtd}^{ov-s}}$ weighted overall participation rate for originally sampled schools only

$\boldsymbol{R_{wtd}^{ov-r}}$ weighted overall participation rate, including originally sampled and first and second substitute schools

$$R_{wtd}^{ov-s} = R_{wtd}^{sc-s} \times R_{wtd}^{st} \quad \text{Eq. 11:9}$$

$$R_{wtd}^{ov-r} = R_{wtd}^{sc-r} \times R_{wtd}^{st} \quad \text{Eq. 11:10}$$

Weighted school, class, student and overall participation rates are computed for each SSES participant using these procedures. The results are shown in Table 11.10 and Table 11.11.

According to the participation rates set forth in the Technical Standards, each Site had to aim to secure a school participation rate of at least 85% of eligible schools, and 80% of eligible students, both in terms of weighted indicators, for an overall response rate of at least 68%.

Table 11.10. School and student response rates – younger cohort

Participating Site	Unweighted Response Rates					Weighted Response Rates				
	School		Student	Overall		School		Student	Overall	
	Sampled only	Sampled and Substitute		Sampled only	Sampled and Substitute	Sampled only	Sampled and Substitute		Sampled only	Sampled and Substitute
Bogota	85.6%	96.7%	98.7%	84.5%	95.4%	87.8%	98.5%	98.7%	86.6%	97.3%
Daegu	87.5%	90.0%	77.7%	68.0%	69.9%	88.4%	91.2%	77.9%	68.8%	71.0%
Helsinki	98.8%	98.8%	90.3%	89.2%	89.2%	99.8%	99.8%	90.2%	90.0%	90.0%
Houston	97.4%	97.4%	91.6%	89.2%	89.2%	98.1%	98.1%	92.0%	90.2%	90.2%
Manizales	97.6%	97.6%	90.0%	87.9%	87.9%	98.9%	98.9%	90.0%	89.0%	89.0%
Moscow	98.7%	98.7%	90.7%	89.6%	89.6%	98.7%	98.7%	90.6%	89.4%	89.4%
Ottawa	97.8%	100.0%	87.3%	85.3%	87.3%	97.9%	100.0%	87.3%	85.5%	87.3%
Sintra	86.0%	86.0%	58.2%	50.1%	50.1%	82.5%	82.5%	58.2%	48.0%	48.0%
Suzhou	100.0%	100.0%	96.5%	96.5%	96.5%	100.0%	100.0%	96.5%	96.5%	96.5%
Istanbul	89.9%	89.9%	84.1%	75.6%	75.6%	96.1%	96.1%	84.7%	81.4%	81.4%

Table 11.11. School and student response rates – older cohort

Participating Site	Unweighted Response Rates					Weighted Response Rates				
	School		Student	Overall		School		Student	Overall	
	Sampled only	Sampled and Substitute		Sampled only	Sampled and Substitute	Sampled only	Sampled and Substitute		Sampled only	Sampled and Substitute
Bogota	88.1%	97.6%	98.6%	86.9%	96.2%	87.8%	97.7%	98.7%	86.7%	96.5%
Daegu	93.6%	97.4%	86.4%	80.9%	84.2%	93.9%	98.0%	86.3%	81.0%	84.5%
Helsinki	89.5%	89.5%	72.9%	65.2%	65.2%	87.6%	87.6%	75.7%	66.3%	66.3%
Houston	100.0%	100.0%	85.2%	85.2%	85.2%	100.0%	100.0%	84.0%	84.0%	84.0%
Manizales	98.6%	98.6%	89.0%	87.8%	87.8%	99.1%	99.1%	89.0%	88.3%	88.3%
Moscow	97.4%	98.7%	92.2%	89.8%	91.0%	97.7%	98.7%	92.6%	90.5%	91.4%
Ottawa	91.9%	91.9%	70.3%	64.6%	64.6%	99.4%	99.4%	70.7%	70.3%	70.3%
Sintra	71.9%	71.9%	44.7%	32.2%	32.2%	74.5%	74.5%	48.8%	36.4%	36.4%
Suzhou	100.0%	100.0%	97.0%	97.0%	97.0%	100.0%	100.0%	96.8%	96.8%	96.8%
Istanbul	95.1%	95.1%	83.6%	79.5%	79.5%	95.9%	95.9%	84.8%	81.4%	81.4%

11.6. Teacher questionnaire response rates

Unweighted response rates for teachers were created using similar methods to those for unweighted student and school response rates – that is, ineligible or excluded teachers were not used in the denominator for the rate calculation. Also, since teacher response is mainly a function of student participation (teachers were selected to provide information on the participating student's background), the numerator represents the number of student background questionnaire filled by teachers, while the denominator corresponds to all eligible students within the school, as shown in Table 11.12. The Technical Standards required a minimum of 80% response rate for teachers. Most cities complied with this standard, except for Daegu, Houston, Ottawa and Sintra.

Table 11.12. Teacher questionnaire response rates

Participating Site	Younger Cohort			Older Cohort		
	Teacher Unweighted Response Rate	Teacher Numerator	Teacher Denominator	Teacher Unweighted Response Rate	Teacher Numerator	Teacher Denominator
Bogota	98.1%	3,392	3,459	99.2%	3,377	3,404
Daegu	68.4%	2,548	3,727	74.4%	2,833	3,807
Helsinki	82.3%	2,763	3,359	84.6%	2,712	3,207
Houston	79.4%	2,889	3,637	72.9%	2,654	3,641
Manizales	85.0%	3,045	3,583	82.2%	3,259	3,966
Moscow	90.3%	3,322	3,680	93.1%	3,446	3,700
Ottawa	49.4%	1,840	3,723	31.3%	969	3,097
Sintra	60.7%	2,036	3,353	48.4%	1,496	3,091
Suzhou	96.4%	3,629	3,764	94.8%	3,530	3,724
Istanbul	85.4%	2,735	3,202	89.5%	3,328	3,718

11.7. Principal questionnaire response rates

The principal of each participating school was asked to respond to a questionnaire. To assess the response rate, the numerator is the number of principals who responded to thequestionnaire, while the denominator corresponds to all eligible schools in the Site and cohort, as shown in Table 11.8 and Table 11.9. The results for both cohorts are presented in Table 11.13.

Table 11.13. Principal questionnaire response rates

Participating Site	Younger Cohort			Older Cohort		
	Principal Unweighted Response Rate	Principal Numerator	Principal Denominator	Principal Unweighted Response Rate	Principal Numerator	Principal Denominator
Bogota	96.7%	87	90	96.4%	81	84
Daegu	86.3%	69	80	85.9%	67	78
Helsinki	76.2%	64	84	68.4%	39	57
Houston	85.5%	65	76	88.9%	40	45
Manizales	85.9%	73	85	87.3%	62	71
Moscow	98.7%	76	77	100.0%	77	77
Ottawa	67.4%	60	89	80.6%	50	62
Sintra	92.0%	46	50	87.5%	28	32
Suzhou	98.7%	75	76	96.0%	72	75
Istanbul	90.9%	90	99	92.6%	75	81

11.8. Parent questionnaire response rates

For the parent questionnaire, the numerator represents the number of parent questionnaires returned and filled with enough responses. The denominator corresponds to all eligible students within the Site. The results for both cohorts are shown in Table 11.14.

Table 11.14. Parent questionnaire response rates

Participating Site	Younger Cohort			Older Cohort		
	Parent Unweighted Response Rate	Parent Numerator	Parent Denominator	Parent Unweighted Response Rate	Parent Numerator	Parent Denominator
Bogota	93.1%	3,222	3,459	89.2%	3,036	3,404
Daegu	34.8%	1,298	3,727	21.1%	803	3,807
Helsinki	27.7%	932	3,359	19.1%	611	3,207
Houston	15.9%	577	3,637	10.6%	385	3,641
Manizales	74.2%	2,657	3,583	69.3%	2,748	3,966
Moscow	84.5%	3,108	3,680	79.0%	2,922	3,700
Ottawa	15.6%	582	3,723	12.1%	375	3,097
Sintra	21.6%	725	3,353	12.6%	390	3,091
Suzhou	95.4%	3,589	3,764	94.3%	3,510	3,724
Istanbul	61.7%	1,975	3,202	54.7%	2,033	3,718

Annex 11.A. School selection process

Following the procedures outlined by LaRoche, Joncas and Foy (2016[2]), SSES employs a random-start fixed-interval systematic sampling approach to draw the school sample, with each school selected with probability proportional to size (PPS). The procedure for selecting schools is as follows.

Step One: Initial setup

To begin, the schools from a stratum are sorted by any specified implicit (that is, within-stratum) stratification variables and by their enrolment size. As an example, the first 50 schools in a hypothetical sampling frame are shown below. For each record, the school's measure of size (MOS)[1] is listed. The cumulative MOS for each successive school is also listed. The sum of the MOS for all schools in the stratum (the Stratum Total Measure of Size) for this example is 97734.

Step Two: Compute the sampling interval

To determine the sampling interval, the Total Measure of Size is divided by the number of schools required for the sample. In the example, 50 schools are required.

Equation 1

$$\mathbf{97734 \div 50 = 1954.68}$$

'Certainty schools'

In some cases, the MOS for a very large school may exceed the sampling interval. Such schools become automatic selections – so-called 'certainty schools' – into the sample. Before proceeding to Step Three, certainty schools are removed from the stratum frame, and the sampling parameters are adjusted accordingly.

Step Three: Generate a random start

The sampling interval is multiplied by a random number, generated from a uniform (0, 1) distribution. The first school in the frame with a cumulative MOS higher than the resulting random start value is designated the first school in the sample.

Equation 2

$$\mathbf{1954.68 \times 0.7764 = 1517.6136}$$

$$\boldsymbol{School\ S1221, with\ a\ cumulative\ MOS\ of\ 1599, is\ the\ first\ school\ in\ the\ sample.}$$

Step Four: Identify the next school in the sample

The sampling interval is added to the number computed in the previous step in order to select the second school in the sample.

Equation 3

$$1517.6136 + 1954.68 = 3472.2936$$
$$School\ S1878, with\ a\ cumulative\ MOS\ of\ 3506, is\ the\ second\ school\ in\ the\ sample.$$

This step is repeated until the required number of schools have been sampled. For example, the third school will be identified using the following:

Equation 4

$$3472.2936 + 1954.68 = 5426.9736$$
$$School\ S2505, with\ a\ cumulative\ MOS\ of\ 5491, is\ the\ third\ school\ in\ the\ sample.$$

Step Five: Identify substitute schools

For each sampled school that is selected using this process, two substitute schools will also be identified. The first substitute (**sub1**) will be the school directly following the sampled school in the sampling frame, and the second substitute (**sub2**) will be the school that directly precedes the sampled school. There are some exceptions for this allocation method. For example, where the sampled school is the very last school in the explicit stratum, the school prior to the sampled school will become **sub1** and the school prior to **sub1** will be allocated as **sub2**.

Table 11.15. School selection process

Sampling Parameters			School Identifier	Implicit Stratum Label	Measure of Size	Cumulative Measure of Size	Sampled and Substitute Schools
Stratum Total MOS	97734		S0386	1,1	223	223	
			S1990	1,1	214	437	
School Sample Size	50		S0022	1,1	214	651	
			S2405	1,1	207	858	
Sampling interval	1954.68		S0747	1,1	192	1050	
			S2417	1,1	184	1234	
Random number	0.7764		S0897	1,1	184	1418	**sub2**
			S1221	1,1	181	1599	**s**
Random start	1517.6136		S2860	1,1	178	1777	**sub1**
			S1950	1,1	171	1948	
Step one			S2165	1,1	171	2119	
Set up frame as shown.			S1691	1,1	160	2279	
Step Two			S0791	1,1	160	2439	
Compute the sampling interval.			S0039	1,1	155	2594	
			S2574	1,1	149	2743	
97734 ÷ 50 = 1954.68			S0833	1,1	149	2892	
Step Three			S1409	1,2	120	3012	
Generate a random start.			S2790	1,2	121	3133	
0.7764 × 1954.68 = 1517.6136			S1797	1,2	124	3257	
			S2310	1,2	124	3381	**sub2**

Step Four (repeat until complete)		S1878	1,2	125	3506	**s**
		S0596	1,2	125	3631	**sub1**
Identify the next school in the sample.		S2303	1,2	125	3756	
		S0794	1,2	129	3885	
1517.6136 + 1954.68 = 3472.2936		S1193	1,2	153	4038	
		S2816	1,2	153	4191	
Step Five		S0444	1,2	153	4344	
Identify substitute schools		S1417	1,2	160	4504	
(sub1, sub2)		S2508	1,2	173	4677	
		S2163	1,3	183	4860	
		S0950	1,3	179	5039	
		S2395	1,3	168	5207	
		S2494	1,3	144	5351	**sub2**
		S2505	1,3	140	5491	**s**
		S1121	1,3	135	5626	**sub1**
		S1850	1,3	134	5760	
		S0389	1,3	130	5890	
		S0698	1,3	124	6014	

Source: Adapted from: (LaRoche, Joncas and Foy, 2016[2]) Appendix 3C.

References

LaRoche, S., M. Joncas and P. Foy (2016), *Sample design in TIMSS 2015*, http://timss.bc.edu/publications/timss/2015-methods/chapter-3.html. [2]

OECD (2017), *PISA 2015 Technical Report*, OECD Publishing, https://www.oecd.org/pisa/data/2015-technical-report/. [1]

Notes

[1] The MOS for larger schools is equal to the enrolment size for the target population. For smaller schools, the MOS is adjusted (see Chapter 4 Sample Design).

Chapter 12. Scaling SSES Data

This chapter describes the analyses of the OECD Survey Social and Emotional Skills 2019 (SSES) Main Study data and the assessment of psychometric properties of items and scales that involved a series of iterative modelling and analysis steps. These steps included the application of confirmatory factor analysis (CFA) to evaluate constructs; multiple-group confirmatory factor analysis (MGCFA) to review measurement equivalence across groups (age cohorts, participating cities and gender groups); and IRT (Item Response Theory) Generalised Partial Credit Model (GPCM) to scale items and generate scores for Study participants. The following section provides an overview of the scaling methodology.

12.1. Introduction

The SSES conceptual framework (Chernyshenko, Kankaraš and Drasgow, 2018[1]) was developed in reference to the 'Big Five Model' (John, Naumann and Soto, 2008[2]) with the aim of measuring social and emotional skills among young people. The SSES scaling work was initially based on a theoretical identification of 19 potential latent constructs that were grouped into five broad domains – Task performance, Emotional regulation, Collaboration, Open-mindedness and Engaging with others (see Chapter 2 for more information). Furthermore, there was the expectation of measuring 'compound skills' that were defined as a combination of two or more skills. After an extensive analysis of Field Test (FT) data, an evaluation of latent constructs and considerations regarding the appropriate length of the SSES assessment in the final data collection, 15 scales (see Table 12.1) across the aforementioned five broad domains were selected for measurement in the Main Study (MS). In addition, two compound skills scales of Self-efficacy (EFF) and Achievement motivation (MOT) were derived based on items selected from the original item pool that also contributed to the measurement of the other scales. The individual items in each scale can be found in Annex B.

Table 12.1. The SSES assessment scales

Domain	Skill Label	Skill
Collaboration	EMP	Empathy
	TRU	Trust
	COO	Cooperation
Emotional regulation	EMO	Emotional control
	OPT	Optimism
	STR	Stress resistance/resilience
Engaging with others	ASS	Assertiveness
	ENE	Energy
	SOC	Sociability
Open-mindedness	CRE	Creativity
	CUR	Curiosity
	TOL	Tolerance
Task performance	PER	Perseverance/Persistence
	RES	Responsibility
	SEL	Self-control
Compound skills	EFF	Self-efficacy
	MOT	Achievement motivation

The analyses of the SSES MS data and the assessment of psychometric properties of items and scales involved a series of iterative modelling and analysis steps. These steps included the application of confirmatory factor analysis (CFA) to evaluate constructs; multiple-group confirmatory factor analysis (MGCFA) to review measurement equivalence across groups (age cohorts, participating cities and gender groups); and IRT (Item Response Theory) Generalised Partial Credit Model (GPCM) to scale items and generate scores for Study participants. The following section provides an overview of the scaling methodology. The procedures for assessing the psychometric properties of the item material and for scaling the assessment data collected from students and parents included the following aspects:[1]

- Reviewing classical item/scale statistics
    a. Item analysis of missing data

   b. Response time analyses
   c. Estimating scale reliabilities

- Estimating CFA models[2]
   a. Estimating models with predefined constructs by cohort and for the combined dataset
   b. Reviewing model outcomes
   c. Refining constructs including item deletions
   d. Re-estimating CFA models
   e. Reviewing scale reliabilities
- Reviewing measurement invariance with MGCFA
   a. Reviewing levels of measurement invariance across three groups: cohort, gender and cities
   b. Reviewing invariance levels controlling for acquiescent response
- Calibrating items with GPCM
   a. Estimating GPCM separately for cities and cohorts
   b. Fitting GPCM to final scales with concurrent international calibration data
- Estimating scale scores
   a. Estimating participant weighted likelihood estimates (WLEs) using GPCM with the concurrent international data
   b. Standardising scale scores
   c. Adjusting scale scores to account for acquiescent response style (ARS)

For the teacher indirect assessment of students, the International Contractor calculated averages of the three raw scores per scale provided by teachers for each assessed student, subsequent to recoding of negatively worded items.

The following sections describe each step in detail and the outcomes of the scaling analyses.

12.2. Data used for analyses

After all data cleaning procedures were completed the data underwent a further review before the scaling analysis was carried out. The target population details and sampling outcomes can be found in Chapter 11. Table 12.2 shows the percentages of students by gender at each Site and at the international level by cohort.

Table 12.2. Percentage of students by gender cohort and Site

Site	Younger cohort			Older cohort		
	Female	Male	Other*	Female	Male	Other*
Bogotá	50.5	48.9	0.6	49.4	49.8	0.8
Daegu	49.2	50.6	0.1	51.0	48.6	0.4
Helsinki	49.6	49.4	1.0	51.8	46.3	1.9
Houston	50.2	48.8	1.0	52.0	46.9	1.1
Manizales	50.5	48.7	0.8	51.7	47.7	0.6
Moscow	48.4	51.6	0.0	49.0	51.0	0.0
Ottawa	48.5	48.0	3.5	48.8	49.6	1.6
Sintra	51.7	46.4	2.0	51.7	46.9	1.3
Suzhou	45.5	54.4	0.1	48.6	51.1	0.2
Istanbul	49.9	50.1	0.0	58.7	41.3	0.0
Pooled sample	49.3	49.8	0.9	51.2	48.1	0.7

Note: *Students who did not identify as either female or male are included in the 'other' category.

The SSES student survey in Sintra (Portugal) did not meet the sample participation requirements for SSES and was not included in the data for estimating the scaling parameters for the student direct assessment. For the parent indirect assessment, only five cities (Bogotá (Colombia), Manizales (Colombia), Moscow (the Russian Federation), Suzhou (People's Republic of China) and Istanbul (Turkey)) met the sample participation requirements and were used for the estimation of scaling parameters for the parent indirect assessment. The eligible Site data were combined to form a set of calibration data using weights that ensured each Site had the same contribution to the parameter estimation. Data from cities that were not used for the estimation of scaling parameters were scaled using the model parameters estimated with the data from the calibration sample. The number of eligible records used for scaling from each Site can be found in Table 10.4 in Chapter 10 Data Management Processes.

12.3. Classical item statistics

All items in the SSES assessments had a Likert-type format with five categories and included both positively and negatively worded statements. The five categories were 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree' and 'strongly agree'. Each item was scored from 0 to 4 for items with positively worded statements and reverse-scored for the negatively worded ones (Table 12.3).

Table 12.3. Item score for the scaling

Response category	Score for positively worded items	Score for negatively worded items
Strongly disagree	0	4
Disagree	1	3
Nether agree nor disagree	2	2
Agree	3	1
Strongly agree	4	0

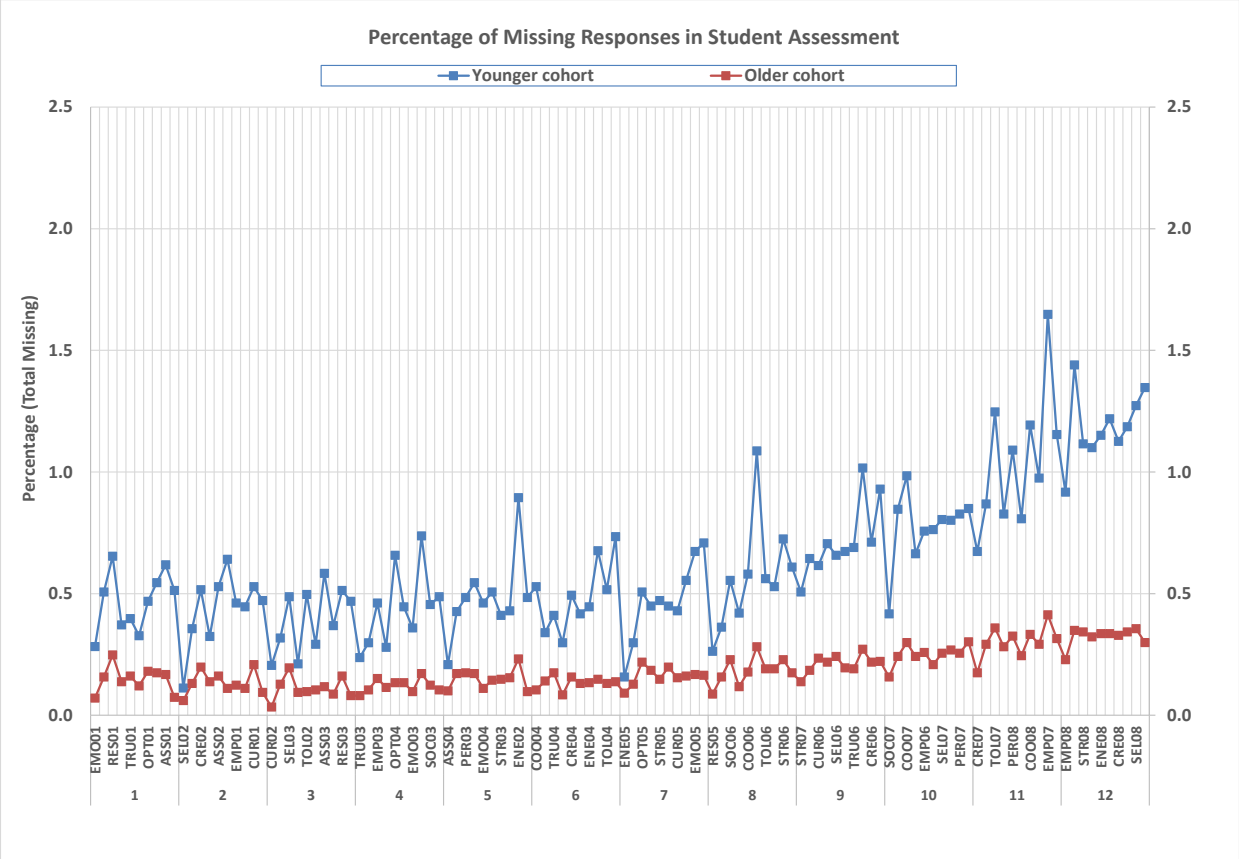
During an initial analysis step, the International Contractor examined percentages of missing item responses and patterns of missing values, the distribution of item scores, item-total correlations and internal scale consistencies (Cronbach's alpha).

12.3.1. Item analysis of missing data

As described in Chapter 10 Data Management, there were three types of missing values. A missing value was classified as 'omitted' when the respondent had the opportunity to answer the question but did not provide a response. This type of missing value was coded as '9' in the database. A missing value was categorised as 'not administered' if an item was not presented and the respondent had no opportunity to respond. The corresponding values were coded as '7'. A third type of missing value relates to cases where a response could not be provided due to an unexpected issue (such as an error in the computer-based delivery system). This type of missing value occurred rarely and was coded as '8'. In the analyses of missing data, only 'omitted' missing values were considered.

The following graphs (Figure 12.1, Figure 12.2 and Figure 12.3) depict the percentages of missing responses in student, parent and teacher assessment data by cohort. Only the missing responses due to omission by respondents were included in these graphs. The items are shown (on the horizontal axis) in the sequential order of how respondents completed the assessment.

Figure 12.1. Percentage of missing responses in student direct assessment by cohort



Overall, the percentages of missing responses were very low. On average across cities, the percentages of missing responses across all 120 assessment items in the student direct assessments were 0.6% and 0.2%, for the younger and older cohorts, respectively. For parents, the average percentages of missing responses were below 1.0% for both cohorts. For teacher data, the average percentages of missing responses per assessed student across

45 items were below 0.2% for both cohorts. These results show that the students from the younger cohort tended to have a somewhat higher percentage of missing responses compared to the students from the older cohort. Not unexpectedly, there were hardly any differences in the percentages of missing values in the data collected from indirect parent and teacher assessments. The number of missing responses increased notably towards the end of the assessment in the data collected from students in the younger cohort, compared to only slight increases from students in the older cohort.

The corresponding results from the indirect assessment among parents also show that percentages of missing responses increased in the second half of the assessment across both cohorts, which suggests there were some (but relatively few) parents who started but did not complete this indirect assessment.

Figure 12.2. Percentage of missing responses in parent indirect assessment by cohort

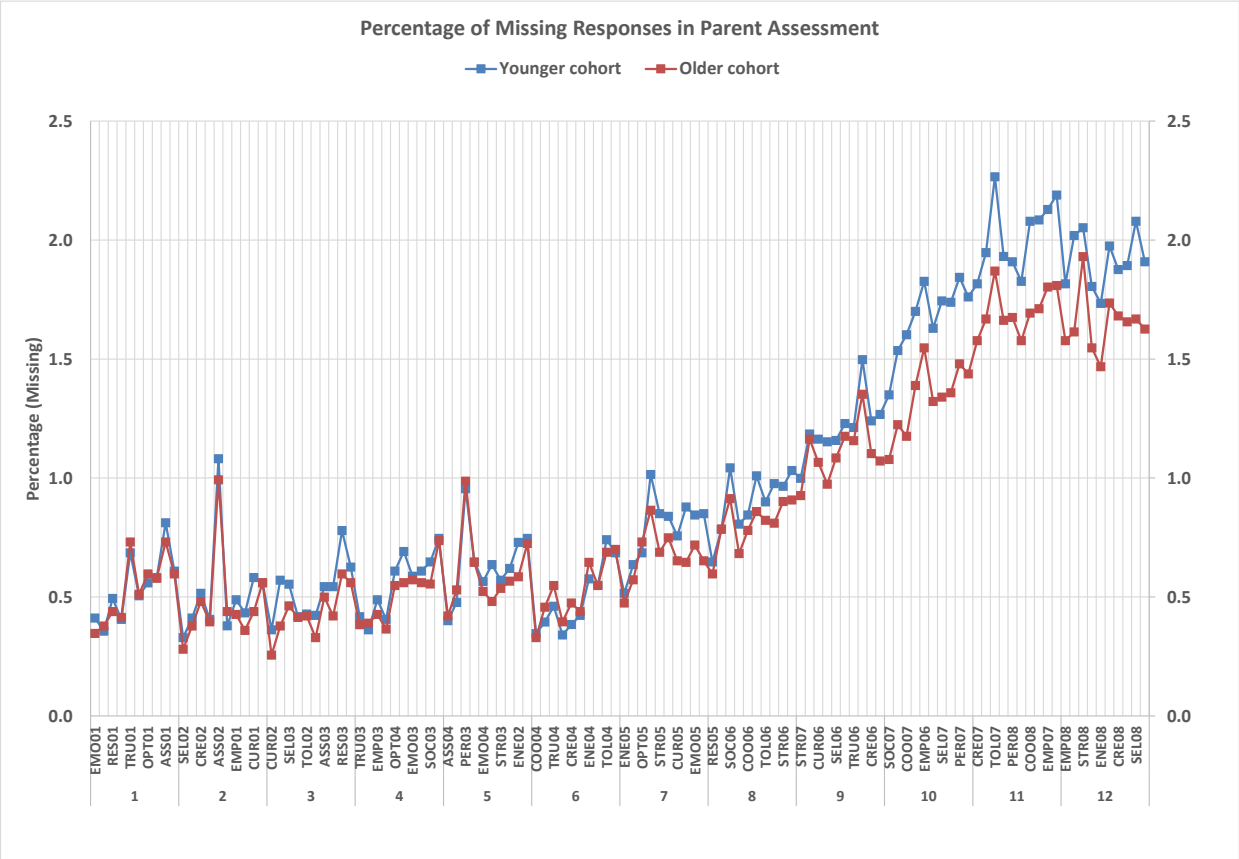
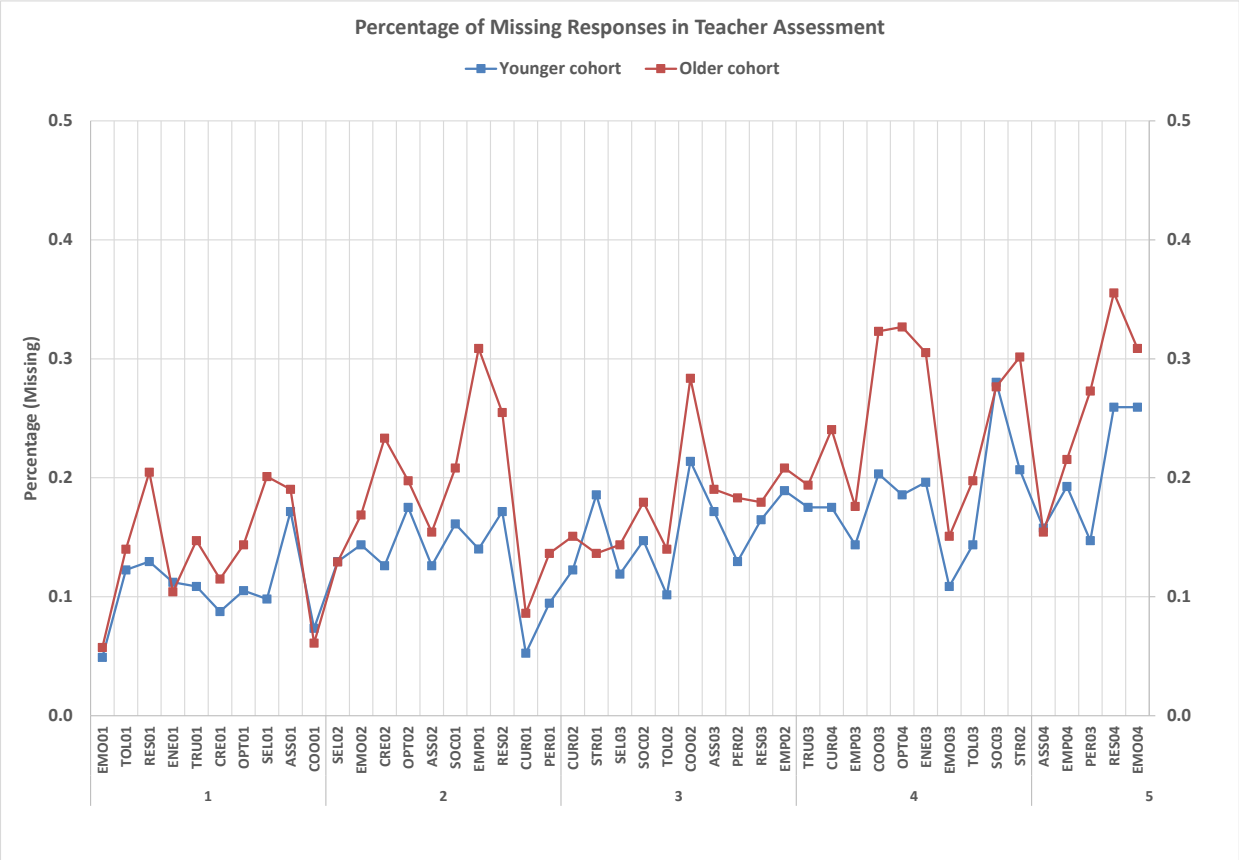


Figure 12.3 Percentage of missing responses in teacher indirect assessment by cohort



12.3.2. Response time analyses

The RM Assessment Master (AM) online delivery platform captured session starting and ending times. This information was used to compute the time spent on the assessment by students. The timing results presented in Table 12.4 are based on valid responses as described previously. Because the student assessment and contextual questionnaire were administered in a single online session, the session time was combined for both parts. Students were expected to complete both the assessment and the contextual questionnaire in about 90 minutes.

An initial analysis showed that the timing distributions were quite skewed, with a number of very short and very lengthy assessment times recorded. Very short response times recorded may have been due to an unexpected technical issue while very long response times could be explained in part by students who did not logout after completing the assessment. For this reason, only those students with an assessment length in the range of 15 to 120 minutes were used to examine student timing information. Table 12.4 shows the percentages of valid students included and excluded from the response time analysis by cohort. It shows that across the pooled SSES dataset, approximately 83% among the younger cohort students and 88% among those from the older cohort had valid students' response timing information that could be used for the analysis. No valid timing information was recorded for 12.6% and 9.9% of students from the younger and older cohorts respectively. As the indirect assessments by parents and teachers were not timed, the International Contractor did not conduct any timing analysis with these data.

Table 12.4. Percentage of response time information

Cohort	Student data included in the analysis (%)	Students with no timing information (%)	Students with extremely short sessions (%)	Students with extremely long sessions (%)
Younger	83.07	12.54	0.14	4.25
Older	87.59	9.89	0.15	2.37

Table 12.5 shows descriptive statistics with timing information for students by cohort. The results are presented for each of the two cohorts separately based on the average percentages for the pooled SSES data. On average, younger cohort students completed the entire assessment in 59.6 minutes while older cohort students spent approximately 14 minutes less and completed the entire assessment in 45.9 minutes. The analysis showed a somewhat larger standard deviation among students from the younger cohort than among those from the older cohort.

Table 12.5. Students timing data of the respondent type

Cohort	Number of students	Min	Q1	Median	Mean	Q3	Max	SD
Younger	25907	15.5	48.5	58.5	59.6	68.3	110.0	16.2
Older	26100	15.1	35.7	43.7	45.9	53.1	119.7	15.1

The International Contractor also reviewed the variation in response times for student data from the 10 cities within each cohort. Table 12.6 shows that the students in Daegu (Korea) on average spent the shortest time when providing responses while there was only less variation in average response time across the other nine cities. Within cities the largest variation of response time was recorded for students in Suzhou (People's Republic of China), in particular those in the older cohort.

Table 12.6. Variability of time used by students by Site

Cohort	Site	Number of students	Min	Q1	Median	Mean	Q3	Max	SD
Younger	Bogotá	2888	16.0	54.3	64.1	65.9	75.4	120.0	16.3
	Daegu	2785	17.4	35.3	41.0	42.5	48.5	116.9	11.0
	Helsinki	1978	18.8	49.2	56.9	56.9	63.9	111.6	11.5
	Houston	3038	18.4	45.5	54.1	54.1	63.2	101.9	11.3
	Manizales	2936	24.5	55.2	64.0	65.9	73.9	119.9	15.7
	Moscow	3097	15.5	52.0	60.4	60.9	68.8	118.1	13.3
	Ottawa	2037	15.6	48.6	57.4	58.0	66.5	118.3	13.1
	Sintra	2088	24.7	54.9	63.2	62.6	69.7	113.5	11.4
	Suzhou	2792	21.3	52.7	64.1	67.8	80.3	119.9	20.2
	Turkey	2268	16.4	49.3	58.2	60.9	69.4	119.4	17.3
Older	Bogotá	2906	15.5	39.9	46.1	47.7	53.9	118.0	11.8
	Daegu	3189	15.1	25.6	29.8	30.9	34.7	109.0	7.8
	Helsinki	1943	15.1	35.9	41.8	43.0	49.2	115.5	10.4
	Houston	2746	15.5	31.8	37.9	40.3	46.5	115.8	12.3
	Manizales	3379	19.2	42.8	48.8	50.9	56.2	119.7	12.5
	Moscow	3128	15.1	37.6	43.7	45.5	51.6	114.9	11.7
	Ottawa	1529	15.9	35.4	41.9	42.8	49.3	113.2	10.6
	Sintra	1564	21.0	43.5	49.9	51.1	57.4	112.8	11.3
	Suzhou	2813	15.2	43.3	55.4	59.0	71.6	119.7	20.8
	Turkey	2903	19.8	37.4	44.5	48.0	54.1	119.7	15.9

12.3.3. Scale reliabilities

The reliabilities of the 17 scales from the student, parent and teacher assessments were reviewed using the pooled SSES dataset as well as each cohort dataset separately. Two sets of scale reliability coefficients were reported: Cronbach's alpha coefficients (Cronbach, 1951[3]) and omega reliabilities (McDonald, 2013[4]). While Cronbach's alpha does not account for differences in item contributions to the measurement of a scale, the omega coefficient provides a better estimate of internal consistency when there are differences in how well the construct is measured across items (Trizano-Hermosilla and Alvarado, 2016[5]). For both measures, reliabilities above 0.7 are typically regarded as satisfactory and values above 0.8 indicate high reliability (Nunnally and Bernstein, 1994[6]).

Table 12.7 through Table 12.9 show (Cronbach's) alpha and omega coefficients for student, parent and teacher assessments of the final scales (after the selection described in the following section), respectively. The reliabilities are shown by scale by cohort and also for pooled data including both cohorts (reported as 'All'). Senate weights were used for estimating the overall scale reliabilities so that each Site meeting sample participation requirements had the same contribution to the analyses. The reliabilities for each Site dataset with student direct and parent indirect assessment data are recorded in Annex Tables G.1 and Annex G.2, respectively. Student or parent weights were used for estimating the scale reliabilities for each Site dataset.

Table 12.7. Reliability coefficients for student direct assessment final scales

Scale	All		Younger cohort		Older cohort	
	Omega	Alpha	Omega	Alpha	Omega	Alpha
ASS	0.85	0.85	0.82	0.82	0.88	0.88
COO	0.79	0.78	0.80	0.79	0.78	0.77
CRE	0.75	0.75	0.72	0.71	0.78	0.78
CUR	0.80	0.78	0.79	0.77	0.81	0.79
EMO	0.77	0.77	0.74	0.73	0.80	0.80
EMP	0.71	0.71	0.70	0.70	0.73	0.73
ENE	0.78	0.77	0.73	0.72	0.80	0.79
OPT	0.83	0.82	0.78	0.77	0.86	0.85
PER	0.82	0.82	0.79	0.78	0.85	0.84
RES	0.73	0.72	0.70	0.70	0.75	0.74
SEL	0.76	0.75	0.75	0.74	0.77	0.77
SOC	0.74	0.72	0.70	0.68	0.77	0.76
STR	0.80	0.80	0.76	0.76	0.82	0.82
TOL	0.74	0.74	0.71	0.71	0.77	0.77
TRU	0.81	0.81	0.78	0.77	0.82	0.82
EFF	0.73	0.72	0.74	0.73	0.72	0.71
MOT	0.76	0.76	0.76	0.76	0.76	0.75

Note: The compound skills scale of MOT contains 6 items.

Table 12.7 shows that all 17 scales had satisfactory reliabilities. The omega coefficients were similar to the alpha coefficients across scales. The scale reliabilities for student direct assessment data in the older cohort were overall slightly higher than those for the younger cohort, however, there were a few exceptions. All 17 scales from the older cohort and all students had reliabilities above 0.70. The reliabilities estimated from data for the younger cohort showed that 14 scales were above 0.70 while the scales Empathy (EMP) and Responsibility (RES) had reliabilities of just below or close to 0.70, and the alpha

coefficient of the scale Sociability (SOC) was the lowest, at 0.68, but its omega coefficient reached 0.70. For 14 out of 17 scales, reliabilities estimated from data for the older cohort tended to be higher than those from data for the younger cohort. However, this was not the case for the scale Cooperation (COO) and the two compound skills scales, Self-efficacy (EFF) and Achievement motivation (MOT).

Table 12.8 shows the scale reliabilities for data from the parent indirect assessment of students. Almost all 17 scales from the parent datasets for younger cohort students, for older cohort students, and for the combined dataset with all students were satisfactory (values above 0.70). The alpha coefficient of the parent data scale Energy (ENE) from data for parents of the younger cohort students was at 0.69 while the omega coefficient was at 0.71. Scale reliabilities estimated from data for parents of the older cohort were slightly higher than those from data of parents of the younger cohort.

Table 12.8. Reliability coefficients for parent indirect assessment final scales

Scale	All		Younger cohort		Older cohort	
	Omega	Alpha	Omega	Alpha	Omega	Alpha
ASS	0.82	0.82	0.80	0.80	0.83	0.83
COO	0.80	0.79	0.80	0.78	0.81	0.80
CRE	0.80	0.80	0.80	0.79	0.81	0.80
CUR	0.83	0.82	0.82	0.81	0.84	0.83
EMO	0.77	0.77	0.76	0.75	0.79	0.79
EMP	0.74	0.74	0.73	0.73	0.76	0.76
ENE	0.74	0.73	0.71	0.69	0.75	0.74
OPT	0.81	0.80	0.79	0.77	0.82	0.81
PER	0.87	0.87	0.87	0.87	0.88	0.88
RES	0.80	0.79	0.78	0.78	0.81	0.80
SEL	0.80	0.80	0.80	0.80	0.80	0.79
SOC	0.77	0.75	0.74	0.73	0.77	0.75
STR	0.81	0.80	0.80	0.80	0.81	0.81
TOL	0.77	0.75	0.75	0.73	0.78	0.77
TRU	0.81	0.80	0.79	0.78	0.82	0.82
EFF	0.78	0.77	0.77	0.76	0.78	0.78
MOT	0.84	0.83	0.83	0.83	0.84	0.84

Table 12.9 shows the scale reliabilities for data from the teacher indirect assessment of students for 15 basic scales. As there were only three items per scale, reliabilities are expected to be lower than for those for the student and parent assessment data. However, there were some scales with very high reliabilities as was the case for the scale Assertive (ASS) at about 0.90 and the scale Persistence (PER) with about 0.93. Scale reliabilities estimated from data from teachers of the older cohort were similar to those based on data from teachers of the younger cohort.

Table 12.9. Reliability coefficients for final scales of the teacher indirect assessment of students

Scale	All		Younger cohort		Older cohort	
	Omega	Alpha	Omega	Alpha	Omega	Alpha
ASS	0.91	0.90	0.92	0.89	0.91	0.90
COO	0.76	0.69	0.77	0.67	0.75	0.72
CRE	0.84	0.82	0.85	0.81	0.84	0.83
CUR	0.86	0.83	0.86	0.82	0.85	0.83
EMO	0.80	0.78	0.81	0.76	0.78	0.80
EMP	0.79	0.76	0.80	0.76	0.78	0.76
ENE	0.81	0.78	0.82	0.77	0.80	0.79
OPT	0.83	0.81	0.83	0.79	0.82	0.81
PER	0.94	0.93	0.94	0.92	0.93	0.93
RES	0.82	0.80	0.83	0.79	0.81	0.81
SEL	0.77	0.74	0.78	0.72	0.75	0.76
SOC	0.84	0.81	0.83	0.82	0.84	0.80
STR	0.71	0.68	0.72	0.67	0.70	0.68
TOL	0.88	0.86	0.88	0.86	0.88	0.86
TRU	0.77	0.75	0.77	0.75	0.76	0.74

12.4. Scaling analyses and procedures for SSES instruments

The items for the SSES assessment were developed to measure a number of constructs reflecting social and emotional skills of students from participating OECD cities and countries in two age cohorts (10-year-old and 15-year-old students). As these personal traits are latent constructs that are not directly observable, they have to be estimated based on the responses to sets of items that are designed to measure the underlying constructs. The sections below describe the scaling analyses and subsequent procedures for the scaling of assessment data provided by students and parents. Each of the student assessments completed by students and parents were designed to measure a total of 17 constructs, resulting in the computation of 15 basic scales and two compound skills scales. While for student and parent data IRT scaling was applied to derive scale scores, for teachers the International Contractor computed a simple raw score average across the three items teachers had responded to for each of 15 basic scales.

The SSES scaling analyses and calibration procedures included the following steps:

- The International Contractor conducted an analysis of measurement invariance using MGCFA and selected items for final scaling based on CFA, MGCFA and IRT analysis.
- Subsequently, MGCFA were estimated to review the robustness of these measures with respect to their comparability across cities, cohorts and gender groups.
- Finally, IRT GPCM were estimated to derive final scale scores for each of the measured constructs.

The following sections describe the methodology used for the analysis of item dimensionality and measurement invariance for the final scaling as well as the corresponding analysis results and scaling parameters.

12.4.1. Confirmatory factor analysis

Exploratory factor analysis (EFA) was conducted as part of the analysis of the SSES FT data as an earlier step to investigate how well the data from the initially developed assessment reflected the predefined latent construct. To confirm the constructs identified in the SSES framework using the assessment data, the International Contractor used CFA – an approach that is part of the more general structural equation modelling (SEM) framework (Kaplan, 2009[7]).[3]

A CFA measurement model that relates unobserved latent variables to observed individual responses (x) can be specified as:

$$x = \mu + \Lambda\eta + \varepsilon \qquad \text{Eq. 12:1}$$

Here $\boldsymbol{\mu}$ is a vector of item responses, Λ denotes a matrix of factor loading, $\boldsymbol{\eta}$ denotes the latent variable and $\boldsymbol{\varepsilon}$ is a vector of unique error variables that are independent of the common factor.

CFA measurement models for SSES items were estimated using the SEM software Mplus7 (Muthén and Muthén, 2012[8]). This software offers a number of different estimators, among others maximum likelihood (ML), robust maximum likelihood (MLR), weighted least squares (WLS) and weighted least-squares – mean and variance adjusted (WLSMV). The estimation method of ML estimation is based on the assumptions that the observed variables are continuous and normally distributed. With this assumption, ML estimation is not appropriate for analysis of categorical data such as SSES assessment data. MLR estimation is better when assumptions about normal distributions are not entirely met. Weighted least-squares estimators (WLS and WLSMV) are deemed most appropriate for categorical items. According to the research studies (Rhemtulla, Brosseau-Liard and Savalei, 2012[9]) (Suh, 2015[10]), however, under the assumption that underlying each categorical variable (x) there are approximately normally distributed continuous variables (x*), items with five or more categories may be treated as quasi-interval and different estimation methods – such as ML, MLR and WLS – yield appropriate modelling outcomes with only negligible differences between the different estimation methods. Since all SSES assessments relied on ordinal Likert-type items with five categories for measurement, it was decided to use MLR for the CFA in this Study.

In order to evaluate the extent to which the SSES data match a model with an assumed a-priori structure in CFA, a number of model fit indices were reported and reviewed. For the SSES analysis, model fit was primarily assessed through reviewing the Root Mean Square Error of Approximation (RMSEA), the Standardised Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI) and the Tucker-Lewis index (TLI) which are less affected by sample size and model complexity than other indices (Bollen and Long, 1993[11]). In addition, the Bayesian Information Criterion (BIC) was also reported; however, it should be noted that BIC is an index that is greatly influenced by sample size.[4]

In order to assist the evaluation of CFA results, threshold values were defined for the review of model fit. RMSEA and SRMR values over 0.10 were defined as having unsatisfactory model fit, values below 0.05 indicate close model fit, values between 0.05 and 0.08 satisfactory model fit, and values between 0.08 and 0.10 marginally satisfactory model fit. Fit indices CFI and TLI vary between 0 and 1. Values below 0.90 indicate poor model fit, values between 0.90 and 0.95 satisfactory model fit, and values greater than 0.95 close model fit. There is no absolute value used for the fit index BIC as it relates to the sample size, however, lower BIC values are preferred when comparing two models. It should be

noted that there are no agreed rules in the literature about the acceptability of model fit, which also may vary depending on the respective estimation method. Therefore, these thresholds were determined as guidelines for the assessment of models and should not be interpreted in the sense of pass/fail marks for CFA models.

In addition to these fit indices, standardised factor loadings and the corresponding item residual variance provided further information about the psychometric characteristics of SSES assessment data. The standardised factor loadings $\boldsymbol{\Lambda}'$ can be interpreted in the same way as standardised regression coefficients if the indicator variable is regressed on the latent factor. The loadings also reflect the extent to which each indicator measures the underlying construct. Squared standardised factor loadings indicate how much variance in an indicator variable can be explained by the latent factor and are related to the (standardised) residual variance estimate $\boldsymbol{\varepsilon}'$ (which provides an indication of the unexplained proportion of variance) as:

$$\varepsilon' = (1 - \lambda'^2) \qquad \text{Eq. 12:2}$$

Confirmatory factor analyses were carried out separately for each of 15 conceptual constructs and the two compound skills constructs by cohort, as well as for the pooled dataset including both cohorts for student direct and parent indirect assessments. All modelling was carried out with data from cities where the SSES sample participation requirements had been met. So-called *senate weights* were calculated for each respondent in a way that each Site was represented with equal weights in the model estimations (see details in Chapter 9).

For the analysis of SSES data, items with factor loadings in the range between 0.40 and 0.60 were considered as moderately related to the underlying latent factor, while those with factor loadings of 0.60 or higher were considered as strongly related to the underlying latent factor. The item standardised factor loadings from the CFA models for student direct and parent indirect assessment data can be found in Annex Table G.3 and Table G.4, respectively. The standardised factor loadings within a scale indicate the strength of the relationship between the items and the scale construct. All negatively worded items had been reversed prior to estimating CFA models.

For student direct assessment data, the standardised factor loadings for the final scales were above 0.40 for the positively worded items[5] in the older cohort and all students data in the 15 predefined scales. Three items in the younger cohort data had slightly lower factor loadings (0.37-0.39). Where possible, negatively worded items were retained in order to maintain the design feature of having both types of items even in cases where the factor loading was lower than 0.40. For parent data, all items had the standardised factor loading higher than 0.40 except for one negatively worded item (COO04).

The CFA model fit indices for student direct and parent indirect assessment data can be found in Annex Table G.5 and Table G.6, respectively. The model fit indices are shown by scale for three different sets of data: from the student or parent data from the younger cohort, from the older cohort data and from the combined two age cohorts. In general, across the different fit indices, the model fit indices for all scales had satisfactory model fit for both student direct and parent indirect assessment data.

12.4.2. The review of measurement invariance

For any international large-scale assessment, it is of crucial importance to ensure the scales developed are equivalent across different national contexts and sub-groups within assessed

populations. In the case of SSES, measurement equivalence was assessed with regard to cities (reflecting different national contexts), cohort (reflecting the two age groups participating in the Study) and gender. To review measurement invariance of SSES data, MGCFA was conducted distinguishing the two cohorts and participating cities for both student and parent data. In addition, MGCFA was also carried out for the two gender groups when reviewing student data.

A multiple-group factor model can be defined as:

$$x_g = \mu_g + \Lambda_g \eta_g + \varepsilon_g \qquad \text{Eq. 12:3}$$

Here **g** is group index that a respondent belongs to, **g=1,2,…,G**.

Factorial invariance holds when factor loadings ($\boldsymbol{\Lambda_g}$) and item intercepts ($\boldsymbol{\mu_g}$) are equivalent across the G groups. The measurement invariance level is then reviewed through comparisons of model fit indices across three models with different constraints:

- Configural invariance reflects a model with no constraints where all item parameters (factor loadings and intercepts) are group specific. For scales reaching configural invariance, neither scores nor their associations can be directly compared across groups.
- Metric invariance refers to a model where factor loadings are constrained to be equal across groups. For a scale with metric invariance it is possible to directly compare results from analysis of associations of this variable with other factors (such as correlation or linear regression). However, means of scales scores derived from a model with metric invariance cannot be directly compared across groups.
- Scalar invariance refers to a model where both factor loadings and intercepts are equal across groups. At this level of measurement equivalence, scale scores can be directly compared across groups.

Generally, model fit is expected to decrease with increasing model constraints because having more constraints in the model leaves less room for accounting for differences between groups. To judge levels of measurement invariance, it is also informative to review relative changes in model fit in addition to the overall fit of each model.

Table 12.10 lists the model fit indices and the changes in these statistics across models with different constraints, which were used to evaluate model invariance levels. The thresholds used as indicators for invariance levels follow recommendations made by Chen (2007[12]) and were also used for the analyses in TALIS 2018 (OECD, 2019[13]). It is important to keep in mind that it is not appropriate to determine measurement invariance levels with static cut-off points because of the merely indicative nature of fit indices for which there are no scientifically established criteria for acceptance or non-acceptance. Furthermore, models with otherwise very similar model fit may be placed below or above such cut-off points, which makes absolute pass- or fail-marks rather questionable. However, when drawing conclusions from scale differences across groups (such as cities or cohorts), scale invariance levels should be kept in mind in order to avoid erroneous interpretations of Study results.

Table 12.10. Difference in model fit indices change in measurement invariance level

Invariance Level	Index	Value
Configural invariance	RMSEA	≤ 0.08
	SRMR	≤ 0.06
	CFI	≥ 0.90
	TLI	≥ 0.90
Metric invariance	ΔRMSEA	≤ 0.015
	ΔSRMR	≤ 0.030
	ΔCFI	≤ 0.010
	ΔTLI	≤ 0.010
	ΔBIC	≤ 0
Scalar invariance	ΔRMSEA	≤ 0.015
	ΔSRMR	≤ 0.010
	ΔCFI	≤ 0.010
	ΔTLI	≤ 0.010
	ΔBIC	≤ 0

When reviewing student direct assessment data, MGCFA modelling was carried out considering three grouping variables: the younger (10-year-olds) and older (15-year-olds) cohorts, female and male gender groups, and cities. When reviewing data from the indirect parent assessment, these analyses were conducted using two grouping variables (cohorts and cities).

The fit of model for each scale was reviewed based on the criteria described above to assess the extent to which the empirical data matched the relationships specified in the model. The levels of measurement invariance were also assessed using MGCFA models. Results suggested that only a few of the original 8-item scales had reasonable fit indices for the overall CFA model, and most scales did not reach metric invariance across cities, even though this level of invariance was met across gender groups and cohorts. Therefore, in order to improve the model fit and to achieve metric invariance, model modifications were applied by removing problematic items. Problematic items were identified by a review of low factor loadings and model modification indices which, together with an examination of item wording, indicated instances of local dependence through the inclusion of highly similar items.

In addition, MGCFA models were estimated using acquiescence response sets as control variables as part of *multiple indicator multiple cause* (MIMIC) models, which for some items sets showed improved model fit and higher levels of measurement invariance.

Acquiescence refers to tendencies among respondents to provide their agreement or disagreement to different positively and negatively worded statements irrespective of the content, wording and direction. Such response styles may result in biased measures and calculation of acquiescence response sets (ARS) has been suggested as way of modelling such response tendencies for Likert-type items (Primi et al., 2020[14]).

One way to control for acquiescence is using a balanced scale in which positively and negatively worded items are paired within scales. One of the design features of the SSES assessment was to have both positively and negatively worded items within each item set measuring a particular construct scale. However, the items were not evenly balanced. In order to derive an acquiescence response set, 25 pairs of items across all scales were selected for both student and parent data as shown in Table 12.11.

Table 12.11. Items used for calculating acquiescence

Positive item	Negative item	Positive item	Negative item	Positive item	Negative item
ASS07	ASS05	ENE08	ENE06	SEL01	SEL05
COO02	COO04	OPT07	OPT01	SEL03	SEL08
CRE02	CRE03	OPT02	OPT08	SOC06	SOC04
CRE06	CRE08	PER02	PER03	SOC07	SOC08
CUR05	CUR06	PER04	PER05	STR01	STR02
EMO01	EMO08	PER08	PER06	TOL08	TOL06
EMO02	EMO06	RES02	RES03	TRU08	TRU05
EMP02	EMP08	RES06	RES04		
ENE01	ENE04	RES05	RES07		

The acquiescence score for a participant ***n*** was calculated as follows:

$$ARS_n = \frac{1}{2}\left(\frac{\sum_{i=1}^{25}(x_{in}^{p}) + \sum_{i=1}^{25}(x_{in}^{n})}{25}\right) \quad \text{Eq. 12:4}$$

Here, $\boldsymbol{x_{in}^{p}}$ indicates the response of participant ***n*** to a positively worded item ***i*** and $\boldsymbol{x_{in}^{n}}$ the response of participant ***n*** to a negatively worded item ***i***.

The results of the MGCFA from student direct assessment data are shown in Annex Table G.7 to Annex Table G.9 for cohort, site and gender, respectively. The results of the MGCFA from parent indirect assessment data are recorded in Annex Table G.10 and Table G.11 for cohort and site, respectively. These tables show the fit indices from the final models with three different levels of constraints and the differences in fit indices between models. The reported results correspond to the MGCFA where ARS was included as a control variable.

The results from the five Annex tables illustrate that in increasing the model constraint levels, the fit of model did not always become less satisfactory. In some cases the configural model had better fit that the metric model. It also shows that the measurement invariance level achieved may have concluded differently depending on the model fit index used. For example, the RMSEA and SRMR differences in most scales between the scalar and metric models met the criteria. However, the difference of the indices CFI and TLI showed more scales failed to meet the criteria.

In summary, models for cohort and gender groups (student data only) typically had scalar invariance, while some scales showed metric invariance. Across sites, most of the scales showed metric invariance, even though for some scales (such as ASS or MOT) there were larger differences between the configural and metric models, suggesting a deterioration of model fit between these invariance levels. Only a few of the 17 scales reached scalar invariance as the level where both intercepts and loadings become equivalent. This means that direct comparisons of most scale means across cities tend to be problematic.

12.4.3. The IRT scaling model

One of the IRT family models, the Generalised Partial Credit Model (GPCM) (Muraki, 1992[15]) was applied to derive each of SSES student direct and parent indirect assessment scales. The GPCM has been utilised in large-scale assessment programs, including the *Programme for International Student Assessment* (PISA).

Details of the GPCM and the model assumptions are provided in the following section.

The Generalised Partial Credit Model

The GPCM is a probabilistic model that is designed to work with items that are partial credit (ordered polytomous items). Items of the SSES assessments are all Likert-style items with five categories, from 'strongly disagree' to 'strongly agree'. Each item is polytomously scored. For item i with m_i categories, the probability of an individual responding in a certain response category on a particular item modelled by the GPCM is defined as:

$$P(x_i|\theta,\beta_i,\alpha_i,\tau_i) = \frac{exp\sum_{j=0}^{x}\alpha_i(\theta_n-\delta_i+\tau_{ij})}{\sum_{h=0}^{m_i}exp\sum_{j=0}^{h}\alpha_i(\theta_n-\delta_i+\tau_{ij})} \quad x_i = 0,1,\dots,m_i \qquad \text{Eq. 12:5}$$

Here, $P(x_i|\theta,\beta_i,\alpha_i,\tau_i)$ denotes the probability of individual n score x_i on item i, θ_n is the estimated latent trait for respondent n, α_i is a discrimination parameter for item i, δ_i is an item parameter which indicates the location of the item i on the latent continuum and τ_{ij} is a parameter indicating step j for item i.

The ACER *ConQuest 4* (Adams, Wu and Wilson, 2015[16]) software provides tools for estimating a variety of item response models and regression models. It was used for calibrating SSES student direct and parent indirect assessment scales and for generating WLEs (Warm, 1989[17]). The marginal maximum likelihood (MML) estimation method was used for item parameter calibrations. Students were weighted with the senate weight, which was computed based on the final student weight (WT2019) in a way that the nine cities meeting sample participation requirements were equally represented in the calibrations. Similarly, parent indirect assessment scales were calibrated with equal representation from the five cities meeting sample participation requirements.

All GPCM models were estimated using scored data (see Table 12.3). All missing responses were treated as missing for calibration and scaling as there are no 'correct' or 'incorrect' answers for this type of rating-scale items that are designed to measure the degree of agreement or disagreement with statements.

Student scales were first calibrated separately by cohort and by Site, and the corresponding item and slope parameters were reviewed with regard to their equivalence between cohorts and across cities. Furthermore, the results were used to identify problematic items together with the results from the CFA models in the adjudication of final item sets for measurement.

Table 12.12 summarises the item deletions in student direct and parent indirect assessments for the 15 original predefined scales with eight items per measured construct. Decisions on removing items were based on reviews of several aspects of scaling analysis outcomes as well as the content of items. Aspects included in this review related to improvements of model fit for CFA, the size of item factor loadings, model modification indices, overall invariance levels across cohort, cities and gender groups, and results from initial calibrations with GPCM, such as a review of item slope (tau) parameters.

Decisions were made in consideration of whether removed items improved the model fit without reducing scale reliability significantly; keeping negatively worded item(s) within a scale in cases where satisfactory model fit was ensured; and removing a minimum number of items within each scale. Item deletions in the parent indirect assessment were based on the results of scaling analyses of parent data where the item adjudication differed from one

of the student direct assessment scales, even though for many scales the same items were used for the final scaling.

Table 12.12. Item deletion in direct student and indirect parent assessment

Scale		Student			Parent	
	Number of Item Kept	Deleted item	Reason	Number of Item Kept	Deleted item	Reason
ASS	7	ASS07	high tau and duplication with 06	7	ASS07	high tau and duplication with 06
COO	7	COO05	duplication with 08	7	COO05	duplication with 08
CRE	6	CRE03	low factor loading, duplication with 03	6	CRE03	low factor loading, duplication with 03
		CRE08	duplication with 03		CRE08	duplication with 03
CUR	6	CUR01	residual covariation	8	Nil	
		CUR03	residual covariation			
EMO	7	EMO05	duplication with 08	6	EMO01	extremely low factor loading
					EMO06	improve model fit
EMP	6	EMP05	duplication with 06	6	EMP02	Improve model fit
		EMP08	low factor loading		EMP08	low factor loading
ENE	7	ENE07	duplication with 06	7	ENE04	duplication with 03
OPT	7	OPT04	residual covariation with 02 and 06	8	Nil	
PER	7	PER06	residual covariation with 05	8	Nil	
RES	6	RES02	low factor loading	6	RES02	low factor loading
		RES07	residual covariation with 02 and 06		RES07	residual covariation with 02 and 06
SEL	6	SEL05	low factor loading, residual covariation with 08	6	SEL05	low factor loading, residual covariation with 08
		SEL06	low factor loading, residual covariation with 04		SEL06	low factor loading, residual covariation with 04
SOC	6	SOC06	residual covariation with 03	6	SOC06	residual covariation with 03
		SOC08	residual covariation with 07		SOC08	residual covariation with 07
STR	6	STR06	residual covariation with 02	7	STR07	improve model fit
		STR08	residual covariation with 04			
TOL	7	TOL08	high tau in older cohort	7	TOL03	improve model fit
TRU	6	TRU03	residual covariation with 02	7	TRU03	improve model fit
		TRU04	residual covariation with 07			

The GPCM item parameters for student direct and parent indirect assessment data for each scale are shown in Annex Table G.12 and Table G.13, respectively. The results reported in these tables were based on the combined datasets from the two cohorts with conditioning on cohort using the grouping variable. The following item parameters are listed in the tables: item slope (discrimination parameter $\boldsymbol{\alpha}_i$ in Eq. 12:5) and item location (δ_i) and item step parameters (τ_{ij}). The last step parameter of each item was not listed as the sum of the step parameters equals zero.

12.4.4. Deriving SSES scale scores

Once the final items sets for each scale had been determined, the final item parameters (location and slope) were estimated for each scale using the combined SSES datasets with equally weighted Site data. WLEs were then estimated by anchoring both item location and slope parameters for all respondents, including those from cities that had not been included in the calibration. The WLEs, originally on a logit metric, were transformed (standardised) to a reporting metric where the scale averages of 500 for equally weighted data from cities

meeting sample participation requirements reflected the results for students from the younger cohort who had chosen average mid-points across items in each scale. Similarly, the standard deviation for the combined dataset with equally weighted Site data was set to 100.

The standardisation procedure was carried out in two steps. In the first step, a linear transformation was applied using the means and standard deviations for each scale among the students from the younger cohort. The means and standard deviations were computed using the calibration sample with so-called senate weights. The linear transformation was achieved as follows:

$$SS_n = 500 + 100\frac{(WLE_n - \overline{WLE})}{\sigma_{WLE}} \qquad \text{Eq. 12:6}$$

Here, $\boldsymbol{SS_n}$ was the initial standardised scale score of individual $\boldsymbol{n}$, $\boldsymbol{WLE_n}$ the estimated WLE in logit matrix for individual $\boldsymbol{n}$, $\boldsymbol{\overline{WLE}}$ the mean WLEs of the younger cohort, and $\boldsymbol{\sigma_{WLE}}$ the standard deviation of WLEs of the younger cohort.

In the second step, the average scores for respondents who had on average chosen mid-points across items in a scale were calculated and given as $\boldsymbol{SS_M}$. Then, for each scale, $\boldsymbol{SS_M}$ was subtracted from 500 and the differences were subsequently added to the initial scale scores $\boldsymbol{SS_n}$ to obtain the final scale scores: ScaleScore_n (Eq. 12:7). Figure 12.4 illustrates the relationship between the value of 500 on the reporting scale and the midpoint of responses across an example item set. It should be noted that there was typically variation in responses across items and this refers to the average midpoint of respondents.

Figure 12.4. Illustration of relationship between average midpoint of items and metric

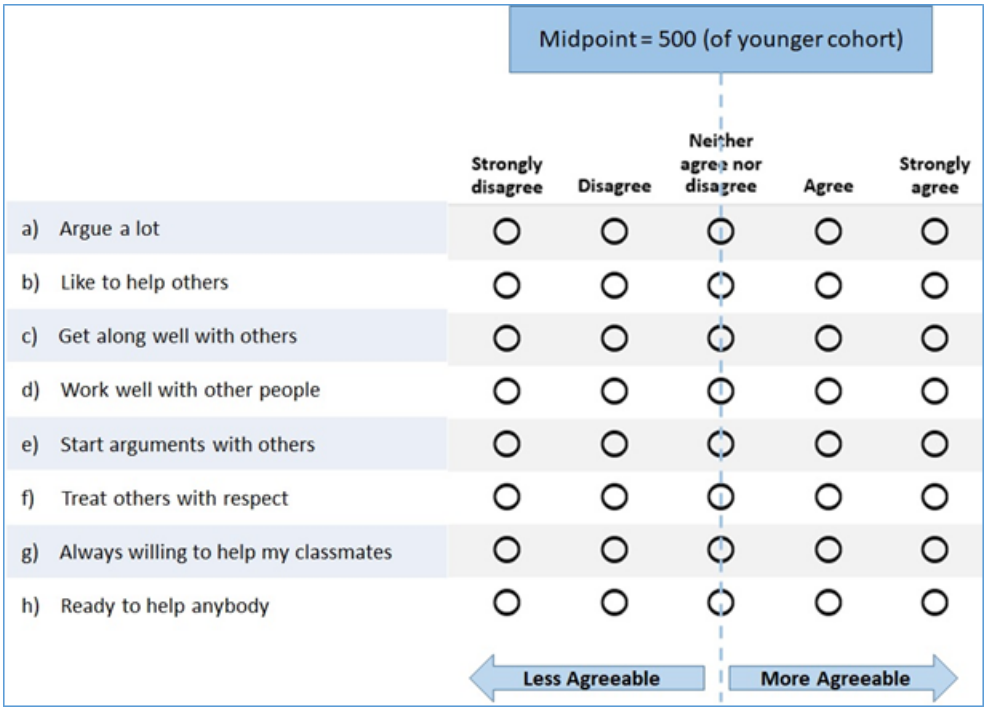


Table 12.13 lists the parameters used for transforming individual WLEs (in logits) to the reporting scale metric for each scale for the student direct and parent indirect assessments.

$$ScaleScore_n = SS_n + (500 - SS_m)$$ Eq. 12:7

Table 12.13. Scale score transforming parameters

Scale	Student			Parent		
	WLE	σ_{WLE}	SS_M	WLE	σ_{WLE}	SS_M
ASS	-0.0008	1.1594	499.7	0.0179	1.1523	453.0
COO	0.0239	1.2575	371.4	-0.0198	1.1058	343.0
CRE	0.0366	1.3108	385.5	0.0134	1.1799	377.1
CUR	0.0118	1.2143	372.2	0.0136	1.1653	353.0
EMO	0.0366	1.3003	443.6	0.0065	1.1496	454.9
EMP	0.0398	1.3592	391.3	0.0088	1.2124	386.1
ENE	0.0503	1.3284	411.1	0.0759	1.4414	401.3
OPT	0.0297	1.2427	399.3	0.0638	1.3354	382.9
PER	0.0270	1.2151	391.1	0.0287	1.1449	434.6
RES	0.0530	1.3789	413.2	0.0181	1.1570	433.3
SEL	0.0348	1.3016	410.1	0.0113	1.1487	434.9
SOC	0.0370	1.3228	392.7	0.0159	1.2044	376.4
STR	0.0048	1.1933	480.7	0.0147	1.1838	462.1
TOL	0.0398	1.2883	403.0	0.0673	1.3401	394.0
TRU	0.0288	1.2717	430.6	0.0155	1.1695	414.6
EFF	0.0457	1.3895	390.7	0.0239	1.2321	373.5
MOT	0.0327	1.3164	388.0	0.0071	1.1121	399.6

12.4.5. Estimated adjusted scale scores

As described previously, for the review of measurement invariance final MGCFA were estimated controlling for acquiescence response sets (ARS). As using IRT models does not include corrections for ARS (Primi et al., 2019[18]), WLE scale scores were adjusted for ARS by relying on a linear regression and implemented in two steps as shown below:

$$ScaleScore = a + b * ARS + \varepsilon$$ Eq. 12:8

$$AdjScaleScore = ScaleScore - (a + b * ARS)$$ Eq. 12:9

In the first step, the calibration sample parameters ***a*** and ***b*** were determined from a linear regression (Eq. 12:8) where scale scores were regressed on the ARS indicator variable. In the second step, adjusted scale score were derived as residuals for such a regression for all participants using Eq. 12:9. The adjusted scale scores were then further adjusted so that 500 corresponded again to the average item score midpoint within each scale for respondents from the younger cohort in the same way as described above.

Table 12.14. Adjusted scale score transforming parameters

Scale	Student			Parent		
	a	*b*	SS_M	*a*	*b*	SS_M
ASS	-2.5078	0.8125	500.2	-2.8804	0.8918	454.1
COO	-3.9081	1.1949	374.3	-3.4502	1.0558	343.1
CRE	-3.8222	1.1639	390.4	-3.8512	1.1825	379.3
CUR	-3.1082	0.9265	375.0	-3.6465	1.0989	353.6
EMO	1.9701	-0.6489	445.0	2.9618	-0.9271	456.9
EMP	-4.0877	1.2972	395.9	-3.8204	1.2012	386.9
ENE	-1.0519	0.2465	410.3	-2.2225	0.6322	401.1
OPT	-2.2187	0.6225	399.6	-3.1317	0.9585	377.9
PER	-1.6774	0.4690	390.3	-0.8219	0.2872	434.1
RES	2.5681	-0.8445	415.7	2.4926	-0.7687	434.8
SEL	-2.0638	0.6280	409.9	-1.8972	0.6389	434.9
SOC	-3.0684	0.9002	393.5	-3.4439	1.0343	378.7
STR	5.0285	-1.6506	481.7	6.0619	-1.8912	460.5
TOL	-3.3082	1.0750	405.6	-3.1636	1.0008	395.1
TRU	-2.2369	0.6201	430.2	-1.6772	0.4836	413.7
EFF	-5.4385	1.6975	397.1	-4.9527	1.5651	378.8
MOT	-4.5006	1.3728	393.1	-3.7905	1.2064	404.3

12.4.6. Anchoring vignettes

Cross-cultural comparability is an important methodological aspect of SSES. Reference bias represents a potential source of cross-cultural incomparability for self-report measures (Kankaraš, 2017[19]). It refers to a situation in which people from different countries answer the same question using different reference standards. In particular, a question such as: "I see myself as someone who tends to be lazy" may be answered differently depending on a person's standards or reference points regarding what it means to be lazy. Therefore, it is possible that national rankings of responses to this question do not correlate with factual measures such as average working hours (Schmitt et al., 2007[20]). Reference bias can be a problem when comparing aggregate data between cultures, but not when comparing individual scores within the same culture (Kyllonen and Bertling, 2013[21]).

One way to try to reduce potential reference bias is by using anchoring vignettes. Anchoring vignettes are designed to identify the reference system used by respondents for evaluating behaviours presented on a given scale. Based on the answers obtained from the anchoring vignettes, respondents' answers to the social and emotional skills are adjusted to account for differences in their reference systems. This adjustment could reduce possible bias introduced by respondents from different cultures using different reference systems for evaluating the same behaviours.

The initial form of the anchoring vignettes was developed based on a comprehensive review of previous forms of anchoring vignettes used in large-scale international studies (Mõttus et al., 2012[22]; Primi et al., 2016[23]; He and Van de Vijver, 2016[24]; OECD, 2014[25]) and in consultation with the SSES Technical Advisory Group. The main construction principle in terms of their content, i.e. situations and behaviours that they depict, was that behavioural examples should largely represent all subdomains/individual skills of the Big Five that are included in the study. Furthermore, an attempt is made to find a balance between concreteness and generalisability (both culturally and socially), referring to a more general aspect of the school environment that should be familiar and relatable to

all school children. The vignettes were first tested during the item trials. In both item trials and field test, for each Big Five domain, a set of three anchoring vignettes were developed, depicting behavioural characteristic for the high, medium and low end of that dimension. Thus, a total set of 15 anchoring vignettes were tested. In the items trials and the field test, all 15 vignettes were administered to all three groups of respondents (students, parents and teachers). However, in order to reduce the response burden, only one of the five randomly chosen sets was administered to respondents in the main study. Table 12.15 below provides examples of the anchoring vignettes for the five domains of the Big Five.

Table 12.15. Examples of anchoring vignettes

		How much do you agree that each of the following students handles his/her emotions well?	**Strongly disagree**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
Emotional regulation	**+**	[John] never looks stressed. He is calm and positive even before an exam. How much do you agree that [John] handles his emotions well?	1	2	3	4	5
	0	[Mary] is usually calm during classes, but there are a few moments when she can get upset and change moods. How much do you agree that [Mary] handles her emotions well?	1	2	3	4	5
	-	[Anna] is frequently in a bad mood and gets upset every time someone does something she doesn't like. How much do you agree that [Anna] handles her emotions well?	1	2	3	4	5

		How much do you agree that each of the following students is a hard-working person?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Conscientiousness	+	[Mia] studies hard and gets very good grades. She always does her homework, finishes her assignments, and is always prepared before starting the class. How much do you agree that [Mia] is a hard-working person?	1	2	3	4	5
	0	[Peter] usually gets good grades. He sometimes has trouble paying attention in class, but usually completes his homework on time. How much do you agree that [Peter] is a hard-working person?	1	2	3	4	5
	-	[Tom] often forgets to do his homework, is rarely prepared before the class starts, and does not care about his grades. How much do you agree that [Tom] is a hard-working person?	1	2	3	4	5

		How much do you agree that each of the following students is a kind person?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Agreeableness	+	[Marc] is nice to almost everyone. He is always willing to help or lend things to his classmates. How much do you agree that [Marc] is a kind person?	1	2	3	4	5
	0	[Wesly] is helpful to most of his friends but there are a few of his classmates that find him unfriendly. How much do you agree that [Wesly] is a kind person?	1	2	3	4	5
	-	[Maria] often argues with her classmates and can be quite rude. She never lends her things to her classmates. How much do you agree that [Maria] is a kind person?	1	2	3	4	5

		How much do you agree that each of the following students is a social and outgoing person?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Extraversion	+	[Jennifer] has many friends and likes to talk with her classmates. She is very active and leads many school activities. How much do you agree that [Jennifer] is a social and outgoing person?	1	2	3	4	5
	0	[Alexandra] likes to be with her friends but there are days when she is quiet and does not talk with other classmates. How much do you agree that [Alexandra] is a social and outgoing person?	1	2	3	4	5
	-	[Jack] is shy and talks very little with his classmates. He does not like to talk in front of other people and prefers to be left alone. How much do you agree that [Jack] is a social and outgoing person?	1	2	3	4	5

		How much do you agree that each of the following students enjoys learning new things?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Openness to Experience	+	[Mario] is very interested in knowing about the world. He is curious about different places and people. He also loves reading about different things. How much do you agree that [Mario] enjoys learning new things?	1	2	3	4	5
	0	[Christine] is sometimes interested in learning about other places and people but she rarely reads about different things. How much do you agree that [Christine] enjoys learning new things?	1	2	3	4	5
	-	[Timothy] is not interested in learning about other places and people. He does not like to explore new places and does not like reading about different things. How much do you agree that [Timothy] enjoys learning new things?	1	2	3	4	5

Anchoring vignettes were analysed seperately for each of the five Big Five domains. Variance decomposition showed that although significant, only a small portion of the total variance can be attributed to the different sites. Furthermore, the partial credit model (PCM) and modified PCM are used to obtain individual skill estimates. These estimates use the TAM package in R. Only marginal differences were found between social and emotional skill scales with and without a correction for the anchoring vignettes. Correlations are generally above 0.90. Therefore, it was decided to not correct for the anchoring vignettes in the final scaling of the social and emotional skills.

Nevertheless, a specific aspect of response styles, acquiescence, is taken into account when scaling the social and emotional skills as explained in section 12.4. Also, even if not used to adjust scores, anchoring vignettes could still help students understand how to respond to this type of questions. In conclusion, although SSES was designed to mitigate reference-group bias through anchoring vignettes, part of this effect may still remain.

References

Adams, R., M. Wu and M. Wilson (2015), *ACER ConQuest: Generalised Item Response Modelling Software*, Australian Council for Educational Research. [16]

Bollen, K. and J. Long (eds.) (1993), *Testing structural equation models.*, Sage Publications, Inc. [11]

Chen, F. (2007), "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 14/3, pp. 464-504, http://dx.doi.org/10.1080/10705510701301834. [12]

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://dx.doi.org/10.1787/db1d8e59-en. [1]

Cronbach, L. (1951), "Coefficient alpha and the internal structure of tests", *Psychometrika*, Vol. 16/3, pp. 297-334, http://dx.doi.org/10.1007/bf02310555. [3]

He, J. and F. Van de Vijver (2016), "Corrigiendo las diferencias de uso de escala entre países de América Latina, Portugal y España en PISA", *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, Vol. 22/1, http://dx.doi.org/10.7203/relieve.22.1.8282. [24]

John, O., L. Naumann and C. Soto (2008), *Paradigm shift to the integrative Big Five taxonomy*. [2]

Kankaraš, M. (2017), "Personality matters: Relevance and assessment of personality characteristics"*, OECD Education Working Papers*, No. 157, OECD Publishing, Paris, https://dx.doi.org/10.1787/8a294376-en. [19]

Kaplan, D. (2009), *Advanced Quantitative Techniques in the Social Sciences: Structural equation modeling: Foundations and extensions*, SAGE Publications, Inc., http://dx.doi.org/10.4135/9781452226576. [7]

Kyllonen, P. and J. Bertling (2013), *Innovative Questionnaire Assessment Methods to Increase Cross-Country Comparability*. [21]

McDonald, R. (2013), *Test Theory*, Psychology Press, http://dx.doi.org/10.4324/9781410601087. [4]

Mõttus, R. et al. (2012), "The Effect of Response Style on Self-Reported Conscientiousness Across 20 Countries", *Personality and Social Psychology Bulletin*, Vol. 38/11, pp. 1423-1436, http://dx.doi.org/10.1177/0146167212451275. [22]

Muraki, E. (1992), "A generalized partial credit model: Application of an EM Algorithm", *ETS Research Report Series*, pp. i-30, https://doi.org/10.1002/j.2333-8504.1992.tb01436.x. [15]

Muthén, L. and B. Muthén (2012), *Mplus User's Guide: Statistical Analysis with Latent Variables*, Muthén & Muthén. [8]

Nunnally, J. and I. Bernstein (1994), *Psychometric Theory*, McGraw-Hill. [6]

OECD (2019), *TALIS 2018 Technical Report*, OECD Publishing. [13]

OECD (2014), *PISA 2012 Technical Report*, OECD Publishing. [25]

Primi, R. et al. (2020), "Classical Perspectives of Controlling Acquiescence with Balanced Scales", in *Springer Proceedings in Mathematics & Statistics, Quantitative Psychology*, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-43469-4_25. [14]

Primi, R. et al. (2019), "Controlling Acquiescence Bias with Multidimensional IRT Modeling", in *Springer Proceedings in Mathematics & Statistics, Quantitative Psychology*, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-01310-3_4. [18]

Primi, R. et al. (2016), "Anchoring Vignettes", *European Journal of Psychological Assessment*, Vol. 32/1, pp. 39-51, http://dx.doi.org/10.1027/1015-5759/a000336. [23]

Rhemtulla, M., P. Brosseau-Liard and V. Savalei (2012), "When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions.", *Psychological Methods*, Vol. 17/3, pp. 354-373, http://dx.doi.org/10.1037/a0029315. [9]

Schmitt, D. et al. (2007), "The Geographic Distribution of Big Five Personality Traits", *Journal of Cross-Cultural Psychology*, Vol. 38/2, pp. 173-212, http://dx.doi.org/10.1177/0022022106297299. [20]

Suh, Y. (2015), "The Performance of Maximum Likelihood and Weighted Least Square Mean and Variance Adjusted Estimators in Testing Differential Item Functioning With Nonnormal Trait Distributions", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 22/4, pp. 568-580, http://dx.doi.org/10.1080/10705511.2014.937669. [10]

Trizano-Hermosilla, I. and J. Alvarado (2016), "Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements", *Frontiers in Psychology*, Vol. 7, http://dx.doi.org/10.3389/fpsyg.2016.00769. [5]

Warm, T. (1989), "Weighted likelihood estimation of ability in item response theory", *Psychometrika*, Vol. 54/3, pp. 427-450, http://dx.doi.org/10.1007/bf02294627. [17]

Notes

[1] Only the final scale results are reported.

[2] The CFA model estimation went through multiple iterative reviews to ensure the scale had a satisfactory model fit.

[3] During the FT data analysis, the data collected across 19 scales were fitted to exploratory factor analysis as well as CFA.

[4] Chi-square (χ^2) test values are also often used for evaluating model fit but the chi-square statistic is overly sensitive to sample size and violations of the assumptions about normal distributions. Therefore, in view of the large sample sizes for SESS data, the chi-square statistic was not considered when reviewing model fit.

[5] The scale STR (stress resistance/resilience) contains 7 negatively worded items and one positively worded item in the original 8 items. Hence the negative items in STR were considered as positive items as in other scales, and the positive item was consider as a negative item for STR with regard to this comment.

Chapter 13. Data adjudication

The Technical Standards for the OECD Survey on Social and Emotional Skills 2019 (SSES) (see Annex A) were developed and revised during the early phases of the project development. The Technical Standards stipulate how SSES (the Study) must be implemented in each participating city (Site). The International Contractor monitored how the Study was implemented in each Site and adjudicated on their respective adherence to each aspect of the Technical Standards.

This chapter describes the processes and procedures used to adjudicate the implementation of the Study in each Site and determine the outcomes of data adjudication, mainly based on the following aspects:

the extent to which each Site met SSES sampling standards

the outcomes of the adaptation, translation and verification process

the outcomes of the quality monitoring visit

the quality and completeness of the submitted data.

13.1. SSES 2019 Technical Standards

The Technical Standards cover the following aspects of the Study.

13.1.1. Data quality standards

- Target population and sampling
- Language of the survey
- Field Test (FT) participation
- Adaptation of skill scale items, questionnaires and manuals
- Translation of skill scale items, contextual questionnaires and manuals
- Survey administration
- Security of the material
- Quality monitoring
- Occupational data response coding
- Data submission
- Survey delivery mode
- Security of personal data.

13.1.2. Management standards

- Communication with the International Contractor
- Schedule for submission of materials
- Drawing of samples
- Archiving of materials.

13.2. Quality assurance - implementing the Technical Standards

The appointed Site Project Managers (SPMs) were responsible for implementing the Study in their respective cities, in accordance with the Technical Standards and the advice of the International Contractor, as detailed in the various SSES operational manuals and guidelines.

Throughout the entire SSES project cycle, ongoing quality assurance activities were built into each task delivery plan and implemented consistently by the International Contractor. First, a quality control mechanism was established for key aspects and tasks of the Study using the operational manuals. Second, quality monitoring activities were carried out through the agreement processes for Site submissions on various aspects of the project deliveries, regular communications and Quality Monitor (QM) school visits. Those agreement processes ensured that the Study was implemented at each Site in accordance with the Technical Standards, and that the International Contractor could provide advice and take early rectifying action to avoid critical errors occurring. Further, the QM visits provided first-hand and systematically-collected evidence of how the assessments were being administered across the participating cities. The quality monitoring data was collected systematically, which reflects the monitored implementation of the survey in relation to the Technical Standards.

Information collected through the quality monitoring activities was used during the data adjudication process to determine the level of compliance with the Technical Standards.

13.3. Collection of quality assurance data

The collection of quality assurance data was embedded in the Study implementation plan, and factored into all activities and stages of the SSES project cycle. The quality assurance data were collected against project milestones and key tasks, including:

- monitoring each participating Site's adherence to the deadlines
- summary of each participating Site's sampling framework and plan
- information from the language adaptation and verification teams
- data from the Quality Monitor reports
- information gathered from online or in-person meetings with the SPMs.

This information formed the basis of the data adjudication database, which provided comprehensive records of:

- indicators of any non-compliance with the Technical Standards and associated agreements or arrangements made at that time
- the point at which a problem that occurred could be easily identified.

The quality monitoring data was collected from these main administrative areas:

- *project administration and management*: information relating to adherence to the timeline, agreement of adaptation spreadsheets, submission of information
- *data analysis:* information from item level reports
- *school-level materials:* information from the agreement of adaptations made to the test administration scripts
- *Main Study (MS) feedback:* information provided by SPMs in the MS feedback online questionnaire
- *Quality Monitors*: coordination activities including QM recruitment
- *Quality Monitors' country reports*: information gathered via the data collection forms and through their interactions with School Coordinators and Study Administrators
- *sampling data:* information such as school and student response rates, exclusion rates and eligibility problems
- *translation*: information relating to the verification and translation process
- *data management*: issues identified during the data cleaning checks and from data cleaners' reports, as well as issues related to the eligibility of students assessed.

13.4. Data adjudication

Data adjudication is a process to critically review and examine the dataset from each Site in accordance with the Technical Standards. Then, with reference to the quality assurance data collected, make a range of possible technical recommendations for the inclusion, exclusion, or treatment of the data.

A data adjudication summary report was compiled for all participating cities by the experts in each respective area. The summary report contained the quality assurance data for each key area of SSES implementation, as well as the detailed Site participation summary. Any key areas of concern identified were documented in alignment with the related Technical Standard for further consideration, and the relevant SPM was contacted for clarification where required.

The adjudication group, formed by representatives of the OECD and of the International Contractor, reviewed the reports and collaborated to reach adjudicated decisions on the quality of the data from each Site.

13.5. Adjudication outcomes

13.5.1. Overview of response rate issues

The SSES school and student response rate requirements are discussed in Chapter 4. One participating Site – Sintra (Portugal) – failed to meet both the school response rate and student response rate (see Table 11.10 and Table 11.11), therefore their dataset for both age cohorts was excluded from the student assessment item scaling.

The older cohorts in Daegu (Korea), Helsinki (Finland) and Ottawa (Canada) fell short of the student response rate standard; however, considering the fact that the achieved response rates were relatively close to the acceptable rate of 80% it was determined, with OECD's agreement, that the data were acceptable (see Table 11.11).

Half of the participating sites – Daegu (Korea), Helsinki (Finland), Houston (United States), Ottawa (Canada) and Sintra (Portugal) – had extremely low response rates for the parent questionnaire in both age cohorts (see Figure 12. 14). It was determined, with OECD's agreement, that the scaling of the parent data would be performed using data from cities with response rates greater than 50% only. However, parent datasets from all participating cities have been kept in the final dataset.

Daegu (Korea), Houston (United States), Sintra (Portugal) and Ottawa (Canada) had unweighted teacher response rates for both cohorts below the standards of 80% (see Table 11.12). Scaling indicators were created using raw scores.

13.5.2. Deviations from standard procedures

The Technical Standards established a best practice benchmark for the Study implementation, and explicitly stipulated expectations for the data collection process and data quality. However, it was anticipated that not all the standards and requirements could be fully achieved in all local contexts and circumstances. It was recognised that participation in the Study could be a quite complex and diverse undertaking, and many factors might influence the process and outcomes. For instance, it may not be possible for sampled small schools to participate in the Study due to national privacy legislation; or it may be that the initially randomly selected school list for quality monitoring visits had to be changed due to scheduling constraints that may be caused either by the school or the unavailability of the QM. Typically, these issues were raised with the International Contractor in advance of the assessment, and alternative approaches were considered jointly with the SPM. Where deviations from the standard procedures were easily determined by the International Contractor and there was minimal risk to the quality of the data collection, the deviations are not reported in the Site summaries below. Sites who fully met the Technical Standards are not listed below.

During the MS implementation, due to the limited size of the target population for the Study as identified in the sampling frame, Manizales (Colombia), Sintra (Portugal), Helsinki

(Finland) (older cohort only), Houston (United States) (older cohort only) and Ottawa (Canada) (older cohort only) took the census approach in their sampling process.

Daegu (Korea)

Daegu (Korea) had a weighted student response of 77.9% for the younger cohort, slightly below the 80% standard requirement. There were fewer than 3000 students in the younger cohort as specified in the Technical Standards (2895). Data were included in the final database.

Helsinki (Finland)

Helsinki (Finland) had a weighted student response rate of 75.7% for the older cohort, slightly below the 80% standard requirement.

There was a total of 8.02% of student exclusions for the younger cohort and 9.04% of student exclusions for the older cohort in Helsinki (Finland), mainly due to a much higher than expected within-school Special Education Needs (SEN) exclusion rates of 7.89% for the younger cohort and 8.82% for the older cohort. This was due to the presence of a significant number of students with language background other than the official language, which was considered by the Study Administrator to be a significant obstacle to completing the assessment.

There were fewer than 3000 students in the older cohort as specified in the Technical Standards (2338). Data were included in the final database.

Houston (United States)

There was a total of 6.2% of student exclusions for the older cohort in Houston (United States), due to a high level of overall school exclusions, equivalent to 5.12% of the student population before sampling. Citing concerns about legal repercussions of potential identification of single student responses in schools with very low student enrolments, the Site requested schools with less than 20 eligible students not be allowed to participate in the Study. As a result, the number of eligible schools was significantly reduced. This required a census be drawn of all remaining schools with an increase in the cluster size or number of students sampled in each school, to yield the required number of student responses. The final sample was drawn with OECD's agreement. Data were included in the final database.

Istanbul (Turkey)

There were fewer than 3000 students in the younger cohort as specified in the Technical Standards (2693).

After the MS school sample had been drawn, the Site requested to draw an additional school sample due to an unexpected shortfall in the number of participating schools and hence a low student yield. With OECD's agreement, a second set of schools was sampled for this cohort with minimum overlap control on the original drawn school sample. Data were included in the final database.

Ottawa (Canada)

Ottawa (Canada) had a weighted student response rate of 70.7% for the older cohort, below the 80% standard requirement.

There were fewer than 3000 students in the older cohort as specified in the Technical Standards (2176). Data were included in the final database.

Sintra (Portugal)

In Sintra (Portugal), the weighted school response rate for the younger cohort (82.5%) fell slightly short of the 85% standard requirement, and for the older cohort (74.5%) also fell below the minimal threshold.

Sintra (Portugal) had a weighted student response rate of 58.2% for the younger cohort and 48.8% for the older cohort, both significantly below the 80% standard requirement.

There were fewer than 3000 students as specified in the Technical Standards (1952 for the younger cohort and 1383 for the older cohort). Data were included in the final database, but not used for student assessment item scaling. The student scores for Sintra (Portugal) were based on the international scale which was calibrated without Sintra's data. Results have been reported separately from other cities.

Chapter 14. Procedures and Construct Validation of Contextual Questionnaire Data

This chapter outlines the procedures for scaling items from the four OECD Survey on Social and Emotional Skills 2019 (SSES) contextual questionnaires (student, parent, teacher and principal). There were three different types of indices: simple indices, item response theory (IRT) scale indices and composite index.

The simple indices derived from the SSES student, parent, principal and teacher questionnaires are presented in detail in the first section. The methodology used for the validation and construction of each IRT scale index is provided in the next section. The methodology used for the construction of the composite index, SES, is provided in detail in the final section.

14.1. Introduction

There were three different types of indices used for scaling items from the four SSES contextual questionnaires (student, parent, teacher and principal):

- *Simple indices:* constructed using an arithmetic transformation or recoding of one or more variables. This can take the form of ratios, averages or binary indicators.
- *Item Response Theory (IRT) scale indices:* derived by combining a group of items which are intended to measure an underlying latent construct.
- *Composite index:* index of socio-economic status (SES) based on data from parental education, parental occupation and home possessions, derived as factor scores from principal component analyses.

14.2. Computation of simple indices

14.2.1. Student Questionnaire

Student age

Student age (AGE_STD) was calculated as the age in months at the time of the questionnaire administration. It is the difference between the date the Student Questionnaire was administered and the student's date of birth. Information from the Student Tracking Forms was used to derive age, except for students where this information was missing. In these cases, information from the student questionnaire (Question 3) provided data for calculating this index. The formula for computing AGE_STD was:

$$AGE\_STD = ((DOQ - DOB)/365.24)*12$$

where DOQ and DOB are, respectively, the date of the questionnaire administration and the student's date of birth. The result was rounded to two decimal places.

14.3. Scaling methodology and construct validation

14.3.1. Preliminary item calibration

Preliminary item calibrations were estimated using rescaled weights to ensure that each site was equally represented in the international sample (senate weights). For each questionnaire, only sites that met minimum participation requirements were included in the calibration of item parameters. For ease of interpretation, all negatively worded items were reverse-coded, so that the highest value represented a higher attribute (e.g. a larger quantity). The International Contractor computed scale scores for all respondents who had answered at least two of the items measuring the respective construct. All missing responses were omitted in both the item calibration stage and in the estimation of scale scores.

14.4. Construct validation

One of the main goals of this study is to develop comparable measures of students', parents' and teachers' background and perceptions. The cross-country validity of these constructs requires closely monitoring the rigorous translation process into different languages (described in Chapter 6). It makes assumptions about measuring similar characteristics and perceptions in different national and cultural contexts. Psychometric techniques can be used

to analyse the extent to which constructs have consistent construct validity across participating countries.

Cross-national validity of these constructs is of particular importance as measures derived from questionnaires are often used to predict differences in student development within and across countries. There are different methodological approaches for validating questionnaire constructs, each with advantages and limitations.

For the analyses and scaling, the International Contractor applied three different approaches:

- a review of internal scale consistencies through the computation of Cronbach's alpha coefficients and reviewing item-total correlations
- a factorial analysis of item dimensionality measurement invariance
- a scaling of items using the IRT Rasch Partial Credit Model (PCM).

Cronbach's alpha was used to measure scale reliability. The reliability or internal consistency compares each scale within each site and compares it between the sites. The alpha coefficient values range between 0 and 1, with higher values indicating higher reliability of the scale. Commonly accepted cut-off values are 0.9 to signify excellent, 0.8 for good and 0.7 for acceptable internal consistency.

The questionnaires were designed to follow the framework, with many items intended to measure an underlying latent construct. In order to derive meaningful indices, a number of procedures were undertaken to validate each scale. For ease of interpretation of the indices, all negatively worded items were reverse coded, so the highest value for each item represents a higher attribute. Items from all potential scales were initially evaluated through an exploratory factor analysis (EFA). The root mean square error of approximation (RMSEA) was used to evaluate the overall fit of the scale. A RMSEA value below 0.1 indicates the scale has an acceptable model fit. Factor loadings for individual items are assessed with the expectation that items belonging to the same scale would load strongly together. Items with poor item statistics were discarded and scales with poor fit were peer reviewed. A CFA was then carried out on the scales, with only acceptable items from the EFA.

Structural equation modelling (SEM) allows the confirmation of theoretically expected dimensions which could have been re-specified at the Field Test stage. When using CFA, researchers acknowledge the need to employ a theoretical model of item dimensionality that can be tested via the collected data. Within the SEM framework, latent variables are linked to observed variables via measurement equations. An observed variable $\boldsymbol{x}$ is thus modelled as:

$$x = \Lambda_x \xi + \delta$$

where $\boldsymbol{\Lambda_x}$ is a q x k matrix of factor loadings, $\boldsymbol{\xi}$ denotes the latent variable(s) and δ is a q x 1 vector of unique error variables. The expected covariance matrix is fitted according to the theoretical factor structure.

When conducting the CFA for SSES questionnaire data, selected model-fit indices provided measures of the extent to which a particular model with an assumed a-priori structure 'fitted the data'. For SSES, the assessment of model fit was primarily conducted through evaluating the RMSEA, the comparative fit index (CFI) and the Tucker Lewis fit index (TLI), all of which are less affected than other indices by sample size and model complexity (Bollen & Long, 1993).

Typically, RMSEA values over 0.10 suggest an unacceptable model fit, those between 0.08 and 0.1 a marginally satisfactory model fit, while values of 0.05 and lower indicate a close model fit (MacCallum, Browne, & Sugawara, 1996). As additional fit indices, CFI and TLI are bound between 0 and 1. Values below 0.90 indicate a non-satisfactory model fit, whereas values greater than 0.95 suggest a close model fit [see (Bentler & Bonett, 1980) (Hu & Bentler, 1999)].

Generally, maximum likelihood estimation and covariance matrices are not appropriate for analyses of categorical questionnaire items because the approach treats items as if they are continuous. Therefore, the SSES analysis relied on robust weighted least squares estimation (WLSMV) (Muthén, du Toit, & Spisic, 1997) (Flora & Curran, 2004) to estimate the confirmatory factor models. The software package used for estimation was Mplus version 7 (Muthén & Muthén, 2012).

Validating the constructs across sites assumes that all constructs are consistent irrespective of the national and cultural contexts. As all the questionnaire items are based on the respondents' self-reports, measurement error stemming from respondent bias is expected.

A multiple-group confirmatory factor analysis (MGCFA) was used to review measurement invariance. The MGCFA model's factor loadings, intercepts, residual variances, means and standard deviations were estimated for each site separately, and constraints on these parameters were then examined simultaneously across groups. However, depending on the level of invariance being investigated, parameters can be restricted so they are equal or they can be allowed to vary across groups. The MGCFA model allows parameters to be constrained to the mean and covariances of the observed variables instead of the raw simple scores (Sörbom, 1974).

Three levels of invariance were examined for SSES – configural, metric and scalar. Configural invariance would hold if each group (e.g. Site) had the same factor structure. Metric invariance would hold if each group had the same factor structure and the same strength of association between the items and their respective factor. Scalar invariance would hold if the conditions of metric invariance are observed, and each group had equivalent intercepts.

Measurement invariance testing was carried out on the final measurement model using a MGCFA in the Mplus software. For the student questionnaire, the MGCFA was evaluated for the following groups; Site, cohort and gender. For the parent, principal and teacher questionnaires, the MGCFA was evaluated for Site and cohort groups only. In testing for measurement invariance, three different models were specified and compared (i.e. configural, metric and scalar models).

Configural invariance is the least constrained model. In this model, it is assumed that the items measuring the underlying latent construct are equivalent across all groups of reference (e.g. sites). If the magnitude of the associations between the groups are the same, then the latent construct is assumed to have the same meaning for all groups (i.e. the structure of the construct is the same).

The metric level of invariance is achieved if the structure of the construct is the same across groups (i.e. configural invariance is achieved) and the strength of the association between the construct and items (factor loadings) is the same across groups.

Scalar level invariance is achieved when metric invariance has been achieved and the intercepts/thresholds for all items across groups are equivalent. When scalar invariance is achieved, it is assumed that differences in scale means across groups are free of any cross-group bias.

In order to determine whether the models were invariant across the groups, fit statistics were examined (RMSEA, CFI and TLI) and the following criteria were used for an indicative review:

- *Configural invariance:* CFI≥0.9 or TLI≥0.9 and RMSEA≤0.1
- *Metric invariance:* if the difference in the fit indices between the metric and configural models met the following: ΔCFI<0.01 or ΔTLI<0.015 and ΔRMSEA≤0.015
- *Scalar invariance:* if the differences in the fit indices between the scalar and metric model met the following: ΔCFI<0.010 or ΔTLI<0.015 and ΔRMSEA≤0.015

In instances where there were only three items for the scale (RELPARENTS, RELTEACH and ANXTEST), the models indicated perfect fit and could not be evaluated due to the limited number of degrees of freedom. Therefore, the MGCFA was evaluated using multi-dimensional models.

14.5. Item calibration and scaling

The SSES item parameters were derived using equally weighted site datasets and restricted to data from sites that met sample participation requirements:

- *Calibration of item parameters for the student questionnaire:* This was done based on a pooled database with equally weighted national samples from nine sites that met sample participation requirements for the student questionnaire.
- *Calibration of item parameters for the parent questionnaire:* This was done based on a pooled database with equally weighted national samples from five sites that met sample participation requirements for the parent questionnaire.
- *Calibration of item parameters for principal questionnaire:* This was done based on a pooled database with equally weighted national samples from nine sites that met sample participation requirements for the principal questionnaire.
- *Calibration of item parameters for teacher questionnaire:* This was done based on a pooled database with equally weighted national samples from eight sites that met sample participation requirements for the teacher questionnaire.

The latent constructs were modelled using IRT methodology. A one-parameter Rasch model (Rasch, 1960) was fitted to dichotomous items. The probability of a respondent selecting category one, instead of category zero is modelled as:

$$P_i(\theta_n) = \frac{exp(\theta_n - \delta_i)}{1 + exp(\theta_n - \delta_i)}$$

where $\boldsymbol{P_i(\theta_n)}$ is the probability of person $\boldsymbol{n}$ to score 1 on item $\boldsymbol{i}$, $\boldsymbol{\theta_n}$ is the estimated latent trait of person $\boldsymbol{n}$ and $\boldsymbol{\delta_i}$ is the estimated location of item $\boldsymbol{i}$ on this dimension. For each item, responses are modelled as a function of the latent trait $\boldsymbol{\theta_n}$.

In the case of items with more than two categories (as, for example, with Likert-type items), this model can be generalized to the (Rasch) partial credit model (Masters & Wright, 1997), which takes the form of:

$$P_{x_i}(\theta_n) = \frac{exp\sum_{j=0}^{x}(\theta_n - \delta_i + \tau_{ij})}{\sum_{h=0}^{m_i} exp\sum_{j=0}^{h}(\theta_n - \delta_i + \tau_{ij})} x_i = 0,1,\dots,m_i$$

Here, $P_{xi}(\theta_n)$ denotes the probability of person n to score x on item i, θ_n denotes the person's ability, the item parameter δ_i gives the location of the item on the latent continuum and τ_{ij} denotes an additional step parameter.

The ACER ConQuest Version 5.9.0 software was used for the estimation of model parameters and subsequent scaling. While only data from sites that had met sample participation requirements were included in the calibration of scaling parameters, the International Contractor computed weighted likelihood estimates (WLEs) scores for each respondent in all 10 sites by anchoring the overall SSES parameters estimated in the calibration procedures.

For all scales, the original logit scores were transformed into an SSES reporting metric in reference to the younger cohort. Using the equally weighted datasets from the sites that met sample participation requirements, each scale was transformed with a mean of 50 and a standard deviation of 10 for the younger cohort. This was achieved by applying the following formula:

$$\theta'_n = 50 + 10\,\frac{\theta_n - \mu_y}{\theta_y}$$

where θ'_n are the scores in the SSES metric, θ_n are the original WLEs in logits and μ_y is the mean of the younger cohort's logit scores using equally weighted data for sites meeting sample participation requirements. θ_y is the corresponding standard deviation of the younger cohort's logit scores using equally weighted site data.

14.6. Student questionnaire

14.6.1. WHO-5 wellbeing

The WHO-5 wellbeing (*WELLBEING*) index was derived from items with five wellbeing-related statements. Students were asked to indicate how they had been feeling over the past two weeks by selecting one of five options ('at no time', 'some of the time', 'more than half of the time', 'most of the time' or 'all the time'). Higher scale scores correspond to greater perceived levels of positive student wellbeing. The fit indices and factor loadings from the CFA are shown in Table 14.1. The overall wellbeing scale had a marginally satisfactory fit (RMSEA = 0.097). The first item ('felt cheerful and in good spirits') had the highest factor loading, while the remaining four items had moderately high loadings.

On average across sites, the scale had a reliability (Cronbach's alpha) of 0.86 as shown in Table 14.11). When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models, suggesting an acceptable level of measurement invariance across cohorts and gender groups. Table 14.2 records the item parameters for the final scaling.

Table 14.1. Confirmatory factor analysis fit indices and factor loadings for WHO-5 wellbeing scale

Fit indices	Pooled sample
RMSEA	0.097
CFI	0.994
TLI	0.987

STQ scale	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
WELLBEING	Cohort	0.090	0.994	0.988	0.066	0.996	0.994	0.063	0.992	0.994
	Site	0.103	0.994	0.988	0.075	0.994	0.993	0.091	0.980	0.990
	Gender	0.099	0.993	0.986	0.070	0.995	0.993	0.056	0.994	0.996

Item		Factor loadings	Item label
WHO-5 Wellbeing Index	STQM02001	**0.87**	Felt cheerful and in good spirits
	STQM02002	**0.78**	Felt calm and relaxed
	STQM02003	**0.78**	Felt active and vigorous
	STQM02004	**0.78**	Woken up feeling fresh rested
	STQM02005	**0.75**	Daily life has been filled with things that interest me

Table 14.2. Item parameters for WHO-5 wellbeing scale

Item		Item Parameters				
		Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
WHO-5 Wellbeing Index	STQM02001	-0.29	-3.10	0.00	0.35	2.75
	STQM02002	-0.04	-2.86	-0.19	0.44	2.61
	STQM02003	-0.08	-2.42	-0.12	0.44	2.11
	STQM02004	0.48	-2.02	0.09	0.24	1.69
	STQM02005	-0.07	-2.36	-0.03	0.44	1.95

14.6.2. Global mindedness

The global mindedness scale (*GLOBALMIND*) was derived from items related to global issues reflecting perceptions of climate change, global health, international conflict, poverty and gender equality. Students were asked how informed they were about the different topics by selecting one of four options ('I have never heard of it', 'I know little about this', 'I know something about this' or 'I know a lot about this'). Students received higher scores on this scale if they indicated greater knowledge of global issues. The fit indices from the CFA are shown in Table 14.3. The overall global mindedness scale had poor model fit (RMSEA = 0.114) and all five items had only moderate factor loadings. The average reliability (Cronbach's alpha) across all sites was 0.77 (see site-level results in Table 14.11).

When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models across gender and cohort groups, suggesting an acceptable level of measurement invariance. With regard to sites, model fit was much poorer for the more constrained models and caution is

warranted with regard to the comparability of this scale across sites. The item parameters used for the final scaling are shown in Table 14.4.

Table 14.3. Confirmatory factor analysis fit indices and factor loadings for global mindedness scale

Fit indices	Pooled sample
RMSEA	0.114
CFI	0.972
TLI	0.944

STQ scale	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
GLOBALMIND	Cohort	0.119	0.969	0.937	0.099	0.970	0.957	0.076	0.970	0.974
	Site	0.120	0.976	0.953	0.134	0.949	0.941	0.162	0.857	0.913
	Gender	0.114	0.972	0.945	0.086	0.978	0.969	0.087	0.963	0.968

Item		Factor loadings	Item label
Global Mindedness	STQM02501	**0.66**	Climate change and global warming
	STQM02502	**0.73**	Global health (e.g. epidemics)
	STQM02504	**0.73**	International conflicts
	STQM02506	**0.66**	Causes of poverty
	STQM02507	**0.64**	Equality between men and women in different parts of the world

Table 14.4. Item parameters for global mindedness scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Global Mindedness	STQM02501	-0.43	-1.45	-0.31	1.76
	STQM02502	0.35	-1.53	-0.26	1.79
	STQM02504	0.32	-1.34	-0.20	1.54
	STQM02506	-0.06	-1.24	-0.23	1.47
	STQM02507	-0.17	-1.00	-0.35	1.35

14.6.3. Students' perceived relationships with parents and friends

Students were asked about their perceived relationship with their parents (*RELPARENTS*) by responding about the extent to which the following statements were true ('almost never, or never true', 'sometimes true', 'often true' or 'almost always, or always true'): 'I get upset easily with my parents', 'It is hard for me to talk with my parents' and 'I feel angry with my parents'. Students received higher scores on this scale if they indicated the statements were true more often.

Students were asked about their perceived relationship with their friends (*RELFRIENDS*) by responding about the extent to which the following statements were true ('almost never, or never true', 'sometimes true', 'often true' or 'almost always, or always true'): 'My friends understand me', 'My friends accept me as I am', 'My friends are easy to talk to' and 'My friends respect my feelings'. Students received higher scores on this scale if they indicated a higher extent to which the statements were true. The fit indices and factor loadings from a two-dimensional CFA are shown in Table 14.5. The overall model had a

good fit (RMSEA = 0.028) and all four items on the *RELFRIENDS* scale had high loadings on the scale. The three items on the *RELPARENTS* scale had moderately high loadings on the scale. The average reliabilities (Cronbach's alpha) across all sites were 0.86 for *RELFRIENDS* and 0.74 for *RELPARENTS* (see site-level results in Table 14.11).

When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models, suggesting high levels of measurement invariance across cohorts and gender groups. However, the fit for more constrained models across sites was less satisfactory and comparisons of this scale across sites should be done with some caution. The item parameters used for IRT scaling are shown in Table 14.6.

Table 14.5. Fit indices for combined relationship with parents and friends scale

Fit indices	Pooled sample
RMSEA	0.028
CFI	0.998
TLI	0.998

STQ scale	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
RELPARENTS and RELFRIENDS	Cohort	0.030	0.998	0.997	0.030	0.998	0.997	0.030	0.997	0.997
	Site	0.035	0.998	0.997	0.065	0.992	0.991	0.080	0.981	0.986
	Gender	0.029	0.998	0.997	0.028	0.998	0.998	0.033	0.997	0.997

Item		Factor1 Loadings	Factor2 Loadings	Item label
RELPARENTS	STQM03001	**0.77**		I get upset easily with my parents
	STQM03002	**0.73**		It is hard for me to talk with my parents
	STQM03003	**0.82**		I feel angry with my parents
RELFRIENDS	STQM03101		**0.81**	My friends understand me
	STQM03102		**0.86**	My friends accept me as I am
	STQM03103		**0.84**	My friends are easy to talk to
	STQM03104		**0.85**	My friends respect my feelings

Table 14.6. Item parameters for relationship with parents and friends scales

Scale/Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Perceived relationships with parents	STQM03001	-0.24	-1.86	0.54	1.32
	STQM03002	-0.08	-0.97	0.26	0.72
	STQM03003	0.32	-1.51	0.64	0.86
Perceived relationships with friends	STQM03101	0.56	-2.49	-0.01	2.50
	STQM03102	-0.37	-2.04	-0.08	2.12
	STQM03103	-0.23	-1.86	-0.25	2.10
	STQM03104	0.03	-2.07	-0.10	2.17

14.6.4. Sense of belonging

The *BELONG* scale consisted of six items, three of which were positively worded while the other three were negatively worded. The positively worded items include 'I make friends easily at school', 'I feel like I belong at school' and 'Other students seem to like me'. The negatively worded items include: 'I feel like an outsider (or left out of things) at school', 'I feel awkward and out of place in my school' and 'I feel lonely at school'. For analysis and scaling purposes, the negatively worded items were reverse-coded. Students indicating a greater sense of belonging obtained higher scores on the scale. Table 14.7 shows the fit indices and the factor loadings from the CFA. The overall scale had a good fit (RMSEA = 0.056). The negatively worded items loaded strongly and the positively worded items loaded moderately on the *BELONG* scale. The average reliability (Cronbach's alpha) across all sites was 0.76 (see site-level results in Table 14.11).

When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models. This suggests an acceptable level of measurement invariance across cohorts and gender, while across sites only the metric invariance model had fit indices that were all within an acceptable range. Table 14.8 displays the item parameters used for the final IRT scaling of these items.

Table 14.7. Confirmatory factor analysis fit indices and factor loadings for sense of belonging scale

Fit indices	Pooled sample
RMSEA	0.054
CFI	0.995
TLI	0.988

STQ scale	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
BELONG	Cohort	0.053	0.996	0.989	0.064	0.991	0.984	0.064	0.985	0.984
	Site	0.073	0.993	0.982	0.075	0.987	0.981	0.100	0.954	0.966
	Gender	0.057	0.995	0.987	0.044	0.996	0.992	0.037	0.995	0.994

Item		Factor loadings	Item label
Sense of belonging at school	STQM03701	**0.80**	I feel like an outsider (or left out of things) at school.
	STQM03702	**-0.48**	I make friends easily at school.
	STQM03703	**-0.44**	I feel like I belong at school.
	STQM03704	**0.73**	I feel awkward and out of place in my school.
	STQM03705	**-0.39**	Other students seem to like me.
	STQM03706	**0.84**	I feel lonely at school.
STQM03705 WITH STQM03702		**0.45**	
STQM03705 WITH STQM03703		**0.36**	
STQM03703 WITH STQM03702		**0.37**	

* The items in red font are reverse coded.

Table 14.8. Item parameters for sense of belonging scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Sense of belonging at school	STQM03701	-0.34	-1.12	-0.47	1.59
	STQM03702	0.16	-1.49	-0.60	2.09
	STQM03703	0.27	-1.38	-0.74	2.12
	STQM03704	-0.14	-1.13	-0.54	1.67
	STQM03705	0.35	-1.63	-0.96	2.59
	STQM03706	-0.30	-0.79	-0.59	1.37

* The items in red font are reverse coded.

14.6.5. Student perceptions of school climate and anxiety towards assessment

A multi-dimensional CFA model was constructed using items related to student perceptions of school climate and anxiety towards assessment. For the first dimension, *BULLY*, students were asked how often they had experienced bullying at school over the past 12 months by reporting on the frequency of the following situations: 'Other students made fun of me', 'I was threatened by other students', 'Other students took away or destroyed things that belonged to me' and 'I got hit or pushed around by other students'. Students were given the following options to indicate how often these situations occurred: 'never or almost never', 'a few times a year', 'a few times a month' or 'once a week or more'. Students received higher scores on this scale if they indicated higher frequencies of these situations.

For the second dimension, *RELTEACH*, students were asked about their perceived relationship with their teachers by indicating how often they experienced the following in the past 12 months: 'Most of my teachers treated me fairly', 'I got along well with most my teachers' and 'Most my teachers were interested in my wellbeing'. Students were given the following response options: 'never or almost never', 'a few times a year', 'a few times a month' or 'once a week or more'.

For the third dimension, *ANXTEST*, students were asked to indicate how anxious they felt about testing by selecting their level of agreement ('strongly disagree', 'disagree', 'neither agree nor disagree', 'agree' or 'strongly agree') to the following statements: 'I often worry that it will be difficult for me taking a test', 'Even if I'm well prepared for a test I feel very anxious' and 'I get very tense when I study for a test'. Students received higher scores on this scale if they indicated higher levels of anxiety.

Table 14.9 shows the fit indices and the factor loadings from the multi-dimensional CFA. The overall model had a good fit (RMSEA = 0.035). Being threatened and hit or pushed around loaded strongly on the *BULLY* scale, whereas being made fun of and having possessions taken away or destroyed had somewhat lower loadings. For the *RELTEACH* dimension, 'I got along well with most of my teachers' loaded more strongly on this factor than the other two items. All three items on the *ANXTEST* dimension had equally high loadings. The average reliability (Cronbach's alpha) across all sites for *BULLY* was 0.77, for *RELTEACH* 0.76 and for *ANXTEST* 0.81 (see site-level results in Table 14.11).

When reviewing measurement invariance across multi-group models with different constraints, the model fit was still acceptable for the more constrained models, suggesting acceptable levels of measurement invariance across all groups (i.e. cohort, site and gender). The final scaling parameters for the three scales are displayed in Table 14.10.

Table 14.9. Confirmatory factor analysis fit indices and factor loadings for student perceptions of school climate and anxiety towards assessment

Fit indices	Pooled sample
RMSEA	0.035
CFI	0.994
TLI	0.991

STQ scale	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
BULLY, RELTEACH and ANXTEST	Cohort	0.032	0.995	0.993	0.035	0.993	0.991	0.042	0.987	0.988
	Site	0.043	0.992	0.989	0.056	0.984	0.981	0.066	0.967	0.974
	Gender	0.036	0.993	0.991	0.035	0.993	0.991	0.032	0.992	0.992

Item		Factor1 Loadings	Factor2 Loadings	Factor3 Loadings	Item label
BULLY	STQM03901	**0.69**			Other students made fun of me
	STQM03902	**0.87**			I was threatened by other students
	STQM03903	**0.78**			Other students took away or destroyed things that belonged to me
	STQM03904	**0.85**			I got hit or pushed around by other students
RELTEACH	STQM04101		**0.76**		Most of my teachers treated me fairly.
	STQM04102		**0.95**		I got along well with most of my teachers.
	STQM04103		**0.69**		Most of my teachers were interested in my well-being.
ANXTEST	STQM04201			**0.79**	I often worry that it will be difficult for me taking a test.
	STQM04202			**0.86**	Even if I am well prepared for a test I feel very anxious.
	STQM04203			**0.80**	I get very tense when I study for a test.

Table 14.10. Item parameters for bullying scale

Scale/Item		Item Parameters				
		Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
Bullying at school	STQM03901	-0.62	-0.96	0.45	0.51	-
	STQM03902	0.35	-0.25	-0.07	0.33	-
	STQM03903	0.22	-0.58	0.11	0.48	-
	STQM03904	0.05	-0.25	0.09	0.16	-
Perceived relationships with teachers	STQM04101	-0.03	-0.61	0.18	0.42	-
	STQM04102	-0.46	-1.18	0.33	0.85	-
	STQM04103	0.49	-1.14	-0.01	1.15	-
School anxiety	STQM04201	0.01	-2.28	-0.45	0.91	1.82
	STQM04202	0.22	-2.30	-0.17	0.69	1.78
	STQM04203	-0.23	-2.09	-0.50	0.66	1.94

14.6.6. Reliabilities for student questionnaire scales

Table 14.11 shows the reliabilities for the student questionnaire scales.

Table 14.11. Reliabilities for student questionnaire scales by site

	Student questionnaire scale							
SITE	WELLBEING	GLOBALMIND	RELPARENTS	RELFRIENDS	BELONG	BULLY	RELTEACH	ANXTEST
BOG	0.86	0.72	0.66	0.81	0.69	0.78	0.72	0.72
DAE	0.92	0.83	0.83	0.89	0.77	0.65	0.80	0.87
HEL	0.86	0.80	0.71	0.87	0.83	0.79	0.73	0.79
HOU	0.86	0.83	0.75	0.85	0.71	0.81	0.80	0.79
MAN	0.86	0.73	0.68	0.83	0.71	0.78	0.74	0.76
MOS	0.84	0.75	0.70	0.87	0.82	0.79	0.73	0.85
OTT	0.84	0.78	0.77	0.88	0.81	0.79	0.80	0.82
SIN	0.83	0.76	0.67	0.84	0.72	0.79	0.77	0.75
SUZ	0.89	0.79	0.84	0.92	0.80	0.78	0.75	0.88
TUR	0.86	0.72	0.77	0.85	0.77	0.75	0.73	0.85
Average	**0.86**	**0.77**	**0.74**	**0.86**	**0.76**	**0.77**	**0.76**	**0.81**

14.7. Parent questionnaire

14.7.1. Community closeness

Parents were asked about the extent to which they agreed to statements regarding the area where they live (*COMM*) by selecting one of the following options: 'strongly disagree', 'disagree', 'neither agree nor disagree', 'agree' or 'strongly agree'. Parents received higher scores on this scale if they indicated higher levels of agreement. The fit indices and factor loadings from the CFA are shown in Table 14.12. The overall scale has a good fit (RMSEA = 0.03) and all items had moderate loadings on the scale. The average reliability (Cronbach's alpha) across all sites was 0.76 (see site-level results in Table 14.20). When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models for cohort, suggesting an acceptable level of measurement invariance across these two groups. However, measurement invariance was not at an acceptable level for more constrained models when using site as a grouping variable, so caution is warranted when comparing scale results across sites. Table 14.13 shows the final IRT scaling parameters for this item set.

Table 14.12. Confirmatory factor analysis fit indices and factor loadings for community closeness scale

Fit indices	Pooled sample
RMSEA	0.030
CFI	1.000
TLI	0.997

PAQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
COMM	Cohort	0.042	0.999	0.995	0.024	0.999	0.998	0.028	0.997	0.998
	Site	0.171	0.989	0.933	0.095	0.988	0.980	0.134	0.916	0.959

Item		Factor loadings	Item label
Community closeness	PAQM01501	**0.77**	People in the area where I live help each other out
	PAQM01502	**0.70**	We watch out for each other's children in the area where I live
	PAQM01503	**0.77**	My child is safe in the area where I live
	PAQM01504	**0.68**	When we encounter difficulties, we know where to go for help in our community

PAQM01503 WITH PAQM01502
-0.33

Table 14.13. Item parameters for community closeness scale

Item		Item Parameters				
		Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
Community closeness	PAQM01501	0.06	-1.55	-0.84	-0.01	2.40
	PAQM01502	0.34	-1.43	-0.36	-0.32	2.10
	PAQM01503	-0.20	-1.57	-0.61	-0.10	2.28
	PAQM01504	-0.20	-1.51	-0.66	-0.40	2.57

14.7.2. Parental subjective wellbeing

Parents were asked to respond to statements about how they had been feeling over the last two weeks (*WELLBEING*) by selecting one of the following options: 'at no time', 'some of the time', 'more than half of the time', 'most of the time' or 'all the time'. Parents received higher scores on the wellbeing scale if they indicated feeling positive more frequently. The fit indices and factor loadings from the CFA are shown in Table 14.14. The overall scale had a good fit (RMSEA = 0.087) and items tended to have strong factor loadings. The average reliability (Cronbach's alpha) across all sites was 0.88 (see site-level results in Table 14.20). When reviewing measurement invariance across multi-group models with different constraints, the model fit was satisfactory for the more constrained models for the cohort grouping variable, while the results suggested only metric invariance across sites. Table 14.15 shows the final IRT scaling parameters.

Table 14.14. Confirmatory factor analysis fit indices and factor loadings for parental subjective wellbeing scale

Fit indices	Pooled sample
RMSEA	0.087
CFI	0.997
TLI	0.992

PAQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
WELLBEING	Cohort	0.087	0.997	0.992	0.053	0.998	0.997	0.034	0.998	0.999
	Site	0.093	0.997	0.992	0.065	0.997	0.996	0.112	0.979	0.988

Item		Factor loadings	Item label
Parental subjective well-being	PAQM02501	**0.83**	I have felt cheerful and in good spirits
	PAQM02502	**0.79**	I have felt calm and relaxed
	PAQM02503	**0.86**	I have felt active and vigorous
	PAQM02504	**0.83**	I have woken up feeling fresh and rested
	PAQM02505	**0.77**	My daily life has been filled with things that interest me
PAQM02502 WITH PAQM02501		**0.27**	

Table 14.15. Item parameters for parental subjective wellbeing scale

Item		Item Parameters				
		Delta	Tau(1)	Tau(2)	Tau(3)	Tau(4)
Parental subjective well-being	PAQM02501	-0.35	-4.18	-0.19	0.60	3.78
	PAQM02502	0.01	-3.87	-0.34	0.55	3.67
	PAQM02503	-0.20	-3.70	-0.37	0.67	3.40
	PAQM02504	0.31	-3.34	-0.22	0.39	3.18
	PAQM02505	0.23	-3.45	0.07	0.43	2.96

14.7.3. Need for encouragement

Parents were asked how much encouragement (*ENCOUR)* their child needed to do the following: 'To study hard and be responsible', 'To stay relaxed and calm even in moments of difficulty', 'To socialize with other people', 'To be kind and help other people' and 'To learn new things' by selecting one of the following options: 'none at all', 'little', 'somewhat', 'most of the time' or 'a lot'. Parents received higher scores on the encouragement scale if they indicated their child needed more encouragement. The fit indices and factor loadings from the CFA are shown in Table 14.16. The CFA for the pooled dataset indicated only poor model fit (RMSEA = 0.104). Three of the five items had relatively higher loadings on the scale than the other two items. The average reliability (Cronbach's alpha) across all sites was 0.84 (see site-level results in Table 14.20). When reviewing measurement invariance across multi-group models with different constraints, the model fit was similar for the more constrained models, suggesting the fit for the measurement model was similar across cohorts and sites; however, it should be noted that the model in general had poor or marginally poor fit. Table 14.17 displays the final IRT scaling parameters.

Table 14.16. Confirmatory factor analysis fit indices and factor loadings for level of necessary encouragement scale

Fit indices	Pooled sample
RMSEA	0.104
CFI	0.996
TLI	0.987

PAQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
ENCOUR	Cohort	0.105	0.996	0.987	0.082	0.996	0.992	0.074	0.994	0.994
	Site	0.114	0.996	0.986	0.084	0.995	0.993	0.099	0.986	0.989

Item		Factor loadings	Item label
Need for encouragement	PAQM02901	**0.57**	To study hard and be responsible
	PAQM02902	**0.60**	To stay relaxed and calm even in moments of difficulty
	PAQM02903	**0.86**	To socialize with other people
	PAQM02904	**0.96**	To be kind and help other people
	PAQM02905	**0.84**	To learn new things
	PAQM02902 WITH PAQM02901	**0.33**	
	PAQM02905 WITH PAQM02901	**0.29**	

Table 14.17. Item parameters for level of necessary encouragement scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Need for encouragement	PAQM02901	-0.66	-1.61	-0.24	1.85
	PAQM02902	-0.51	-2.07	-0.31	2.38
	PAQM02903	0.55	-1.38	-0.02	1.41
	PAQM02904	0.57	-1.01	0.22	0.79
	PAQM02905	0.06	-1.17	-0.11	1.27

14.7.4. Parental engagement at home

Parents were asked how often they or someone else at home engaged with their child (*ENGAGE*) by responding ('never or hardly ever', 'once or twice a year', 'once or twice a month', 'once or twice a week' or 'every day or almost every day') to the following: 'Discuss how well my child is doing at school', 'Eat the main meal with my child around the table', 'Spend time just talking to my child' and 'Help my child with his/her homework'. Parents received higher scores on the engagement scale if they indicated they engaged with their child more frequently. The fit indices and factor loadings from the CFA are shown in Table 14.18. The overall scale had a marginally satisfactory fit (RMSEA = 0.085). The item 'Spend time just talking to my child' had a higher loading than other items measuring this scale. The average reliability (Cronbach's alpha) across all sites was not satisfactory with 0.58 (see site-level results in Table 14.20). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was somewhat less satisfactory for the more constrained model indicating a notable lack of measurement invariance across cohorts and sites, in particular at the scalar level. Table 14.19 shows the final IRT scaling parameters.

Table 14.18. Confirmatory factor analysis fit indices and factor loadings for parental engagement at home scale

Fit indices	Pooled sample
RMSEA	0.085
CFI	0.988
TLI	0.963

PAQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
ENGAGE	Cohort	0.059	0.993	0.980	0.120	0.952	0.918	0.170	0.809	0.837
	Site	0.077	0.988	0.963	0.060	0.984	0.978	0.135	0.810	0.886

Item		Factor loadings	Item label
Parental engagement - at home	PAQM03001	**0.78**	Discuss how well my child is doing at school.
	PAQM03002	**0.58**	Eat <the main meal> with my child around a table.
	PAQM03003	**0.83**	Spend time just talking to my child.
	PAQM03004	**0.54**	Help my child with his/her homework.

Table 14.19. Item parameters for parental engagement at home scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Parental engagement - at home	PAQM01501	-0.20	-0.35	-0.32	0.67
	PAQM01502	-0.58	0.43	-0.74	0.31
	PAQM01503	-0.20	-0.32	-0.44	0.76
	PAQM01504	0.97	0.25	-0.74	0.49

14.7.5. Reliabilities for parent questionnaire scales

Table 14.20 shows the reliabilities for the parent questionnaire scales.

Table 14.20. Reliabilities for parent questionnaire scales by site

	Parent questionnaire scale			
SITE	COMM	WELLBEING	ENCOUR	ENGAGE
BOG	0.60	0.85	0.85	0.64
DAE	0.80	0.91	0.83	0.67
HEL	0.78	0.86	0.75	0.50
HOU	0.85	0.92	0.91	0.66
MAN	0.67	0.88	0.86	0.66
MOS	0.71	0.86	0.81	0.50
OTT	0.82	0.85	0.80	0.46
SIN	0.76	0.87	0.87	0.50
SUZ	0.80	0.91	0.87	0.70
TUR	0.81	0.87	0.85	0.54
Average	**0.76**	**0.88**	**0.84**	**0.58**

14.8. Principal questionnaire

14.8.1. Promotion of social and emotional skills

Principals were asked ('yes' or 'no') whether their schools promoted student development of social and emotional skills (*PROMSSES)* through eight different activities and practices. Higher scores on the scale were obtained by schools with greater promotion of social and emotional skills. The fit indices and factor loadings from the CFA are shown in Table 14.21. The overall scale had a satisfactory fit (RMSEA = 0.045) and there was some variation in the strength of factor loadings across items measuring this scale. The average reliability (Cronbach's alpha) across all sites was 0.65 (see site-level results in Table 14.27). Since the items were dichotomous, measurement invariance could not be fully assessed using multiple-group models, however, the results for the scalar model across cohorts indicated an acceptable level of invariance. Measurement invariance was not performed across sites due to insufficient data. Table 14.22 displays the final IRT scaling parameters for this scale.

Table 14.21. Confirmatory factor analysis fit indices and factor loadings for promotion of SE skills scale

Fit indices	Pooled sample
RMSEA	0.045
CFI	0.968
TLI	0.953

PRQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
PROMSSES	Cohort	0.047	0.965	0.948	N/A	N/A	N/A	0.041	0.970	0.962

	Item	Factor loadings	Item label
Promotion of SE skills	PRQM01401	**0.85**	Teachers are asked to promote the development of students' social and emotional skills as part of their work
	PRQM01402	**0.73**	The development of social and emotional skills is one of the objectives included in the <school educational plan>
	PRQM01403	**0.47**	We have separate classes or school activities dedicated specifically to the development of these skills
	PRQM01404	**0.48**	As part of special classes aimed specifically at developing these skills
	PRQM01405	**0.68**	By how we implement our school's disciplinary rules
	PRQM01406	**0.86**	By means of our general school practices
	PRQM01407	**0.52**	By organising extracurricular activities
	PRQM014018	**0.67**	By providing feedback and advice to parents about their children's social and emotional skills
	PRQM01404 WITH PRQM01403	**0.57**	

Table 14.22. Item parameters for promotion of SE skills scale

Item		Delta
Promotion of SE skills	PRQM01401	-1.56
	PRQM01402	-0.25
	PRQM01403	1.45
	PRQM01404	1.99
	PRQM01405	0.05
	PRQM01406	-1.42
	PRQM01407	0.22
	PRQM01408	-0.48

14.8.2. School climate – teachers' disruptive behaviours

Principals were asked about the extent ('not at all', 'a little', 'to some extent' or 'a lot') to which student learning was hindered by disruptive behaviours among teachers (*TEABEHA)*. Given low frequencies, the response categories 'to some extent' and 'a lot' were collapsed into one category for analysis and scaling. Higher scores on the scale were obtained by principals indicating student learning was hindered to a greater extent by disruptive teacher behaviour. The fit indices and factor loadings from the CFA are shown in Table 14.23. The overall scale had only marginally satisfactory fit (RMSEA = 0.093) and most items had strong factor loadings. The average reliability (Cronbach's alpha) across all sites was 0.75 (see site-level results in Table 14.27). When reviewing

measurement invariance using multiple-group models with different constraints, the model fit was satisfactory for the more constrained model indicating acceptable levels of measurement invariance across cohorts. Measurement invariance was not performed across sites due to insufficient data. Table 14.24 shows the final IRT scaling parameters.

Table 14.23. Confirmatory factor analysis fit indices and factor loadings for teachers' disruptive behaviours scale

Fit indices	Pooled sample
RMSEA	0.093
CFI	0.993
TLI	0.985

PRQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
TEABEHA	Cohort	0.091	0.993	0.986	0.077	0.994	0.990	0.067	0.994	0.992

Item		Factor loadings	Item label
School Climate - teachers' (disruptive) behaviours	PRQM02701	**0.86**	Teachers not meeting individual students' needs
	PRQM02702	**0.82**	Teacher absenteeism
	PRQM02703	**0.81**	Staff resisting change
	PRQM02704	**0.73**	Teachers being too strict with students
	PRQM02705	**0.90**	Teachers being late to classes
	PRQM02706	**0.84**	Teachers not being well prepared for classes
PRQM02704 WITH PRQM02703		**0.42**	
PRQM02705 WITH PRQM02701		**-0.66**	

Table 14.24. Item parameters for teachers' disruptive behaviours scale

Item		Item Parameters		
		Delta	Tau(1)	Tau(2)
School Climate - teachers' (disruptive) behaviours	PRQM02701	-1.17	-1.57	1.57
	PRQM02702	0.24	-1.24	1.24
	PRQM02703	-0.72	-1.20	1.20
	PRQM02704	0.07	-1.79	1.79
	PRQM02705	1.12	-1.29	1.29
	PRQM02706	0.47	-1.47	1.47

14.8.3. School climate – students' disruptive behaviours

Principals were asked about the extent ('not at all', 'a little', 'to some extent' or 'a lot') to which student learning was hindered by students' disruptive behaviours (*STUBEHA*). Higher scores on the scale were obtained by principals indicating that student learning was hindered to a greater extent by disruptive student behaviour. The fit indices and factor

loadings from the CFA are shown in Table 14.25. The overall scale had only marginally satisfactory fit (RMSEA = 0.093) and there was some variation in the strength of factor loadings across items. The average reliability (Cronbach's alpha) across all sites was 0.83 (see site-level results in Table 14.27). When reviewing measurement invariance using across multiple-group models with different constraints for cohort groups, the model fit was similarly poor across all three multiple-group models. Measurement invariance was not performed across sites due to insufficient data. Table 14.26 shows the final IRT scaling parameters.

Table 14.25. Confirmatory factor analysis fit indices and factor loadings for students' disruptive behaviours scale

Fit indices	Pooled sample
RMSEA	0.093
CFI	0.996
TLI	0.990

PRQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
STUBEHA	Cohort	0.113	0.995	0.989	0.140	0.990	0.983	0.115	0.988	0.988

Item		Factor loadings	Item label
School Climate - students' (disruptive) behaviours	PRQM02501	**0.94**	Student truancy
	PRQM02502	**0.84**	Students skipping classes
	PRQM02503	**0.76**	Students arriving late for school
	PRQM02504	**0.68**	Students lacking respect for teachers
	PRQM02505	**0.59**	Disruption of classes by students

PRQM02505 WITH PRQM02504

0.68

Table 14.26. Item parameters for students' disruptive behaviours scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
School Climate - students' (disruptive) behaviours	PRQM02501	-0.34	-1.99	0.22	1.77
	PRQM02502	-0.01	-2.01	0.20	1.81
	PRQM02503	-0.24	-3.31	0.32	3.00
	PRQM02504	0.47	-2.58	0.31	2.26
	PRQM02505	0.12	-2.33	0.16	2.17

14.8.4. Reliabilities for principal questionnaire scales

Table 14.27 shows the reliabilities for the principal questionnaire scales.

Table 14.27. Reliabilities for principal questionnaire scales by site

	Principal questionnaire scales		
SITE	PROMSSES	TEABEHA	STUBEHA
BOG	0.56	0.83	0.88
DAE	0.72	0.79	0.88
HEL	0.60	0.63	0.76
HOU	0.47	0.84	0.85
MAN	0.56	0.84	0.84
MOS	0.69	0.48	0.76
OTT	0.66	0.72	0.69
SIN	0.71	0.63	0.81
SUZ	0.85	0.93	0.95
TUR	0.68	0.84	0.86
Average	**0.65**	**0.75**	**0.83**

14.9. Teacher questionnaire

14.9.1. Active learning pedagogies

Teachers were asked to indicate how often ('never or almost never', 'some lessons', 'many lessons' or 'every lesson or almost every lesson') various active learning pedagogies (*ACTPED)* were occurring in their lessons. For analyses and scaling the response categories 'never or almost never' and 'some lessons' were combined given relatively low frequencies in these categories. Teachers received higher scores on this scale if they indicated they applied these learning pedagogies more often.

The fit indices and factor loadings from the CFA are shown in Table 14.28. The overall scale had an acceptable fit (RMSEA = 0.083) and there was some variation in the strength of factor loadings across items. The average reliability (Cronbach's alpha) across all sites was 0.78 (see site-level results in Table 14.38). When reviewing measurement invariance using across multiple-group models with different constraints for the grouping variable cohorts, acceptable levels of measurement invariance were observed. However, the analyses showed a lack of measurement invariance across sites. Table 14.29 displays the final IRT scaling parameters for these items.

Table 14.28. Confirmatory factor analysis fit indices and factor loadings for active learning pedagogies scale

Fit indices	Pooled sample
RMSEA	0.083
CFI	0.985
TLI	0.973

TCQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
ACTPED	Cohort	0.091	0.982	0.966	0.078	0.982	0.975	0.076	0.979	0.976
	Site	0.073	0.995	0.990	0.186	0.948	0.937	0.202	0.917	0.926

Item		Factor loadings	Item label
Active learning pedagogies	TCQM01301	**0.67**	Students are given opportunities to explain their ideas.
	TCQM01302	**0.81**	A small group discussion between students takes place.
	TCQM01303	**0.84**	A whole class discussion takes place in which I participate.
	TCQM01304	**0.75**	I discuss questions that students ask.
	TCQM01305	**0.59**	Students present something to the rest of the class.
	TCQM01306	**0.55**	Students discuss materials from a textbook.
	TCQM01306 WITH TCQM01305	**0.39**	

Table 14.29. Item parameters for active learning pedagogies scale

Item		Item Parameters		
		Delta	Tau(1)	Tau(2)
Active learning pedagogies	TCQM01301	-1.32	-1.13	1.13
	TCQM01302	0.13	-0.98	0.98
	TCQM01303	0.23	-0.75	0.75
	TCQM01304	-0.43	-1.00	1.00
	TCQM01305	0.83	-0.81	0.81
	TCQM01306	0.56	-0.72	0.72

14.9.2. Teacher pedagogies

Teachers were asked to indicate the extent ('not at all', 'to some extent', 'quite a bit' or 'a lot') to which they could apply various teaching pedagogies (*PEDAGOG*). The response categories 'not at all' and 'to some extent' were combined together into one category. Teachers received higher scores on this scale if they indicated they were able to apply the teaching pedagogies to a greater extent. The fit indices and factor loadings from the CFA are shown in Table 14.30. The overall scale had only marginally satisfactory fit (RMSEA = 0.097) and there was some variation in the strength of factor loadings across items. The average reliability (Cronbach's alpha) across all sites was 0.85 (see site-level results in Table 14.38). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was satisfactory for the more constrained model indicating acceptable levels of measurement invariance across cohorts. However,

model fit was poor for the more constrained models across sites. Table 14.31 shows the final IRT parameters used for scaling.

Table 14.30. Confirmatory factor analysis fit indices and factor loadings for teacher pedagogies scale

Fit indices	Pooled sample
RMSEA	0.097
CFI	0.987
TLI	0.980

TCQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
PEDAGOG	Cohort	0.097	0.987	0.980	0.083	0.989	0.985	0.078	0.988	0.987
	Site	0.105	0.990	0.984	0.098	0.988	0.986	0.109	0.981	0.983

Item		Factor loadings	Item label
Teacher pedagogies	TCQM01601	**0.85**	Get students to believe they can do well in school work.
	TCQM01602	**0.90**	Help my students to value learning.
	TCQM01603	**0.74**	Control disruptive behaviour in the classroom.
	TCQM01604	**0.84**	Motivate students who show low interest in school work.
	TCQM01605	**0.80**	Make expectations about student behaviour clear.
	TCQM01606	**0.78**	Help students think critically.
	TCQM01607	**0.73**	Get students to follow classroom rules.
	TCQM01607	WITH TCQM01603	
		0.47	

Table 14.31. Item parameters for teacher pedagogies scale

Item		Item Parameters		
		Delta	Tau(1)	Tau(2)
Teacher pedagogies	TCQM01601	-0.24	-1.65	1.65
	TCQM01602	-0.23	-1.49	1.49
	TCQM01603	0.15	-1.54	1.54
	TCQM01604	0.77	-1.39	1.39
	TCQM01605	-0.34	-1.51	1.51
	TCQM01606	0.39	-1.43	1.43
	TCQM01607	-0.51	-1.69	1.69

14.9.3. School climate – students' disruptive behaviours

The teachers were asked to what extent ('not at all', 'very little', 'to some extent' or 'a lot') student learning was hindered by students' disruptive behaviours (*STDISR)*. Teachers received higher scores on this scale if they indicated student learning was hindered by disruptive behaviour to a greater extent. The fit indices and factor loadings from the CFA

are shown in Table 14.32. The overall scale had a good fit (RMSEA = 0.042) and all items had high factor loadings. The average reliability (Cronbach's alpha) across all sites was 0.84 (see site-level results in Table 14.38). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was less satisfactory for the more constrained model indicating a lack of measurement invariance across cohorts and sites. Table 14.33 shows the final IRT scaling parameters.

Table 14.32. Confirmatory factor analysis fit indices and factor loadings for students' disruptive behaviour scale

Fit indices	Pooled sample
RMSEA	0.042
CFI	1.000
TLI	0.999

TCQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
STDISR	Cohort	0.044	1.000	0.999	0.073	0.999	0.997	0.102	0.995	0.995
	Site	0.093	0.998	0.995	0.149	0.991	0.987	0.186	0.970	0.979

Item		Factor loadings	Item label
School climate - students' (disruptive) behaviour	TCQM02301	**0.82**	Student truancy
	TCQM02302	**0.80**	Students skipping classes
	TCQM02303	**0.76**	Students arriving late for school
	TCQM02304	**0.80**	Students lacking respect for teachers
	TCQM02305	**0.79**	Disruption of classes by students
TCQM02305 WITH TCQM02304		**0.68**	
TCQM02302 WITH TCQM02301		**0.57**	

Table 14.33. Item parameters for students' disruptive behaviour scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
School Climate - students' (disruptive) behaviours	TCQM02301	0.08	-1.24	-0.09	1.33
	TCQM02302	0.20	-1.37	-0.18	1.56
	TCQM02303	-0.03	-2.33	-0.05	2.38
	TCQM02304	0.12	-1.80	0.02	1.78
	TCQM02305	-0.37	-1.78	-0.06	1.84

14.9.4. School climate – quality of relationships

Teachers were asked about the school environment (*CLIMATE*) by indicating their level of agreement ('strongly disagree', 'disagree', 'neither agree nor disagree', 'agree' or 'strongly agree') with various statements. The categories 'strongly disagree' and 'disagree' were combined together into one category for analysis and scaling due to low frequencies in

these categories. Teachers received higher scores on this scale if they indicated a more supportive school environment. The fit indices and factor loadings from the CFA are shown in Table 14.34. The overall scale had poor model fit (RMSEA = 0.101) and one of four items had a relatively lower factor loading compared to the other items. The average reliability (Cronbach's alpha) across all sites was 0.82 (see site-level results in Table 14.38). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was satisfactory for the more constrained model indicating an acceptable level of measurement invariance across cohorts. However, across sites only metric but not scalar invariance was recorded. Table 14.35 displays the item parameters used for the final scaling.

Table 14.34. Confirmatory factor analysis fit indices and factor loadings for quality of relationships scale

Fit indices	Pooled sample
RMSEA	0.101
CFI	0.997
TLI	0.992

TCQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
CLIMATE	Cohort	0.099	0.997	0.992	0.053	0.999	0.998	0.041	0.998	0.999
	Site	0.090	0.999	0.996	0.069	0.998	0.998	0.108	0.990	0.995

Item		Factor loadings	Item label
School climate - quality of relationships	TCQM02601	**0.82**	In this school, teachers and students usually get on well with each other.
	TCQM02602	**0.90**	Most teachers in this school believe that the students' well-being is important.
	TCQM02603	**0.90**	Most teachers in this school are interested in what students have to say.
	TCQM02604	**0.73**	If a student from this school needs extra assistance, the school provides it.

Table 14.35. Item parameters for quality of relationships scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
School climate - quality of relationships	TCQM02601	-0.05	-1.98	-1.29	3.26
	TCQM02602	-0.34	-1.58	-1.08	2.66
	TCQM02603	0.16	-2.03	-0.88	2.91
	TCQM02604	0.23	-2.02	-0.66	2.68

14.9.5. Bullying

Teachers were asked how often ('less than once a month', '1-5 times per month', 'once a week' or 'more than once a week') they had been informed about situations related to bullying in the current school year (*BULLY*). Due to low frequencies, the categories 'once a week' and 'more than once a week' were combined for analyses and scaling purposes.

The fit indices and factor loadings from the CFA are shown in Table 14.36. The CFA for the combined dataset showed an acceptable model fit (RMSEA = 0.08) and there was some variation in the strength of item loadings for this scale. The average reliability (Cronbach's alpha) across all sites was 0.89 (see site-level results in Table 14.38). When reviewing measurement invariance using across multiple-group models with different constraints, the model fit was less satisfactory for the more constrained model indicating a lack of measurement invariance across cohorts and sites. Table 14.37 displays the parameters used for the IRT scaling of these items.

Table 14.36. Confirmatory factor analysis fit indices and factor loadings for bullying scale

Fit indices	Pooled sample
RMSEA	0.080
CFI	0.994
TLI	0.991

TCQ scales	Group	Configural			Metric			Scalar		
		RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
BULLY	Cohort	0.081	0.994	0.991	0.109	0.987	0.983	0.108	0.983	0.983
	Site	0.094	0.995	0.992	0.118	0.989	0.987	0.105	0.986	0.990

Item		Factor loadings	Item label
Bullying	TCQM02701	**0.77**	A student informed you about aggressive or destructive behaviours by other students.
	TCQM02702	**0.91**	A student informed you that he/she was <bullied> by another student.
	TCQM02703	**0.94**	A teacher informed you that a student was <bullied> by other students.
	TCQM02704	**0.90**	A teacher informed you that a student helped another student who was being <bullied>.
	TCQM02705	**0.68**	A student informed you that he/she was <bullied> by a teacher.
	TCQM02706	**0.85**	A parent informed you that his/her son/daughter was <bullied> by other students.
	TCQM02707	**0.74**	A teacher informed you that he/she was <bullied> by students.
	TCQM02708	**0.84**	You witnessed students' <bullying> behaviours.

TCQM02707 WITH TCQM02705
0.50
TCQM02702 WITH TCQM02701
0.46

Table 14.37. Item parameters for bullying scale

Item		Item Parameters			
		Delta	Tau(1)	Tau(2)	Tau(3)
Bullying	TCQM02701	-1.66	-2.08	0.98	1.10
	TCQM02702	-0.88	-1.97	0.97	1.00
	TCQM02703	-0.30	-1.88	0.80	1.08
	TCQM02704	0.13	-1.80	0.90	0.90
	TCQM02705	1.48	-1.11	0.68	0.43
	TCQM02706	0.41	-2.36	0.92	1.43
	TCQM02707	0.99	-1.16	0.62	0.54
	TCQM02708	-0.16	-1.59	0.70	0.89

14.9.6. Reliabilities for teacher questionnaire scales

Table 14.38 shows the reliabilities for the teacher questionnaire scales.

Table 14.38. Reliabilities for teacher questionnaire by site

	Teacher questionnaire scales				
SITE	ACTPED	PEDAGOG	STDISR	CLIMATE	BULLY
BOG	0.77	0.82	0.81	0.80	0.87
DAE	0.89	0.90	0.86	0.85	0.92
HEL	0.72	0.81	0.80	0.77	0.86
HOU	0.79	0.88	0.86	0.78	0.89
MAN	0.80	0.85	0.82	0.81	0.88
MOS	0.63	0.78	0.84	0.83	0.87
OTT	0.69	0.81	0.74	0.77	0.84
SIN	0.79	0.86	0.87	0.75	0.88
SUZ	0.91	0.93	0.94	0.93	0.95
TUR	0.80	0.85	0.89	0.86	0.91
Average	**0.78**	**0.85**	**0.84**	**0.82**	**0.89**

14.10. Computation of composite index

14.10.1. Socio-economic status

A measure of parental socio-economic status (SES) was derived for each site, based on three indices: highest level of parental occupation, highest level of parental education and household possessions.

Occupational data was collected using open-ended questions in both the parent and student questionnaires. The responses were coded to four-digit ISCO codes and then mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom, de Graaf, & Treiman, 1992). The highest occupational status of parents (HISEI) corresponds to the higher ISEI score of either parent or to the only available parent's ISEI score. In instances where there was no information from the parent questionnaire, data from the student questionnaire was used. A higher ISEI score indicates higher levels of occupational status.

In the parent questionnaire, respondents were asked about the highest level of education of each of the student's parents with questions using nationally appropriate terms according to the International Standard Classification of Education scheme (ISCED) (UNESCO, 2011). Respondents were asked to select from eight levels ranging from ISCED level 1, through to ISCED level 8. A condensed version of this question was asked in the student questionnaire (with nationally adapted options given to respondents 'ISCED 3 and below', 'ISCED 4 or 5' and 'ISCED 6 and above'). A variable, HISCED was derived by taking the highest level of education of either parent from the parent questionnaire. If the data was only available for one parent, then that is used as the highest level. In instances where there was no information from the parent questionnaire, data from the student questionnaire was used. For each site, the number of years typically spent at each ISCED level was converted into a continuous variable on the number of years spent in formal education (PAREDYRS) (see Annex D for details). In order to obtain consistency between the parent and student data, the computation of PAREDYRS using data from the parent questionnaire was capped at the number of years for ISCED 3, ISCED 4 or 5 and ISCED 6. For example, if a

respondent indicated that one parent completed an ISCED level 8 qualification, the appropriate number of years for formal education for an ISCED level 6 qualification was recorded for PAREDYRS.

The household possessions index consists of student-reported possessions at home, resources available in the home and the number of books in the home. Site-specific wealth items were also included in the computation of the HOMEPOS index (see Annex E for details).

Missing values for respondents with missing data for only one variable were imputed with predicted values plus a random component based on a regression of the other two variables. If there were missing data on more than one variable, SES was not computed for that student and a missing value was assigned for SES. Variables with imputed values were then used for a principal component analysis at the site level.

The SES scores were obtained as component scores for the first principal component with zero being the score of an average respondent within each site and one being the standard deviation. Table 14.39. shows the standardised factor loadings from the principal component analysis (PCA) results, by cohort and site. The table also includes the SES scale reliability (Cronbach's alpha) for the z-standardised variables.

Table 14.39. Factor loadings and reliability (Cronbach's Alpha) of SES, by cohort and site

	Younger Cohort				Older Cohort			
Site	PAREDYRS	HISEI	HOMEPOS	Cronbach's Alpha	PAREDYRS	HISEI	HOMEPOS	Cronbach's Alpha
BOG	0.83	0.83	0.64	0.65	0.83	0.84	0.73	0.73
DAE	0.65	0.73	0.73	0.48	0.71	0.72	0.72	0.53
HEL	0.62	0.78	0.70	0.48	0.68	0.80	0.74	0.58
HOU	0.64	0.70	0.70	0.42	0.77	0.78	0.75	0.65
MAN	0.82	0.81	0.63	0.62	0.84	0.84	0.73	0.73
MOS	0.77	0.78	0.55	0.49	0.75	0.78	0.65	0.56
OTT	0.72	0.75	0.61	0.47	0.78	0.78	0.64	0.58
SIN	0.74	0.78	0.64	0.53	0.79	0.82	0.72	0.67
SUZ	0.81	0.77	0.68	0.62	0.80	0.76	0.74	0.65
TUR	0.81	0.85	0.71	0.69	0.82	0.85	0.74	0.72

References

Bentler, P., & Bonett, D. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. doi:10.1037/0033-2909.88.3.588

Bollen, K. A., & Long, S. J. (Eds.). (1993). *Testing structural equation models.* Newbury Park, CA: Sage Publications.

Flora, D., & Curran, P. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods, 9*(4), 466-491. doi:10.1037/1082-989x.9.4.466

Ganzeboom, H., de Graaf, P., & Treiman, D. (1992). A standard international socioeconomic index of occupational status. *Social Science Research, 21*, 1-56. doi:10.1016/0049-089X(92)90017-B

Hu, L.-t., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. doi:10.1037/1082-989x.1.2.130

Masters, G., & Wright, B. (1997). The partial credit model. In W. van der Linden, R. Hambleton, & (Eds), *Handbook of modern item response theory* (pp. 101-122). New York/Berline/Heidelberg: Springer.

Muthén, B. O., du Toit, S. H., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical outcomes.* Retrieved from http://www.statmodel.com/bmuthen/articles/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide: Statistical Analysis with Latent Variables* (7th ed.). Los Angeles, CA: Muthén & Muthén.

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

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*(2), 229-239. doi:10.1111/j.2044-8317.1974.tb00543.x

UNESCO. (2011). *ISCED Mappings*. Retrieved from http://uis.unesco.org/en/isced-mappings

Chapter 15. International Data Products

This chapter describes the final databases for the OECD Survey on Social and Emotional Skills 2019 (SSES) produced after completing the scaling analysis. These include the student database, parent database, teacher indirect assessment database, teacher questionnaire database and principal database.

15.1. Introduction

After finalising the scaling and scoring of the SSES data, the International Contractor prepared the final data files in SPSS® data format and prepared the corresponding codebooks. Both data files and codebooks form the international data product, which was delivered to the OECD. These data files are available on the OECD website.

This chapter provides details about the final SSES database.

15.2. The SSES international data products

15.2.1. Data files and codebooks

SSES consisted of two parts, an assessment of social and emotional skills as well as contextual questions. Both parts were administered to students, teachers and parents, while school principals were asked to provide contextual data regarding the sampled schools. There are five data sets with their corresponding codebooks. The codebook provides details, including valid codes, for each of the variables. Each data set contains data collected from the online platform, derived variables, weights and scores wherever applicable.

There were various conditions to determine if a respondent was eligible according to the SSES Technical Standards (see Annex A). Several variables were created to determine the eligibility of a respondent. These variables were: Resp_STA and Pct50_STA, which indicate the count of valid responses and whether a respondent had at least 50% of valid responses for the assessment items, respectively. The variable SchPart indicates whether the school had at least 25% or more eligible students who provided at least 50% of valid responses to the assessment items at the school level.

Student data

The student data file consists of ID variables, background variables, assessment data (including the response to the anchoring vignettes, cognitive items and assessment items), the students' final weights and replicate weights, two sets of scale scores for each of the 17 scales (including the acquiescence score), other derived variables and national variables. A variable name with a prefix of 'TF' is a variable merged from the Student Tracking Form (STF). UserName_Std is a 9-digit unique student ID which is a composition of Cohort ID, SiteID, SchID and StdID. As all records extracted from the online platform were included in the final datasets regardless of whether a student had met the eligible criteria, two derived variables (SchPart=1 and inSSES_STD=1) can be used to select eligible students, in addition, an ineligible student would have a zero final weight.[1]

There were multiple variables from different sources providing the same information. For example, three variables relate to student gender status, TF_gender sourced from the STF, STQM00401 is a questionnaire item and Gender_Std is the final student gender variable that is computed from TF_gender and STQM00401 (see Chapter 10 for details of how Gender_Std was computed). Similarly, DOB_Std and Age_Std are the variables to be used for student date of birth and age, respectively.

In order to determine a student's eligibility, various factors need to be taken into account: the number of items a student provided with a valid response (Resp_STA and Pct50_STA), student eligible status recorded in the STF (TF_Eligible) and school response rate (pct_SchPart and SchPart).

Student final weight (WT2019) and 76 replicate weight variables (rwgt1-rwgt76) were included in the student data file.

Two sets of scale scores were provided in the student data: <Scale>[2]_WLE and <Scale>_WLE_Adj. Each set contains 17 variables, one for each skill scale. <Scale>_WLEs are the scale scores based on the final selection items for each scale. <Scale>_WLE_Adj are the scale scores after controlling for acquiescence and are recommended to be used for analysis.[3] Two acquiescence variables are included: ARS is the computed acquiescence score and ARS_PAIRS is the number of item pairs used for calculating the acquiescence score that a student has provided with valid responses.

Scale scores were provided for the eight student questionnaire contextual indices, identified by the prefix 'ST'.

Parent data

The parent data file consists of ID variables, parent responses to assessment items, parent final weight (WT2019_PA) and two sets of scale scores for each of the 17 scales (including the acquiescence response set indicator).

UserName_PA is a unique 10-digit parent ID which is a composition of student ID UserName_Std and '9' at the end. In order to match easily with student data, UserName_Std was added into this data. An eligible parent record requires the parent to provide valid responses to 50% of the items (Resp_PAA and Pct50_PAA) and the indirect assessed student is from a school with at least 25% of eligible students with at least 50% of STA valid responses at the school level (schPart=1).

Similar to the student data, two sets of scale scores were provided in the parent data: <Scale>_PA_WLE and <Scale>_PA_WLE_Adj. <Scale>_PA_WLE are the scale scores based on the final selected items for each scale in parent data, and <Scale>_PA_WLE_Adj are the scale scores after controlling for Acquiescence and are recommended to be used for analysing parent data. Two acquiescence variables ARS_PA and ARS_PA_Pairs are also included in the parent data. ARS_PA is the computed parent acquiescence score and ARS_PA_Pairs is the number of item pairs used for calculating the parent acquiescence score that the parent has provided with valid responses.

Scale scores were provided for the four parent questionnaire contextual indices, identified by the prefix 'PA'.

Teacher data

There are two datasets provided for teacher data: the first dataset contains the teachers' indirect assessment responses that consists of ID variables, teacher responses to the assessment items, the teachers' final weights (WT2019_TC) and raw score averages (one for each of the 15 original scales). In this data file, one teacher may contribute one or more records, one for each of the students assessed by the teacher. The second data file contains the teacher contextual questionnaire responses that consists of ID variables, the teacher background variables and responses to the teacher contextual questionnaire items. This data file has one record per surveyed teacher.

In both teacher data files, UserName_TC is a unique teacher ID variable. In the first data file, the student unique variable UserName_std is also included to identify the indirect assessment provided to a student. An eligible teacher is required to provide valid responses to 50% of the indirect assessment items (Resp_TCA and Pct50_TCA) per student and with schPart=1. In addition, four variables: Pct50_TCATot, TCA_std_Tot, TCA_SENexclTot and Pct_TCA_overall indicate how many valid responses a teacher provided across the

students who were provided with indirect assessment. The average of raw score per scale (over three items) for the 15 base scales were included in the first teacher data file. In the second teacher data file, variable Resp_TCV is the number of valid responses provided by a teacher to the anchoring vignettes item. Itm3_TCV is a flag variable that indicates whether a teacher has provided three valid responses to the three anchoring vignettes items that were assigned to the teacher.

Scale scores were provided for the five teacher questionnaire contextual indices, identified by the prefix 'TC'.

Principal data

The school principal data file contains principal ID (Username_PR), other ID variables for cohort, Site and school, variables relating to the number of students within the school, the percentage of valid responses provided by the principal (Pct_PRQ_complete) and if it has reached 50% (Pct50_PRQ), the principal responses to the contextual questionnaire items and the principal weight (WT2019_PR).

Scale scores were provided for the three principal questionnaire contextual indices, identified by the prefix 'PR'.

15.2.2. Variables used for link data files and derived variables

There were a number of different ID variables in each of the data files such as the Site ID, school ID, cohort ID or participant ID. All five data files can be merged through different ID variables. The student data, teacher assessment data and parent data can be linked through the student ID variable UserName_STD which consists of 9 digits. UserName_STD is the combination of cohort indicator (1=Younger cohort and 2=Older cohort), Site ID (2-digit), school ID (3-digit) and student ID (3-digit). The two teacher data files can be linked using the teacher ID variable, UserName_TC. The school principal data file can be linked to other data files using the cohort ID, Site ID and school ID.

15.2.3. Variable Naming

The naming of variables followed a convention to ensure that across different groups of respondents variables would have a prefix or suffix that represents the corresponding respondent group. The prefix (or suffix) conventions include the use of STA for the student related variables and student assessment items, STQ for the contextual student questionnaire items, PAA (PA) for the parent-related contextual variables and parent assessment items, TCA (TC) for the teacher-related variables and teacher assessment items, TCQ for the contextual teacher questionnaire items; and PrQ for the contextual school-related variables.

Notes

[1] Sintra's data was excluded from the scaling, but a weight is provided for their valid students.

[2] < Scale > represents a 3-letter scale abbreviation.

[3] In addition, a third set of scores, < Scale >_WLE_8Item, were added into the Site data, which was delivered to each Site. These are the WLE in logits based on an 8-item per scale as required by OECD.

Annex A. Technical standards

Purpose of document

This document lists the set of standards for the Survey on Social and Emotional Skills (SSES) data collection activities for each participating city, region or site (referred as 'site' in this document).

The standards for data collection and submission were developed with three major, and inter-related, goals in mind: (1) consistency, (2) precision and (3) generalisability of the data. Furthermore, the standards ensure a timely progression of the project in general.

- **Consistency:** Data should be collected in an equivalent fashion across all sites, using equivalent assessment materials so that a comparable sample of the student population performs under assessment conditions that are as similar as possible. Given consistent data collection (and sufficiently high response rates), assessment results are comparable across sites. The assessment results in different sites will reflect differences in the performance of the students measured, and will not be caused by factors that are un-related to performance.
- **Precision:** Data collection and submission practices should leave as little room as possible for spurious variation or error. This holds for both systematic and random error sources, e.g. when the study administration environment differs from one group of students to another, or when sampling procedures leave room for interpretation. An increase in precision relates directly to the quality of results one can expect: The more precise the data, the more powerful the (statistical) analyses, and the more trustworthy the results to be obtained.
- **Generalisability:** Data are collected from specific individuals, in a specific situation, and at a certain point in time. Individuals to be assessed, assessment materials and tasks etc. should be selected in a way that will ensure that the conclusions reached from a given set of data do not simply reflect the setting in which the data were collected but hold for a variety of settings and are valid in the target population at large. Thus, collecting data from a representative sample of the population, for example, will lead to results that accurately reflect the level of social and emotional skills of fifteen-year-old students in a site.
- **Timeliness**: Consistency, precision and generalisability of the data can be obtained in a variety of ways. However, the tight timelines and budgets in SSES do not allow for developing and monitoring local solutions, and harmonizing them at a later stage in the project. Therefore, the standards specify one clear-cut path along which data collection and data submission should progress.

This document establishes a collective agreement of mutual accountability among sites, and of the International Contractor towards the sites. This document details each standard, its rationale, and the quality assurance data that need to be collected to demonstrate that the standard has been met.

Where standards have been fully met, data will be recommended for inclusion in the international SSES dataset. Where standards have not been fully met, an adjudication process will determine the extent to which the quality and international comparability of the data have been affected. The result of data adjudication will determine whether the data will be recommended for inclusion in the international SSES dataset.

Since attaining the various standards is cumulative and potentially co-dependent (i.e. not attaining standard X can affect standards Y and Z), in principle each dataset should be evaluated against all standards jointly. In addition, it is possible that sites' proposed plans for implementation are not, for various and often unforeseen circumstances, actually implemented (e.g. national teacher strike affecting not only response rates but also the assessment conditions; unforeseen site budget cuts which have an impact on data management quality). Therefore, the final evaluation of standards needs to be made with respect to the data as submitted since this is the definitive indication of what may appear in the released international dataset.

If any issues with attaining standards are identified, the International Project Director will initiate communication with the site as soon as possible. The priority in the communication is to rectify the identified issues.

The SSES standards act as a benchmark of best practice. As such, the standards are designed to assist sites and the International Contractor by explicitly indicating the expectations of data quality and study implementation, and by clarifying the timeliness of the activities involved. The standards outline levels of attainment, while timelines and feedback schedules of both the participating sites and the International Contractor are defined in the SSES operations manuals.

Where the technical standards stipulate that variations from the standards require agreement between participating sites and the International Contractor, Site Project Managers are asked to initiate the process of negotiation and to undertake everything possible to facilitate an agreement. Where agreement between Site Project Managers and the Consortium cannot be reached, the OECD will adjudicate and resolve the issues. The OECD will also adjudicate any issues resulting from non-compliance with the technical standards that cannot be resolved between participating sites and the International Contractor.

There are two types of standards in this document, each with a specific purpose:

- Data quality standards refer to aspects of study implementation that directly concern the quality of the data or the assurance of that quality.
- Management standards are in place to ensure that all SSES operational objectives are met in a timely and coordinated manner.

Format of the document

The standards are grouped into sections that relate to specific tasks in the SSES data collection process. For every section, a rationale is given explaining why standard setting is necessary. The standards in each section consist of three distinct elements. First, there are the standards themselves that are numbered and are shown in shaded boxes. Second, there are notes that provide additional information on the standards directly. The notes are listed after the standards in each section where appropriate. Third, there are the quality assurance measures that will be used to assess if a standard has been met or not. These are listed at the end of each section. In addition, the standards contain words that have a defined meaning in the context of the standards. These words are clarified in the definitions section at the end of the document, where the terms are listed alphabetically.

Data quality standards

1. Target population and sampling

Rationale: Meeting the standards specified in this section will ensure that in all sites, the assessed students come from the same target population, and are in a nearly equivalent age range. Therefore, the results obtained will not be confounded by potential age effects. Furthermore, to be able to draw conclusions that are valid for the entire population of ten-year-old students and of fifteen-year-old students, a representative sample shall be selected for participation in the assessment for each age group. The size of the representative sample should not be too small, in order to achieve a certain precision of measurement in all sites. For this reason, minimum numbers of participating students and schools are specified.

Standard 1.1 The SSES Desired Target Population is agreed upon through negotiation between the Site Project Manager and the International Contractor within the constraints imposed by the definition of the SSES Target Population (see "Definitions" section).

Standard 1.2 Unless otherwise agreed upon only SSES-eligible students participate in the assessment.

Standard 1.3 Schools are sampled using agreed upon, established and professionally recognised principles of scientific sampling.

The sampling design in the SSES is a two-stage stratified sample design. The first-stage sampling units consist of individual schools having 10- or 15-year-old students.

Schools are sampled systematically from a school sampling frame, with probabilities that are proportional to a measure of size. The measure of size is a function of the estimated number of SSES-eligible 10- and 15-year-old students enrolled in the school. This is referred to as systematic Probability Proportional to Size (PPS) sampling. Sampling procedures for both cohorts are identical

Standard 1.4 Students are sampled using agreed upon, established and professionally recognised principles of scientific sampling and in a way that represents the full population of SSES-eligible students.

The second-stage sampling units consist of students belonging to the schools selected in the first-stage sampling. The second stage sampling is conducted in cases where number of students in selected schools belonging to the target populations exceeds the maximum cluster size. In all other cases, all students in the selected schools that belong to the target populations are selected in the sample

Standard 1.5 The SSES Defined Target Population covers 95% or more of the SSES Desired Target Population. That is, school-level exclusions and within-school exclusions combined do not exceed 5%.

Standard 1.6 The student sample size is a minimum of 3,000 assessed students in each cohort for SSES participants, or the entire SSES Defined Target Population where the SSES Defined Target Population is below 3,000.

Standard 1.7 The school sample size is a minimum of 75 schools or, in sites where the number of schools with students in the SSES Defined Target Population is below 75, all schools that have students in the SSES Defined Target Population.

Standard 1.8 The school response rate is at least 85% of sampled schools. If a response rate is below 85% then an acceptable response rate can still be achieved through agreed upon use of replacement schools**.**

Standard 1.9 The student response rate is at least 80% of all sampled students across responding schools.

The teacher response rate is at least 80% of all nominated teachers across responding schools.

Note 1.1 The target population and sampling standards apply to the Main Study but not the Field Test.

Note 1.2 In cases where minority language populations that are not offered their language versions exceed 5% of the target population, an agreement can be arranged to regard these languages as reduced population coverage.

Note 1.3 A student is regarded as a participant if they have responded to at least 50% of the assessment items.

Note 1.4 Data from schools where the student response rate is greater than 25% will be included in the SSES dataset.

Note 1.5 For the purpose of calculating school response rates, a participating school is defined as a sampled school in which more than 50 % of sampled students respond.

Note 1.6 Guidelines for acceptable exclusions that do not affect standard adherence are as follows:

- School level exclusions that are due to geographical inaccessibility, or where administration of SSES would be not feasible within the school and that total to less than 0.5% of the SSES Desired Target Population
- School level exclusions that are due to a school containing only students that would be within-school exclusions and that total to less than 2.0% of the SSES Desired Target Population
- Within-school exclusions that total to less than 2.5% of the SSES Desired Target Population – these exclusions could include, for example, students not able to do the assessment because of a functional disability.

Note 1.7 Principles of scientific sampling include, but are not limited to:

- The identification of appropriate stratification variables to reduce sampling variance and facilitate the computation of non-response adjustments.
- The incorporation of a target cluster size of 50 SSES-eligible students which upon agreement can be increased, or reduced to a number not less than 20.

Note 1.8 As a measure aimed at reducing the effects of non-response bias, the International Contractor will identify potential schools as substitutes for non-cooperating, originally sampled schools. cities may replace non-cooperating schools with a substitute chosen by the International Contractor. The International Contractor will assign substitute schools at the same time it selects the main sample schools. Each sampled school is assigned up to two designated substitute schools. Only non-responding schools may be substituted; out-of-scope schools will not be substituted. The use of substitutes for sampled non-responding schools diminishes the quality of the sample and must only be used when necessary. The International Contractor will weight substitute schools as if they had been selected in the main sample.

Note 1.9 In the event data from a site fall short of the sampling metrics, the International Contractor may request additional information from the site to determine whether the non-response reflects a systematic regularity that would compromise the credibility of that site's data. The International Contractor, the site and OECD will discuss the case and whether the data, despite missing the metrics, are nonetheless acceptable, and how the international report and dataset will treat the site's data. This review will take into account the reason a student is not assessed as recorded on the date of the assessment as well as the other questionnaires (teacher, principal and parent) and the translation and sampling processes.

Quality Assurance

- Sampling procedures as specified in the SSES operations manuals
- School sample drawn by International Contractor

- Student sample drawn through provided Sampling Tool
- Sampling forms submitted to the International Contractor
- Field Test Review Quality Assurance Survey

2. Language of the assessment

Rationale: Using the language of instruction will ensure analogous assessment conditions for all students within a site, thereby strengthening the consistency of the data. It is assumed that the students assessed have reached a level of understanding in the language of instruction that is sufficient to be able to work on the SSES assessment without encountering linguistic problems. Thus, the level of social and emotional skills can be assessed without interference due to a critical variation in language proficiency.

Standard 2.1 The SSES assessment is administered to a sampled student in a language of instruction provided by the sampled school.

- Agreement with the International Contractor will be subject to the principle that the language options provided should be languages that are common in the community and are common languages of instruction in schools in that site.

3. Field Test participation

Rationale: The Field Test gives sites the opportunity to try out the logistics of their assessment procedures and allows the International Contractor to make detailed analyses of the items so that only suitable ones are included in the Main Study.

Standard 3.1 All SSES sites participating in the SSES Main Study will have successfully implemented the Field Test.

- A Field Test should occur in an assessment language if that language group represents more than 10% of the target population.
- The school sample size for the Field Test is a minimum of 15 schools.
- The minimum student sample size for the Field Test is 500 students.
- For languages that apply to more than 10% of the target population, the Field Test student sample should be a minimum of 200 students per item.

Note 3.1 Sites will use the same schools for the Field Test as the Main Study.

4. Adaptation of assessments, contextual questionnaires and manuals

Rationale: In order to compare findings across sites, equivalent content needs to be delivered. If the nature of the content differs, then it is unclear whether response differences across sites reflect real variation or whether they just mirror disparities in the content across sites. Therefore, to validly compare response differences across sites, all assessment and contextual questionnaires have to be made as equivalent as possible. This should also be reflected in an equivalent set of instructions given to respondents and also equivalent procedures of data-collection where possible. To achieve this goal, other individuals who play a key role in the data-collection process, i.e. the study administrators and school coordinators, should receive the same information in all participating sites.

Standard 4.1 All student assessment instruments are conceptually equivalent to the source versions. Agreed-upon adaptations to the local context are made if needed.

Standard 4.2 The contextual questionnaire instruments are equivalent to the source versions. Agreed upon adaptations to the local context are made if needed.

Standard 4.3 Participating sites are allowed to add site specific content to the instruments, provided that

- the International Contractor and the OECD has agreed upon the content
- no more than 10 extra responses are added per instrument
- additional content is added at the end of the instruments.

Quality Assurance

Site Adaptation and Verification Form (SAVF) in which adaptations to items (mainly contextual questionnaire items) are documented and agreed upon. Adaptations will be checked for compliance with the SSES Translation and Adaptation Guidelines by international verifiers, and the verifiers' recommendations will be vetted by the translation referee. This form is also used for recording and approval of site-specific additions to the instruments.

- Final Optical Check Report
- Field Test Review Quality Assurance Surveys
- Psychometric properties of items and scales

5. Translation of assessments, contextual questionnaires and manuals

Rationale: To be able to compare the performance of students across sites, and of students with different instruction languages within a site, the linguistic equivalence of all materials is central. While Standards 4.1 to 4.2 serve to ensure that instruments in the source language (English) are equivalent in all sites involved, in general, the following Standards 5.1 and 5.2 emphasise the importance of correct translation of the original source material into the local languages used in participating sites. Again, the goal is to ensure that the selected social and emotional skills will be assessed equivalently in all sites, and that no bias is introduced by differences in the translation of materials.

Standard 5.1 The following documents are translated into the assessment language in order to be linguistically equivalent to the international source versions:

- All administered student assessment instruments
- All administered contextual questionnaires

Standard 5.2 Each question is independently double translated. After an adjudication process, the final translation is submitted for external verification by the International Contractor.

Standard 5.3 Unless otherwise agreed upon, the following documents are translated into the assessment language to make them linguistically equivalent to the international source versions.

- The Study Administrator Manual
- The School Coordinator Manual

In the case of the manuals, only specified parts undergo the process of double translation and verification. Other parts are only translated once and are not verified.

Note 5.1 The quality assurance requirements for this standard apply to instruments that are in a language that is administered to more than 10% of the target population.

Note 5.2 Sites will apply the 'team approach' when managing the translation process. Each translator will translate 2/3 of the material as well as reviewing the translations of the other two translators. After the site translators have reached consensus on their translation, they will have the site's subject matter expert review the translation and terminology used to ensure that it is conceptually equivalent to the international source version.

Note 5.3 The 'specified parts' of manuals referred to in Standard 5.3 for which checking of the linguistic equivalence to the source versions would be undertaken are the study administration scripts.

Quality Assurance

Translation Guide developed to provide detailed instructions on the translation process, which would require double translation by independent translators from two source versions.

- Review by international translation referee of all verifiers' suggested changes.
- Agreed upon Site Adaptation and Verification Form (SAVF)
- Translation Management Sheet (TMS) in which translation process of the assessment items are documented.
- SoNET translation and verification system in which translations will be checked for compliance with the SSES Translation and Adaptation Guidelines by international verifiers, and the verifiers' recommendations will be vetted by the translation referee.
- Final Optical Check Report
- Field Test Review Quality Assurance Surveys
- Psychometric properties of items and scales

6. Assessment administration

Rationale: Certain variations in the assessment procedure are particularly likely to affect assessment performance. Among them are session timing, the online administration of assessment instruments, the instructions given prior to assessment, the rules for excluding students from the assessment, etc. A full list of relevant assessment conditions is given in the SSES operations manuals. To ensure that the data are collected consistently, and in a comparable fashion, for all participants, it is therefore very important to keep the chain of action in the data-collection process as constant as possible across all SSES participants.

It is of utmost importance to assign the correct login details to the participants specified beforehand. The student tracking form is central in monitoring whether this goal has been achieved.

The study administrator plays a central role in all of these issues. Special consideration is therefore given to the training of the study administrators, ensuring that as little variation in the data as possible is caused by random or systematic variation in the activities of study administrators.

An important part of the assessment situation is the relationship between study administrators and assessment participants. Therefore, any personal interaction between study administrators and students, either in the past or in the assessment situation, counteracts the goal of collecting data in a consistent fashion across sites and participants. Strict objectivity of the study administrator, on the other hand, is instrumental in collecting data that reflect the level of social and emotional skills obtained, and that are not influenced by factors un-related to social and emotional skills. The results based on these data will be representative for the population under consideration.

Standard 6.1 Unless otherwise agreed upon, the assessment period

- is no longer than eight consecutive weeks in duration
- will finish by 30 November 2018 in the Field Test and 30 November 2019 in the Main Study.

Standard 6.2 All assessment sessions follow international procedures as specified in the SSES operations manuals, particularly the procedures that are:

- relating to assessment session timing
- for maintaining assessment conditions
- for student tracking
- for assigning online login details for the assessment.

Standard 6.3 Study administrators are trained in person unless a suitable alternative is agreed upon.

Standard 6.4 The relationship between study administrators and participating students must not compromise the credibility of the assessment session. In particular, the study administrator should not be the instructor of any student in the assessment sessions he or she will administer for SSES.

Note 6.1 Principals, teachers and parents are allowed two additional weeks to submit their responses. Note 6.2 Study administrators should preferably not be school staff.

Note 6.3 Preferred training procedures for study administrators are described in the SSES operations manuals.

Quality Assurance

- Study administrator's Assessment Session Report Forms
- SSES quality monitors
- Field Test Review Quality Assurance Survey

7. Security of the material

Rationale: The goal of the SSES assessment is to measure the social and emotional skills of the students. Prior familiarisation with the assessment materials, or training to the assessment, may affect the comparability of the data. In order to be able to assess the social and emotional skills of the students rather than short-term learning success, and to make valid international comparisons, confidentiality is very important.

Standard 7.1 SSES materials designated as secure are kept confidential at all times. Secure materials include all assessment materials, data, and draft materials. In particular:

- no one other than approved project staff and participating students during the assessment session is able to access and view the assessment material online or in paper format
- no one other than approved project staff will have access to secure SSES data and embargoed material
- formal confidentiality arrangements will be in place for all approved project staff.

Quality Assurance

- Security arrangements as specified in the SSES operations manuals
- Signed Confidentiality Agreement
- Site quality monitoring
- Field Test Review Quality Assurance Surveys.

8. Quality monitoring

Rationale: To obtain valid results from the assessment, the data collected have to be of high quality, i.e. they have to be collected in a consistent, reliable and valid fashion. This goal is implemented first and foremost by the study administrators, who are seconded by the quality monitors. The quality monitors provide site-wide supervision of all data-collection activities.

Standard 8.1 Trained independent quality monitors visit sites to monitor SSES study administration.

Standard 8.2 Each SSES participating site agrees on a number of site visits to observe study administration sessions. There will be 5 quality monitor visits in Field Test and 10 quality monitor visits in the Main Study.

Standard 8.3 Study administration sessions are randomly selected for site visits.

Note 8.1 A failure to meet the quality monitoring standards in the Main Study will lead to a significant lack of quality assurance data for other standards.

Note 8.2 The quality monitoring standards apply to both the Field Test and the Main Study.

Note 8.3 Implementing the quality monitoring process is the responsibility of the International Contractor. The
quality monitors are independent of the site project teams in all aspects.

Note 8.4 The site provides the International Contractor the assistance required to implement the site visits effectively.

Quality Assurance

- Curricula Vitae of the SSES quality monitor nominees forwarded by the Site Project Manager to the International Contractor
- SSES Quality Monitor Reports
- Site Quality Monitor Report

9. Occupational data response coding

Rationale: To ensure the comparability of the data, open-ended occupational response data from participants across sites have to be coded following one single coding scheme. Therefore, all coding procedures have to be standardised, and coders have to follow training materials to master this task.

Standard 9.1 Sites should code all occupational responses according to the instructions specified in coding scheme by the International Contractor.

Standard 9.2 Coders are recruited and trained following agreed procedures.

Note 9.1 Preferred procedures for recruiting and training coders are outlined in the SSES operations manuals

Quality Assurance

- Field Test and Main Study Review Quality Assurance Surveys

10. Data submission

Rationale: The timely progression of the project, within the tight timelines given, depends on the quick and efficient submission of all collected data. Therefore, one single data submission format is proposed for each form or instrument. Data collected offline will need to be entered into the tools provided by the International Contractor.

Standard 10.1 Each SSES participating site submits its assessment and contextual questionnaire data via the SoNET assessment platform, including data collected by phone or through paper instruments, unless otherwise agreed upon.

Standard 10.2 All other study data are submitted in the format or tool provided by the International Contractor and described in the SSES operations manuals.

Standard 10.3 Data for all instruments are submitted. This includes the assessment data, questionnaire data, and tracking data as described in the SSES operations manuals.

Standard 10.4 Unless agreed upon, all data are submitted without recoding any of the original response variables.

11. Assessment Delivery Mode

Rationale: To ensure comparability across all participating sites, the assessment delivery mode has to be consistent, i.e. participants have to complete the instruments in a same fashion. The primary mode of the assessment delivery is online using devices that meet the minimal requirements of the online delivery platform.

Standard 11.1 All SSES participants complete the instruments online using available devices that meet the minimal requirements, unless otherwise agreed upon.

Standard 11.2 Where administration is delivered in paper format, the site is responsible for entering all response data in the SoNET delivery platform using participants' logins.

Note 11.1 Paper format of the assessment instruments should be prepared as a back-up option, and only be used when required due to the failure of online mode or the preference of the respondents.

12. Security of personal data

Rationale: The International Contractor being a public company incorporated in Australia is committed to protecting all of the personal data that it collects, processes, analyses and reports on both in digital and paper-based formats. We strive to ensure that we are following best practice and in compliance with relevant legislation and regulations in the countries in which International Contractor conducts its work.

In pursuit of the purpose of the SSES project, International Contractor will be required to process pseudonymised data received from participating sites. Pseudonymisation is the processing of personal data in such a manner that they can no longer be attributed to a specific data subject without the use of additional information. This assumes that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

Pseudonymised data should be of equal quality to data that can be attributed to identified or identifiable natural persons (un-pseudonymised data) with respect to data analysis and reporting.

Standard 12.1 International Contractor, in its role as data processor, will receive pseudonymised personal data from participating sites for the purpose of preparing for, analysing and reporting on the performance of the student populations undertaking the SSES assessment in accordance with regional or national regulations (such as GDPR).

Standard 12.2 Any disclosure of personal data to International Contractor by participating sites will occur in the format International Contractor directs.

Standard 12.3 In collecting, handling and transferring personal data to International Contractor, participating sites will have the responsibility and cost of complying with the relevant privacy legislation in their own countries.

Standard 12.4 Any changes made to the data in the process of pseudonymisation should not affect the quality of the data and the scope of analysis and reporting plans, unless otherwise agreed upon.

Note 12.1 Unique IDs assigned by the International Contractor link students to schools, teachers and parents. In case a site is required to assign new IDs, the links need to remain intact and IDs need to be unique.

Note 12.2 More information about security of personal data is in the SSES Data Security Statement.

Management standards

13. Communication with the International Contractor

Rationale: Given the tight schedule of the project, delays in communication between the sites and the International Contractor should be minimised. Therefore, sites need continuous access to the resources provided by the International Contractor.

Standard 13.1 The International Contractor ensures that qualified staff are available to respond to requests by the sites during all stages of the project. The qualified staff:

- are authorised to respond to site queries
- acknowledge receipt of site queries within one working day
- respond to queries from participating sites within three working days, or, if processing the query takes longer, give an indication of the amount of time required to respond to the query.

Standard 13.2 All communication between sites and the International Contractor will be in English.

Note 13.1 Response timelines and feedback schedules for the sites and the International Contractor are further specified in the SSES communication protocols.

14. Schedule for submission of materials

Rationale: To meet the requirements of the work programme, and to progress according to the timelines of the project, the International Contractor will need to receive a number of materials on time.

Standard 14.1 The following items are submitted to the International Contractor in accordance with the timeline:

- Sampling forms (see Standard 1)
- Site Adaptation and Verification Form (SAVF)
- Field Test Reviews
- Other documents as specified in the SSES operations manuals.

Standard 14.2 Assessments and contextual questionnaires are submitted for linguistic verification only after all adaptations have been agreed upon.

Standard 14.3 The 'administration scripts' of the Study Administrator manual requiring verification, including linguistic verification as specified in Standard 5.2, are submitted only after all adaptations have been agreed upon.

Quality Assurance

- International Contractor records
- Assessment and contextual questionnaire materials are submitted for linguistic verification in the SoNET system with corresponding SAVF filled in by the site

15. Drawing samples

Rationale: The mode of drawing the samples used in the study is crucial to data quality. The goal of the project is to collect data that are representative for the population at large. To reach this goal, the sampling procedures have to follow established scientific rules. Furthermore, the comparability of the data across sites is guaranteed if the same procedure is used for all national samples. If different sampling procedures are used, then the equivalence of the sampling quality has to be determined.

Standard 15.1 For efficient and effective quality assurance provision the International Contractor will draw the school sample for the Field Test and the Main Study.

Standard 15.2 For efficient and effective quality assurance provision, the site will use the Sampling Tool provided to draw the student sample using the list of eligible students provided for each school.

16. Management of data

Rationale: Consolidating and merging the sites' databases is a time-consuming and difficult task. To ensure the timely and efficient progress of the project, the International Contractor needs continuous access to national resources helping to rule out uncertainties and to resolve discrepancies. This standard aims to prevent substantial delays to the whole project which could result from a delay in processing the data of a small number of participating sites.

Standard 16.1 The timeline for submission of national databases to the International Contractor is within six weeks of the last day of student assessments for the Field Test and within eight weeks of the last day of student assessments for the Main Study, unless otherwise agreed upon.

Standard 16.2 Sites make a data manager available upon submission of the Main Study database. The data manager:

- is authorised to respond to International Contractor data queries
- is available for a three-month period immediately after the database is submitted unless otherwise agreed upon
- is able to respond to International Contractor queries within three working days
- is able to resolve data discrepancies.

Note 16.1 Each participating site has access to and can publish its own data after a date that is established for the publication of the initial OECD publication of the survey results.

Note 16.2 The OECD Secretariat will not release a site's data to other sites until participating sites have been given an opportunity to review and comment on their own site data and until the release of such data has been approved by the national authorities.

Note 16.3 The OECD and the International Contractor will decide upon the deadline and procedures for withdrawing a site's data from the international micro-level SSES database (the "international database"). Sites can withdraw data only prior to obtaining access to data from other sites. Withdrawn data will not be made available to other sites.

Note 16.4 The OECD Secretariat will discuss with participating sites whose data manifests technical anomalies as to whether the data concerned can be included in the international database. The decision of the OECD Secretariat will be final. However, participating sites may continue to use data that are excluded from the international database at the national level.

Note 16.5 The International Contractor will then compile the international database, which will comprise the complete set of national SSES databases, except those data elements that have been withdrawn by participating sites or by the OECD Secretariat at the previous stage. The international database will remain confidential until the date on which the initial international OECD publication is released.

Note 16.6 National data from all participating sites represented in the international database will be made available to all participating sites from the date on which the initial international OECD publication is released.

Note 16.7 After release of the initial international OECD publication, the international database will be made publicly available on a cost-free basis, through the OECD Secretariat. The database may not be offered for sale. The international database will form the basis for OECD reports and publications.

Note 16.8 The International Contractor will have no ownership of instruments or data and will be subject to the confidentiality terms set in the agreement.

Note 16.9 The OECD establishes rules to ensure adherence to the above procedure and to the continued confidentiality of the SSES data and materials until the agreed release dates. These include confidentiality agreements with all individuals that have access to the SSES material prior to its release.

Note 16.10 As guardian of the process and producer of the international database, the OECD will hold copyright in the database and in all original material used to develop, or be included in, the SSES Field Test and SSES Main Study (among them the assessment materials, field manuals, and coding guides) in any language and format.

Quality Assurance

- International Contractor Records

17. Archiving of materials

Rationale: During the entire cycle of the study, the Site centres will maintain an archive of all the assessment materials and field manuals electronically and in hard copies (if applicable).

Standard 17.1 Unless otherwise requested, Site Centres will archive all Field Test materials until the beginning of the Main Survey, and all Main Survey materials until the publication of the international report. Materials to be archived include:

- all respondents' test booklets and questionnaires (in paper format, (if applicable),
- sampling forms,
- student lists,
- student tracking instruments, and
- all data submitted to the International Contractor.

Definitions

Site – a site, geographic region, or similarly defined population, for which the International Contractor fully implements quality assurance and quality control mechanisms and endorses, or otherwise, the publication of separate SSES results.

Agreed procedures – procedures that are specified in the SSES operations manuals, or variations that are agreed upon between the Site Project Manager and the International Contractor.

Agreed-upon – variations and definitions agreed upon between the Site Project Manager and the International Contractor. Agreed-upon variations will be available to Site Project Managers on their site folder on the International Contractor Project Portal.

International Contractor portal –

https://collaboration.acer.edu.au/projects/SSES. This portal contains the source versions of instruments, manuals and other documents and information relating to sites.

SSES defined target population – all SSES-eligible students in the schools that are listed on the school sampling frame. That is, the SSES Desired Target Population minus exclusions.

SSES desired target population - the SSES Target Population defined for a specific site. It provides the most exhaustive coverage of SSES-eligible students in the site as is feasible.

SSES-eligible Students – students who are in the SSES Target Population.

SSES operations manuals – manuals provided by the International contractor, that is the following:

- Site Project Manager's Manual,
- Study Administrator Manual,
- School Coordinator Manual,
- Sampling Guidelines, and
- all other key documents referenced within the Site Project Manager's manual.

The preparation of the SSES operations manuals will be carried out by the International contractor and will describe procedures developed by the International Contractor. The manuals will be prepared following consultation with the OECD Secretariat.

SSES participant - an administration centre, commonly called a site that is managed by a person, commonly called a Site Project Manager, who is responsible for administering SSES in a site.

SSES quality monitor – a person nominated by the Site Project Manager and employed by the International Contractor to monitor assessment administration quality in a site.

SSES target population –

Older cohort: students aged between 15 years and 3 (completed) months and 16 years and 2 (completed) months at the beginning of the testing period, attending educational institutions located within the site boundaries, and in grade 7 or higher. The age range of the population may vary up to one month, either older or younger, but the age range must remain 12 months in length. That is, the population can be as young as between 15 years and 2 (completed) months and 16 years and 1 (completed) month at the beginning of the testing period; or as old as between 15 years and 4 (completed) months and 16 years and 3 (completed) months at the beginning of the testing period.

Younger cohort: students aged between 10 years and 3 (completed) months and 11 years and 2 (completed) months at the beginning of the testing period, attending educational institutions located within the site boundaries, and in grade 2 or higher. The age range of the population may vary up to one month, either older or younger, but the age range must remain 12 months in length. That is, the population can be as young as between 10 years and 2 (completed) months and 11 years and 1 (completed) month at the beginning of the testing period; or as old as between 10 years and 4 (completed) months and 11 years and 3 (completed) months at the beginning of the testing period.

School level exclusions – exclusion of schools from the sampling frame because:

- of geographical inaccessibility (but not part of a region that is omitted from the SSES Desired Target Population)
- of an extremely small size
- administration of the SSES assessment within the school would not be feasible
- all students in the school would be within-school exclusions, or
- of other reasons as agreed upon.

Source versions – documents provided in English by the International Contractor.

Target cluster size - the number of students that are to be sampled from schools where not all students are to be included in the sample.

Assessment period – the period during which data are collected in a site.

Within-school exclusions – exclusion of students from potential assessment because of one of the following:

- They are functionally disabled in such a way that they cannot take the SSES assessment. Functionally disabled students are those with a moderate to severe permanent physical disability.
- They have a cognitive, behavioural or emotional disability confirmed by qualified staff, meaning they cannot take the SSES assessment. These students are cognitively, behaviourally or emotionally unable to follow even the general instructions of the assessment.
- They have insufficient assessment language experience to take the SSES assessment. Students who have insufficient assessment language experience are those who have received less than one year of instruction in the assessment language.
- They cannot be assessed for some other reason as agreed upon.

Annex B. Item Pool Classification

Information on item pool classification can be found in the excel document called “Item Pool Classification SSES” on the SSES website: https://www.oecd.org/education/ceri/social-emotional-skills-study/

Annex C. Standard errors of the student performance mean estimate

Table C.1. Standard errors of the student performance mean estimate, by city and skill

All students

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	2.07	2.02	1.34	1.58	1.25	1.58	1.66	1.87	1.76	1.43	1.78	1.98	1.44	1.82	1.70	1.31	1.70
Daegu	1.40	1.71	1.69	1.62	1.70	1.86	1.58	1.70	1.60	1.56	1.31	1.89	1.65	1.30	1.29	1.84	1.63
Helsinki	2.23	2.09	1.57	2.14	1.39	1.92	1.56	1.75	1.95	1.62	1.59	1.75	1.79	1.59	2.08	1.56	1.93
Houston	1.76	1.69	1.58	1.46	1.50	1.74	1.72	1.62	1.72	1.53	1.55	1.55	1.51	1.25	1.91	1.57	1.83
Manizales	1.31	2.50	2.03	1.57	1.36	1.57	1.95	2.64	2.54	2.54	1.66	1.77	1.52	2.35	1.55	2.13	2.82
Moscow	1.29	0.93	1.16	1.17	1.30	0.96	1.27	1.09	1.42	1.27	1.43	1.16	0.96	1.29	1.17	1.12	1.16
Ottawa	1.46	1.19	1.30	1.53	1.41	1.22	1.40	1.18	1.31	1.07	1.18	1.26	1.43	1.44	1.31	1.08	1.23
Sintra	2.04	2.39	1.62	2.10	2.55	1.17	1.55	1.89	2.39	2.05	2.10	1.25	1.56	1.53	1.98	1.44	1.96
Suzhou	1.48	2.57	2.09	2.18	2.72	2.53	2.11	2.08	2.37	2.58	2.34	2.07	2.41	2.22	2.44	2.44	2.32
Istanbul	2.19	4.17	3.49	4.22	3.33	2.88	3.02	3.24	2.37	2.44	2.74	3.37	3.35	1.68	1.75	2.50	2.54

Younger Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	2.62	3.55	1.93	2.46	1.93	2.49	2.08	2.64	2.54	2.48	2.81	2.74	1.76	2.40	2.58	2.16	2.62
Daegu	2.19	2.62	3.02	2.58	2.99	3.15	2.76	3.23	2.79	2.30	2.69	3.15	2.45	2.37	2.39	3.20	2.99
Helsinki	2.86	2.72	2.01	2.11	1.86	2.60	1.95	2.29	2.66	2.26	2.16	2.31	1.95	2.06	2.66	1.97	2.56
Houston	2.34	2.08	2.17	1.88	2.14	2.34	1.96	1.78	2.16	2.15	1.99	2.00	2.02	1.65	2.21	2.07	2.11
Manizales	2.16	3.86	4.05	3.10	1.84	2.83	2.91	3.31	3.09	3.06	2.30	2.74	1.70	3.42	1.71	3.43	3.71
Moscow	1.70	1.44	1.91	1.80	1.90	1.50	1.66	1.67	2.26	1.94	2.07	1.84	1.35	1.75	1.79	1.64	1.80
Ottawa	1.81	1.84	1.92	2.30	2.12	2.11	1.90	1.75	1.96	1.85	1.91	1.75	1.66	1.97	1.65	1.98	1.73
Sintra	1.36	3.50	3.15	2.78	3.11	2.13	2.45	3.91	3.54	3.53	2.73	2.75	2.08	1.93	2.55	2.77	3.35
Suzhou	2.11	3.38	2.66	2.64	3.65	3.32	2.78	2.67	3.10	3.47	3.18	2.71	3.30	2.95	3.43	3.21	2.98
Istanbul	3.05	5.39	4.73	5.37	4.36	4.19	3.38	3.65	2.87	3.17	3.39	4.14	4.49	2.28	2.50	3.29	3.17

Older Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	1.99	1.75	1.92	2.11	1.73	1.72	1.86	1.99	1.95	1.37	1.86	2.02	1.89	2.81	1.35	1.59	1.92
Daegu	1.95	2.20	2.28	2.38	2.40	2.44	2.20	2.02	2.35	2.12	2.14	2.46	2.22	1.71	1.93	2.51	2.18
Helsinki	3.54	2.65	2.37	3.80	2.87	2.49	1.90	1.94	2.98	2.17	2.32	1.89	3.28	2.85	2.17	2.72	2.85
Houston	3.11	2.21	2.42	1.78	1.76	2.64	2.13	1.72	2.09	2.24	2.23	1.82	2.11	1.86	2.35	2.56	2.49
Manizales	1.62	2.01	1.37	1.56	1.86	1.60	1.63	2.29	2.00	2.06	1.53	1.51	1.84	2.79	1.41	1.50	2.27
Moscow	2.07	1.60	1.40	1.86	1.63	1.81	1.64	1.39	1.60	1.57	1.87	1.51	1.66	2.10	1.48	1.48	1.49
Ottawa	2.14	1.89	2.13	2.11	1.86	1.43	2.39	2.08	1.90	1.52	1.74	1.90	2.75	2.23	2.24	1.71	1.75
Sintra	3.82	2.91	1.55	2.47	3.26	1.71	2.40	4.23	3.02	2.32	2.98	2.27	2.25	2.33	4.06	1.51	2.49
Suzhou	1.59	2.31	1.71	1.81	1.45	1.99	1.53	1.80	1.79	1.85	1.81	1.60	1.68	2.46	1.95	2.01	1.63
Istanbul	2.31	2.10	1.99	1.60	2.29	2.19	1.76	1.93	2.45	2.07	2.09	1.84	2.58	2.81	1.77	1.82	2.17

Table C.2. Sample sizes by city and by domain

	All Students			Younger			Older		
	student sample size	School sample size	within-school sample size	student sample size	School size	within-school sample size	student sample size	School size	within-school sample size
Bogota	6771	154	43.97	3415	87	39.25	3356	82	40.93
Daegu	6334	132	47.98	3008	77	39.06	3326	78	42.64
Helsinki	5482	97	56.52	3034	83	36.55	2448	55	44.51
Houston	6434	102	63.08	3333	74	45.04	3101	45	68.91
Manizales	6757	85	79.49	3226	83	38.87	3531	70	50.44
Moscow	6792	78	87.08	3363	77	43.68	3429	77	44.53
Ottawa	5440	123	44.23	3250	89	36.52	2190	58	37.76
Sintra	3860	49	78.78	2224	48	46.33	1636	29	56.41
Suzhou	7246	122	59.39	3633	76	47.80	3613	75	48.17
Istanbul	5869	101	58.11	2701	91	29.68	3168	80	39.60

Table C.3. Intraclass correlation by city and skill

Younger Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	298	540	318	320	116	440	149	293	267	156	265	584	83	306	386	255	264
Daegu	111	324	429	261	282	475	319	387	398	282	290	362	170	256	276	459	417
Helsinki	346	191	196	81	63	170	42	106	134	83	45	131	76	127	171	53	128
Houston	450	191	247	58	257	318	78	68	148	141	197	143	76	170	267	169	154
Manizales	93	357	621	354	44	305	207	250	177	140	65	377	78	334	115	316	285
Moscow	33	0	159	46	51	37	45	17	142	60	35	67	23	67	91	39	65
Ottawa	78	76	174	179	207	76	25	40	84	85	117	79	96	145	131	84	76
Sintra	14	106	146	80	63	23	114	122	96	113	62	159	20	26	97	48	156
Suzhou	332	830	758	581	784	743	532	501	622	813	581	497	609	596	618	848	563
Istanbul	150	616	582	540	291	708	374	322	323	382	352	332	279	132	92	422	424

Older Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	221	199	70	179	100	212	93	61	121	22	84	132	115	468	47	82	117
Daegu	128	211	148	185	329	210	194	205	70	83	100	238	259	122	141	218	108
Helsinki	196	113	121	424	73	80	127	103	92	33	71	39	326	433	28	92	147
Houston	159	131	115	54	114	167	75	13	30	56	29	141	139	161	137	73	58
Manizales	126	142	45	76	72	112	71	111	56	51	13	83	92	344	38	41	62
Moscow	108	46	51	138	82	62	89	3	5	12	59	30	6	145	20	33	15
Ottawa	108	95	167	169	116	24	138	100	11	58	74	5	114	246	32	109	66
Sintra	20	128	2	53	64	61	33	56	87	34	33	44	76	115	66	11	61
Suzhou	63	177	180	261	75	152	97	104	161	164	191	49	99	289	67	194	131
Istanbul	177	97	101	83	151	63	118	113	115	62	85	62	274	454	43	88	108

Table C.4. School variance estimate by city and skill

Younger Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	0.04	0.06	0.03	0.04	0.02	0.06	0.02	0.04	0.03	0.02	0.03	0.06	0.01	0.03	0.05	0.03	0.03
Daegu	0.01	0.04	0.04	0.03	0.02	0.05	0.03	0.03	0.04	0.03	0.03	0.03	0.02	0.03	0.03	0.04	0.04
Helsinki	0.03	0.02	0.03	0.01	0.01	0.03	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02
Houston	0.06	0.02	0.03	0.01	0.03	0.04	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.03	0.02	0.02
Manizales	0.01	0.04	0.06	0.04	0.01	0.04	0.02	0.03	0.02	0.02	0.01	0.04	0.01	0.03	0.02	0.03	0.03
Moscow	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.01
Ottawa	0.01	0.01	0.02	0.02	0.02	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01
Sintra	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.00	0.00	0.01	0.01	0.02
Suzhou	0.05	0.07	0.06	0.06	0.05	0.06	0.05	0.04	0.05	0.06	0.06	0.04	0.05	0.05	0.04	0.07	0.06
Istanbul	0.01	0.05	0.05	0.05	0.03	0.06	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.01	0.01	0.04	0.03

Older Cohort

	Assertive	Cooperation	Creativity	Curiosity	Emotion	Empathy	Energy	Optimism	Persistence	Responsibility	Self Control	Sociability	Stress Resistance	Tolerance	Trust	Efficacy	Motivation
Bogota	0.03	0.03	0.01	0.03	0.01	0.03	0.01	0.01	0.02	0.00	0.01	0.02	0.02	0.04	0.01	0.01	0.02
Daegu	0.01	0.03	0.02	0.02	0.03	0.02	0.02	0.03	0.01	0.01	0.01	0.03	0.03	0.02	0.02	0.02	0.01
Helsinki	0.02	0.02	0.02	0.07	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.03	0.06	0.01	0.02	0.03
Houston	0.02	0.02	0.02	0.01	0.01	0.03	0.01	0.00	0.00	0.01	0.00	0.02	0.01	0.02	0.02	0.01	0.01
Manizales	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.03	0.01	0.01	0.01
Moscow	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.01	0.00
Ottawa	0.01	0.02	0.03	0.02	0.01	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.03	0.01	0.02	0.01
Sintra	0.00	0.02	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.01
Suzhou	0.01	0.03	0.02	0.05	0.01	0.02	0.02	0.01	0.03	0.02	0.03	0.01	0.01	0.03	0.01	0.03	0.03
Istanbul	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.04	0.01	0.01	0.01

Annex D. Mapping of ISCED Levels to Years of Education

Table D.1. ISCED Levels to Years of Education

	Completed ISCED levels						
Site	**1**	**2**	**3**	**4 or 5**	**6**	**7**	**8**
Bogota	5	9	11	13.5	15.5	17.5	19.5
Daegu	6	9	12	14	16	18	19
Helsinki	6	9	12	14	15.5	18.5	19.5
Houston	6	9	12	13.5	16	18	21
Manizales	5	9	11	13.5	15.5	17.5	19.5
Moscow	4	9	11	13	15	16.5	18
Ottawa	6	9	12	14	15.5	17	18.5
Sintra	6	9	12	13.5	15.5	17	20
Suzhou	6	9	12	14	15.5	18.5	18.5
Istanbul	4	8	12	14	16	17.5	20

Note: For the purpose of calculating the years of parental education as part of the socio-economic index, ISCED levels 1, 2, 3, ISCED levels 4, 5 and ISCED levels 6, 7, 8 were collapsed in accordance with the Student Questionnaire data collection.

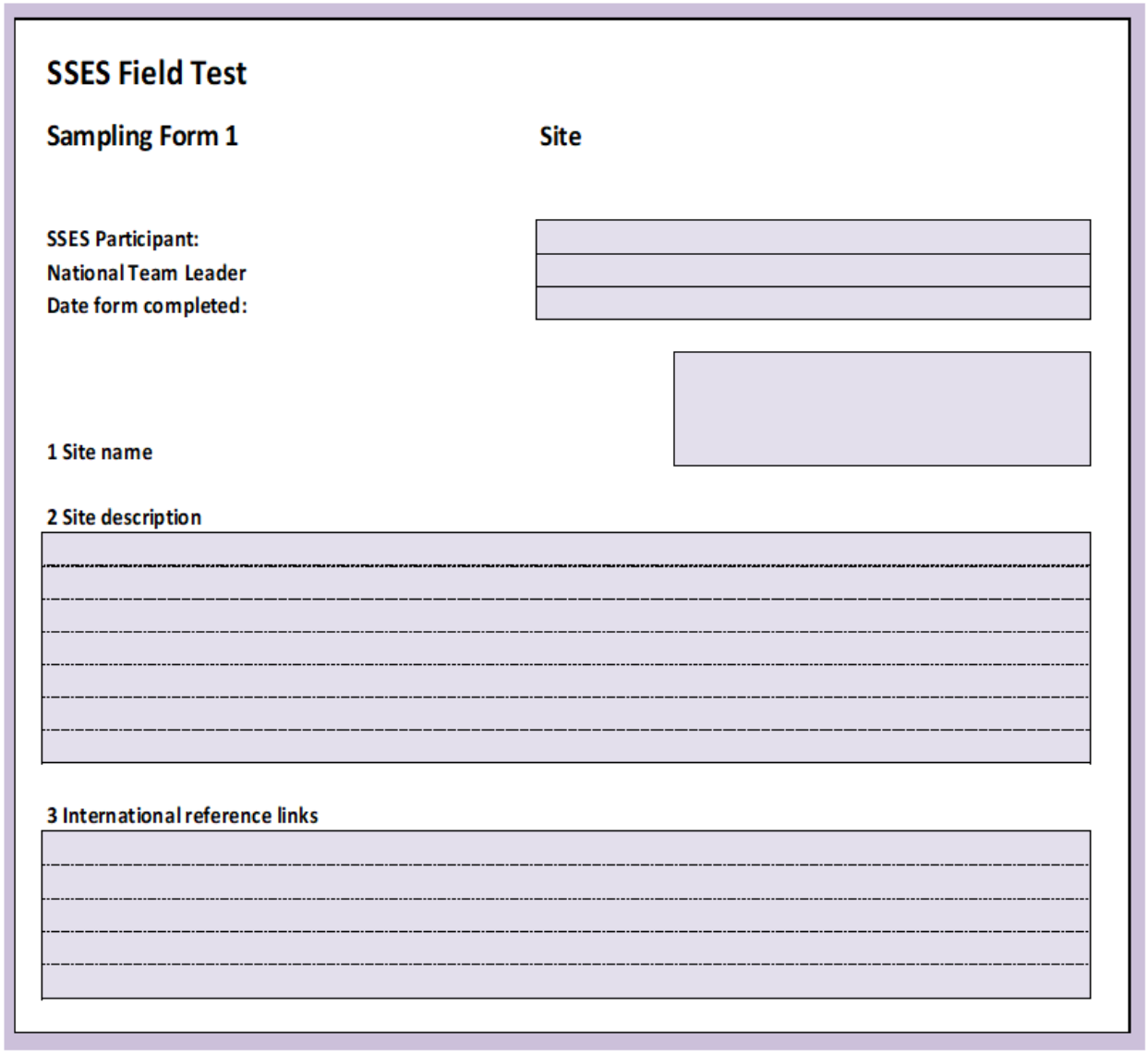
Annex E. National Household Possession Items

Table E.1. National Household Possession Items

Site	STQM00814	STQM00815	STQM00816
Bogota	Digital camera	Encyclopedia	Tablets (for example, iPad®)
Daegu	Air conditioner	Screen and stereo for watching movie	Dishwasher
Helsinki	Laptop	Home alarm system	N/A
Houston	Educational apps for iPad® or tablet	N/A	N/A
Manizales	Digital camera	Encyclopedia	Tablets (for example, iPad®)
Moscow	N/A	N/A	N/A
Ottawa	IPOD®/An MP3 player	A subscription to a daily newspaper	Central air conditioning
Sintra	Cable television or satellite dish	Plasma or LCD TV	Air conditioning
Suzhou	Vacuum cleaner	Digital camera or digital video camera	Juicer
Istanbul	Air conditioning heating-cooling system	Video camera	Home theatre system

Annex F. Structure of target population by site

Figure F.1. Sampling form templates



SSES Field Test

Sampling Form 1 **Site**

SSES Participant:
National Team Leader
Date form completed:

1 Site name

2 Site description

3 International reference links

SSES Field Test

Sampling Form 2

Site Target Population

SSES Participant:
Name:
Date form completed:

1 Please select the expected start date for each assessment period.

Field Test	Main Study
Oct-2018	Oct-2019

2 Please select the age ranges for the field test populations. The default or preferred option is marked with a double asterisk (**). If you choose another option please provide a statement explaining why the one month older or younger option is required.

Older Cohort	Younger Cohort	Statement
15 years, 3 months-16 years, 2 months**	10 years, 3 months-11 years, 2 months**	
Date of Birth of Eligible Students	**Date of Birth of Eligible Students**	
01/07/2002 to 30/06/2003	01/07/2007 to 30/06/2008	

3 Please select the age ranges for the main study populations. The default or preferred option is marked with a double asterisk (**). If you choose another option please provide a statement explaining why the one month older or younger option is required.

Older Cohort	Younger Cohort	Statement
15 years, 3 months-16 years, 2 months**	10 years, 3 months-11 years, 2 months**	
Date of Birth of Eligible Students	**Date of Birth of Eligible Students**	
01/07/2003 to 30/06/2004	01/07/2008 to 30/06/2009	

4 State the name of the grade in your Site Target Population that corresponds to 7 years of schooling counting from the first year of ISCED Level 1. Students within the older cohort age range that are in grades below this level are not eligible for SSES.

5 State the name of the grade in your Site Target Population that corresponds to 2 years of schooling counting from the first year of ISCED Level 1. Students within the younger cohort age range that are in grades below this level are not eligible for SSES.

6 Briefly describe the size of the target population, that is eligible for the SSES Older Cohort.

7 Briefly describe the size of the target population, that is eligible for the SSES Younger Cohort.

8 Include any references or links to documentation used to estimate the eligible site populations.

9 Indicate the site language(s) in which the survey will be administered.

SSES Field Test

Sampling Form 3

Coverage and exclusions

SSES Participant:
Name:
Date form completed:

		Older Cohort	Younger Cohort
Total enrolment in the cohort	[a]		

1 School-level exclusions

Exclusion Category	Description	# of students	# of students
A-9a: Geographically inaccessible			
A-9b: Assessment administration infeasible			
TOTAL		0	0

1a Percentage of school-level exclusions			
A-9c: All students are within-school exclusions			
1b Percentage of school-level exclusions			
2 Total percentage of school-level exclusions	[b]	0.00%	0.00%

3 Estimated within-school student level exclusions (if applicable)

Description of exclusions		# of students	# of students
TOTAL	[c]	0	0

4 Expected percentage of within-school student level exclusions			
5 Expected percentage of reduced coverage and exclusions:	[d]	0.00%	0.00%

SSES Field Test

Sampling Form 4 **Stratification**

SSES Participant:
Name:
Date form completed:

1 List and describe proposed variables to be used for stratification in their order of importance. For each variable list the number of levels.

Stratification Variables						
	Variable name	Description	Value labels	# of levels	Older Cohort	Younger Cohort
					☐	☐
					☐	☐
					☐	☐
					☐	☐
					☐	☐
					☐	☐
					☐	☐
					☐	☐
Additional Stratification Information						

2 If the site has plans to supplement the international sample with additional schools and students for local purposes, describe these plans here.

Table F.1. Participating site population definitions

SSES Participant	Site Name	Site Description
Bogota	Bogotá, Capital District (Colombia)	Bogotá is the capital of Colombia located in the centre of the country, in the eastern mountain range. It has an approximate area of 33 kilometres from south to north and 16 kilometres from east to west and is located at the following coordinates: North Latitude: 4° 35'56" and West Longitude of Greenwich: 74° 04'51"'. In 2017 there were 1858 schools in Bogotá, with attending students in the two stage cohorts for SSES, comprising 694 primary schools, 45 secondary schools, and 1119 combined schools (both primary and secondary). The estimated 10 year-old and 15 year-old student populations are 98,121 and 85,951, respectively.
Daegu	Daegu(Korea)	Daegu, officially known as the Daegu Metropolitan City, is a city located in south-eastern Korea, the fourth-largest city after Seoul, Busan, and Incheon, and the third-largest metropolitan area in the nation with over 2.5 million residents. The city comprises 8 administrative areas. The primary administrative division of Daegu consists of seven 'gu', or districts, and a gun, or county. The 8 administrative divisions and districts of Daegu are referred to as: Suseong-gu, Jung-gu, Dong-gu, Nam-gu, Dalseo-gu, Seo-gu, Buk-gu, and Dalseong-gun. As of 1 April 2018, Daegu had a total of 459 schools including 229 primary schools, 125 middle schools, 93 high schools, and 12 specialised schools. Among these schools are 9 special schools for students with disabilities and 2 schools offering particular programs for students attending general schools were excluded from SSES. (The details are described in sampling form 3). The estimated population of 10 year-old students (equivalent to 4th and 5th grade) is 21,792 and 15 year-old students (equivalent to 9th and 10th grade) is 22,793.
Helsinki	City of Helsinki	Helsinki is the capital of Finland located in Southern Finland. 101 schools are maintained by the City of Helsinki providing for students in the two age cohorts for SSES comprising 55 primary schools, 4 secondary schools, 35 combined schools and 7 special education schools. Only public schools participated in the SSES Field Test. There are also 26 private schools in Helsinki. These are mainly local schools that do not collect tuition fees, so are similar to public schools. The City of Helsinki has no authority over these schools however liaised with them regarding participation in the SSES Main Study. The resident population of 10-14 year-olds was 27,270 on 31 December 2017, and 27,250 for 15-19 year–olds. School participation rates are close to 100% for SSES-aged children.
Houston	Houston Independent School District	Houston Independent School District (HISD) serves the vast majority of students in grades Pre-Kindergarten through Grade 12 in the City of Houston, Texas, USA. The public school district serves approximately 210,000 students at 283 campuses. HISD is the largest school district in Texas and the seventh largest in the United States. Within HISD's boundaries, there are public charter schools with whom HISD has partnered to serve students (included in the sample). Private schools (approximately 15% of the children living in the City of Houston) cannot be included in the sample because they are outside the jurisdiction of HISD; HISD is the participating site through HERC, not the City of Houston. The vast majority of the geographic area of the City of Houston falls inside of HISD's boundaries (illustrated in map provided to the International Contractor).
Manizales	Manizales, Caldas (Colombia)	Manizales is the capital city of the department of Caldas, located in the centre of Colombia. It is a political administrative entity and defined by well recognised city borders. The city has a total population of 400,136 inhabitants (2018). In Colombia, children and young people are considered to be of school age between 5 and 16 years. In Manizales, the estimated resident population of 5 to 16 year-olds is 65,317 (2018), and between 10 and 14 years the estimated population is 27,644. The rate of participation in primary school is 90.13% and 100% for high school. There is a total of 97 schools in the city, 87 of them attended by students in elementary and 73 attended by high school students. It is important to note that most of the city's schools serve both populations. Some of these 97 schools have multiple buildings/ headquarters providing education to students from both cohorts. Manizales has a total of 167 school headquarters serving elementary and high schools students.
Moscow	Moscow	Moscow is the capital of the Russian Federation, located in the central western part of the country. Its administrative borders are defined by the jurisdiction of the Moscow City Department of Education. The total enrolment of students in schools is 1,346,693. According to the available information there are 784 institutions providing education for students of the two SSES-age cohorts. In most cases, one institution is a complex of several formal schools headed by one administration team. A complex usually includes all levels of education: primary school (grades 1-4), secondary school (grades 5-9) and high school (grades 10-11). The average school size is 1930 students, ranging from 25 to 8614 students.
Ottawa	Ottawa, Ontario, Canada	Ottawa is the capital city of Canada, and the political centre of the country. It is located in the province of Ontario and borders the province of Quebec. There are 333 publicly funded schools in Ottawa across four school boards (English Public, English Catholic, French Public, and French Catholic) that serve the two age cohorts: 268 elementary schools and 65 high schools. There are also 20 private schools serving approximately 3600 students in Kindergarten through to Grade 12.

		Together, these schools serve about 153,000 students. For the Field Test, the Site will include the four publicly funded school boards (English Public, English Catholic, French Public, and French Catholic). Together these four school boards account for 97% of students in Ottawa. The private schools will not be included as adding these schools would result in a considerable amount of work and logistical coordination for a very small proportion of eligible students. We acknowledge OECD and the International Contractor's preference to include all students in the geographic area, however, for the reasons above, it is not possible nor practical at this time.
Sintra	Sintra, Portugal	Sintra is a town and a municipality situated 33 km from Lisbon, the capital city of Portugal. In 2016, the Sintra municipality had 383,946 inhabitants, 54,069 (14%) of them aged 6 to 17 years. In the school year 2016/2017, Sintra had 136 primary and secondary schools, 106 (78%) of them public schools. Sintra had a total of 46,427 students enrolled in primary and secondary education, nearly half of them in each level of education (23,761 in primary and 22,666 in lower and upper secondary). The proportion of students enrolled in public schools is approximately 90%.
Suzhou	Suzhou (Jiangsu Province, China)	Suzhou is a city of Jiangsu Province located in the east of the People's Republic of China. It is a political administrative entity and defined by well recognised city borders. In 2017 there were 658 schools in Suzhou providing for students in the two age cohorts for SSES comprising 323 primary schools, 297 high schools (including 53 nine-year-schools and ten 12-year-schools), 26 vocational schools and 12 special education schools. The estimated resident population of 6-11 year-olds was 693,657 in 2016. For 12-19 year-olds it was 394,619. School participation rates in primary and high schools are close to 100% for SSES-aged children.
Istanbul	Istanbul	Istanbul, with a population of 15 million, is the most crowded city in Turkey. With immigration from all parts of the country, Istanbul has become a major metropolitan city. According to the Ministry of National Education's educational statistics report for the 2017/2018 education year, Istanbul has 6864 institutions with 151,326 teachers serving 3,103,439 students. In Istanbul, there are approximately 173,019 students at the older age cohort range and 231,875 students at the younger age cohort range of SSES. 12 years of education is mandatory in Turkey.

Table F.2. Participating site exclusions

Older cohort

SSES Participant	School Level Exclusions	Within-school Student Level Exclusions
Bogota	NA	Cognitive disability
		Functional disability
Daegu	Special schools for students with disabilities	NA
Helsinki	Hospital school	Students with severe handicap, extended compulsory education
	Reformatory homes, not in Helsinki	
	Special education schools	
	All students are within-school exclusions (max. 2%)	
Houston	Geographically inaccessible School (Distance Education)	
	Assessment administration infeasible	
	Schools with fewer than 20 students	
	All students are within-school exclusions	
Manizales	Geographically inaccessible	Physical disability
		Intellectual disability
Moscow	All students are within-school exclusions	NA
Ottawa	Professional School	Students with physical, language or motor disability (special needs)
	All students are within-school exclusions	
Sintra	Schools with a non-Portuguese Curricula	SEN Students with 'individual specific curriculum'
		Students enrolled in 'Portuguese as a second language'
Suzhou	NA	Physical disability
		Intellectual disability
Istanbul	Special education schools	Individuals under temporary protection (Syrian, Afghan etc. refugee children)
	All students within school are excluded (i.e. special schools)	

Younger cohort

SSES Participant	School Level Exclusions	Within-school Student Level Exclusions
Bogota	NA	Cognitive disability
		Functional disability
Daegu	Special schools for students with disabilities	NA
Helsinki	Hospital school	Students with severe handicap, extended compulsory education
	Special education schools	
	All students are within-school exclusions (max. 2%)	
Houston	Geographically inaccessible School (Distance Education)	See School Exclusions in Houston Sampling Frame
	Assessment administration infeasible	
Manizales	Geographically inaccessible	Physical disability
		Intellectual disability
Moscow	All students are within-school exclusions	NA
Ottawa	All students are within-school exclusions	Students with physical, language or motor disability (special needs)
Sintra	Schools with a non-Portuguese Curricula	SEN Students with 'individual specific curriculum'
Suzhou	All students are within-school exclusions	Physical disability
		Intellectual disability
Istanbul	Special education schools	Individuals under temporary protection (Syrian, Afghan etc. refugee children)
	All students within school are excluded (i.e. special schools)	

Table F.3. Participating site stratification variables

Older cohort

SSES Participant	Variable name	Description	Value labels	# of levels
Bogota	School Type	School sector type	Private/ Public / Concession	3
	School Zone	School zone type	Urban/ Rural	2
Daegu	Administrative division	School location	Jung-gu / Dong-gu / Seo-gu / Nam-gu / Buk-gu / Suseong-gu / Dalseo-gu/ Dalseong-gun	8
	School level	School level	Middle school or lower secondary school (Grade 7-9) / High school or upper secondary school (Grade 10-12) / Secondary school (Grade 7-12)	3
	School type	Education purpose	General / Special-Purpose / Vocational / Alternative / Autonomous	5
	Gender composition	Sexual composition of students attending the school	Male and Male / Male and Female/ Female and Female	3
Helsinki	Language	Schools teaching language	Finnish / Swedish	2
	School type	Level of education	Primary/ Secondary/ Combined	3
	S2 student status	Percentage of S2 (Finnish as second language, only Finnish schools) speakers in basic education students in school	<10% (level 1)/ 10-20% (level 2)/ 21-34% (level 3)/ >35%(level 4)	4
Houston	Magnet program	Specialised program at school	Yes/ No	2
	High % Not EcoDis	% Not Economically Disadvantaged	Yes/ No	2
	High % White	% White (9% districtwide)	Yes/ No	2
Manizales	School Administration	School sector type	Public/ Private	2
	School educational model	School program	Active/ Regular	2
	School Location	Geographic location	Rural/ Urban	2
Moscow	Financing	Sources of financing	Private / Government	2
Ottawa	Language	Language of education	English/ French	2
	School type	School type	Public/ Catholic	2
	School Level	Level of education	Primary/ Secondary	2
Sintra	School Type	School administration	Public/ Private	2
Suzhou	School Funding	School sector type	Public / Private	2
	School SES	School socio-economic status of the area the school is in	Low / Medium / High	3
	School Type	Level of education	Primary / Combined / Secondary	3
	Secondary Stream	Type of secondary school	General / Vocational	2
Istanbul	School Type	Type of education	anatolian imam and preacher high school/ anatolian high school/ anatolian vocational high school/ fine arts high school/ science high school/ imam and preacher high school/ imam	11

			and preacher junior high school/ vocational high school/ sports high school/ high school of music and performing arts/ social sciences high school	

Note: Shaded variables correspond to explicit stratification variables.

Younger cohort

SSES Participant	Variable name	Description	Value labels	# of levels
Bogota	School Type	School sector type	Private/ Public/ Concession	3
	School Zone	School zone type	Urban/ Rural	2
Daegu	Administrative division	School location	Jung-gu/ Dong-gu/ Seo-gu/ Nam-gu/ Buk-gu/ Suseong-gu/ Dalseo-gu/ Dalseong-gun	8
	School type	A type of establishment	Public/ Private/ National	3
Helsinki	Language	Schools teaching language	Finnish/ Swedish	2
	School type	Level of education	Primary/ Secondary / Combined	3
	S2 student status	Percentage of S2 (Finnish as second language, only Finnish schools) speakers in basic education students in school	<10% (level 1)/ 10-20% (level 2)/ 21-34% (level 3)/ >35%(level 4)	4
Houston	Magnet program	Specialised program at school	Yes/ No	2
	High % Not EcoDis	% Not Economically Disadvantaged	Yes/ No	2
	High % White	% White (9% districtwide)	Yes/ No	2
Manizales	School Administration	School sector type	Public/ Private	2
	School educational model	School program	Active/ Regular	2
	School Location	Geographic location	Rural/ Urban	2
Moscow	Financing	Sources of financing	Private / Government	2
Ottawa	Language	Language of education	English/ French	2
	School type	School type	Public/ Catholic	2
	School Level	Level of education	Primary/ Middle / Intermediate-Secondary	2
Sintra	School Type	School Administration	Public/ Private	2
Suzhou	School Funding	School sector type	Public / Private	2
	School SES	School socio-economic status of the area the school is in	Low / Medium / High	3
	School Type	Level of education	Primary / Combined / Secondary	3
Istanbul	School Type	Type of education	elementary school/ middle school/ imam and preacher high school/ anatolian imam and preacher high school/ imam and teacher junior high school/ music and ballet lower secondary school	6

Note: Shaded variables correspond to explicit stratification variables.

Table F.4. Field Test sample sizes

Older cohort

SSES Participant	Language	Planned Sample Size	
		Schools	Students
Bogota	Spanish	18	752
Daegu	Korean	17	789
Helsinki	Finnish	12	534
	Swedish	5	240
Houston	English / Spanish	17	808
Manizales	Spanish	20	769
Moscow	Russian	18	799
Ottawa	English	10	500
	French	8	400
Sintra	Portuguese	18	816
Suzhou	Chinese	15	750
Istanbul[1]	Turkish	81	4050

[1] The original plan for Istanbul was to draw a nation-wide sample of all students in Turkey. However, this was deemed impractical for the Main Study and it was limited to Istanbul.

Younger cohort

SSES Participant	Language	Planned Sample Size	
		Schools	Students
Bogota	Spanish	18	733
Daegu	Korean	17	794
Helsinki	Finnish	12	531
	Swedish	7	277
Houston	English/Spanish	17	797
Manizales	Spanish	27	809
Moscow	Russian	18	796
Ottawa	English	14	552
	French	10	412
Sintra	Portuguese	22	790
Suzhou	Chinese	15	750
Istanbul[1]	Turkish	77	3850

[1] The original plan for Istanbul was to draw a nation-wide sample of all students in Turkey. However, this was deemed impractical for the Main Study and it was limited to Istanbul.

Table F.5. Sample frame data sources by city

SSES Participant	Source
Bogota	Ministry of Education Colombia - SIMAT and DUE (Ministry of Education) https://www.mineducacion.gov.co/portal/secciones/English-version/ https://www.mineducacion.gov.co/1759/w3-article-168883.html?_noredirect=1 https://medellin.edu.co/analisis-del-sector-educativo/sistemas-de-informacion/269-directorio-unico-de-establecimientos-educativos-due
Daegu	Korea Educational Statistics Service https://kess.kedi.re.kr/eng/index
Helsinki	Ministry of Education https://minedu.fi/en/general-education https://www.hel.fi/kasvatuksen-ja-koulutuksen-toimiala/en
Houston	Houston Independent School District http://www.houstonisd.org https://kinder.rice.edu/houston-education-research-consortium District student information system (Cognos)
Manizales	National Integrated Enrolment System database. https://terridata.dnp.gov.co/#/perfiles http://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion
Moscow	Register of the accredited educational institutions located in territory of constituent entities of the Russian Federation Department of Education and Science of Moscow https://www.mos.ru/authority/territory/ https://data.mos.ru/opendata/7719028495-register-of-the-accredited-educational-institutions-located-in-territory-of-constituent-entities-of-the-russian-federation https://data.mos.ru/opendata/7719028495-use-results-flat
Ottawa	Most recent admission records at each of the participating school boards, collected by the NTL. https://www.app.edu.gov.on.ca/eng/sift/secondary.asp https://www.app.edu.gov.on.ca/eng/sift/elementary.asp http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CSD&Geo2=PR&Code2=01&Data=Count&SearchType=Begins&SearchPR=01&TABID=1&B1=All&Code1=3506008&SearchText=ottawa
Sintra	Portuguese Education Statistics – General Directorate of Science and Education Statistics http://www.dgeec.mec.pt
Suzhou	Suzhou Education Bureau http://www.suzhou.gov.cn/
Istanbul	Republic of Turkey Ministry of National Education. https://www.meb.gov.tr/meb_haberindex.php?dil=en

Annex G. Item parameters for each scale

Table G.1. Scale reliability in the student assessment final scales by site

Scale	Site	All		Younger		Older	
		Omega	Alpha	Omega	Alpha	Omega	Alpha
ASS	Bogotá	0.81	0.80	0.76	0.75	0.87	0.86
	Daegu	0.89	0.89	0.87	0.88	0.91	0.91
	Helsinki	0.90	0.90	0.87	0.87	0.91	0.91
	Houston	0.80	0.80	0.77	0.77	0.86	0.85
	Manizales	0.80	0.79	0.73	0.71	0.87	0.86
	Moscow	0.85	0.84	0.82	0.81	0.88	0.88
	Ottawa	0.86	0.86	0.84	0.85	0.88	0.88
	Sintra	0.85	0.85	0.82	0.82	0.89	0.89
	Suzhou	0.79	0.79	0.78	0.77	0.85	0.84
	Istanbul	0.84	0.83	0.82	0.81	0.88	0.88
COO	Bogotá	0.76	0.75	0.76	0.74	0.76	0.74
	Daegu	0.80	0.79	0.81	0.79	0.80	0.78
	Helsinki	0.82	0.81	0.83	0.82	0.82	0.81
	Houston	0.80	0.79	0.80	0.79	0.81	0.80
	Manizales	0.76	0.74	0.77	0.74	0.75	0.73
	Moscow	0.77	0.76	0.77	0.76	0.77	0.76
	Ottawa	0.79	0.77	0.80	0.78	0.78	0.77
	Sintra	0.76	0.74	0.76	0.74	0.76	0.75
	Suzhou	0.83	0.83	0.82	0.82	0.81	0.81
	Istanbul	0.81	0.80	0.83	0.82	0.74	0.72
CRE	Bogotá	0.71	0.70	0.66	0.63	0.77	0.77
	Daegu	0.80	0.80	0.80	0.79	0.80	0.80
	Helsinki	0.76	0.75	0.75	0.74	0.78	0.78
	Houston	0.68	0.66	0.64	0.62	0.74	0.73
	Manizales	0.72	0.70	0.67	0.64	0.76	0.76
	Moscow	0.75	0.74	0.71	0.70	0.79	0.79
	Ottawa	0.73	0.73	0.70	0.69	0.75	0.75
	Sintra	0.74	0.74	0.71	0.69	0.77	0.77
	Suzhou	0.80	0.80	0.77	0.76	0.82	0.82
	Istanbul	0.75	0.74	0.74	0.73	0.76	0.76
CUR	Bogotá	0.72	0.66	0.71	0.64	0.73	0.68
	Daegu	0.82	0.81	0.82	0.80	0.83	0.81
	Helsinki	0.83	0.81	0.82	0.81	0.83	0.81
	Houston	0.78	0.76	0.76	0.75	0.79	0.77
	Manizales	0.72	0.66	0.71	0.65	0.72	0.66
	Moscow	0.79	0.78	0.77	0.75	0.81	0.81
	Ottawa	0.82	0.81	0.81	0.80	0.83	0.82
	Sintra	0.78	0.77	0.78	0.75	0.79	0.78
	Suzhou	0.81	0.80	0.78	0.77	0.81	0.80
	Istanbul	0.79	0.76	0.79	0.75	0.78	0.77

EMO	Bogotá	0.70	0.70	0.61	0.60	0.78	0.78
	Daegu	0.83	0.82	0.82	0.82	0.83	0.82
	Helsinki	0.76	0.77	0.75	0.75	0.79	0.79
	Houston	0.74	0.74	0.71	0.70	0.80	0.79
	Manizales	0.71	0.70	0.59	0.57	0.79	0.79
	Moscow	0.77	0.76	0.75	0.74	0.80	0.79
	Ottawa	0.80	0.80	0.80	0.80	0.81	0.81
	Sintra	0.76	0.76	0.71	0.71	0.81	0.81
	Suzhou	0.81	0.80	0.78	0.77	0.82	0.81
	Istanbul	0.72	0.71	0.70	0.69	0.75	0.74
EMP	Bogotá	0.65	0.65	0.62	0.61	0.71	0.71
	Daegu	0.78	0.78	0.79	0.79	0.77	0.77
	Helsinki	0.74	0.74	0.73	0.73	0.77	0.77
	Houston	0.68	0.68	0.65	0.65	0.73	0.73
	Manizales	0.67	0.65	0.66	0.64	0.68	0.68
	Moscow	0.71	0.71	0.70	0.70	0.72	0.72
	Ottawa	0.74	0.74	0.74	0.74	0.74	0.74
	Sintra	0.67	0.65	0.66	0.63	0.68	0.67
	Suzhou	0.78	0.77	0.78	0.76	0.76	0.75
	Istanbul	0.73	0.73	0.74	0.74	0.70	0.70
ENE	Bogotá	0.73	0.72	0.67	0.66	0.78	0.77
	Daegu	0.82	0.81	0.81	0.79	0.82	0.81
	Helsinki	0.80	0.80	0.76	0.75	0.81	0.80
	Houston	0.76	0.75	0.71	0.71	0.80	0.79
	Manizales	0.73	0.72	0.63	0.63	0.78	0.77
	Moscow	0.79	0.78	0.74	0.73	0.83	0.82
	Ottawa	0.80	0.80	0.77	0.77	0.79	0.79
	Sintra	0.74	0.73	0.68	0.67	0.79	0.78
	Suzhou	0.79	0.78	0.74	0.73	0.81	0.80
	Istanbul	0.74	0.74	0.72	0.71	0.76	0.76
OPT	Bogotá	0.79	0.77	0.73	0.69	0.84	0.83
	Daegu	0.85	0.84	0.84	0.83	0.84	0.83
	Helsinki	0.84	0.83	0.80	0.78	0.88	0.87
	Houston	0.82	0.81	0.78	0.77	0.86	0.85
	Manizales	0.80	0.78	0.73	0.67	0.85	0.84
	Moscow	0.83	0.82	0.78	0.77	0.87	0.86
	Ottawa	0.87	0.87	0.83	0.83	0.88	0.88
	Sintra	0.81	0.80	0.73	0.69	0.85	0.85
	Suzhou	0.83	0.82	0.79	0.78	0.86	0.85
	Istanbul	0.84	0.83	0.81	0.80	0.86	0.86
PER	Bogotá	0.74	0.73	0.68	0.66	0.80	0.79
	Daegu	0.87	0.86	0.86	0.85	0.87	0.87
	Helsinki	0.82	0.80	0.80	0.77	0.85	0.84
	Houston	0.79	0.79	0.76	0.75	0.84	0.83
	Manizales	0.75	0.74	0.66	0.63	0.81	0.81
	Moscow	0.87	0.87	0.84	0.84	0.89	0.89
	Ottawa	0.86	0.85	0.84	0.83	0.87	0.87
	Sintra	0.79	0.78	0.74	0.74	0.84	0.83
	Suzhou	0.85	0.84	0.82	0.81	0.85	0.84
	Istanbul	0.81	0.80	0.80	0.79	0.85	0.84
RES	Bogotá	0.65	0.64	0.63	0.62	0.67	0.67
	Daegu	0.75	0.74	0.75	0.74	0.76	0.75

	Helsinki	0.68	0.68	0.65	0.65	0.72	0.72
	Houston	0.66	0.66	0.63	0.62	0.72	0.72
	Manizales	0.63	0.62	0.57	0.56	0.68	0.68
	Moscow	0.79	0.78	0.77	0.76	0.81	0.80
	Ottawa	0.74	0.74	0.72	0.71	0.75	0.75
	Sintra	0.71	0.71	0.67	0.67	0.75	0.75
	Suzhou	0.78	0.77	0.75	0.74	0.80	0.79
	Istanbul	0.74	0.73	0.72	0.71	0.78	0.77
SEL	Bogotá	0.75	0.74	0.72	0.71	0.78	0.77
	Daegu	0.76	0.76	0.78	0.77	0.75	0.75
	Helsinki	0.75	0.75	0.75	0.74	0.76	0.76
	Houston	0.77	0.77	0.76	0.76	0.80	0.79
	Manizales	0.74	0.73	0.69	0.67	0.79	0.78
	Moscow	0.79	0.79	0.78	0.78	0.80	0.80
	Ottawa	0.79	0.79	0.79	0.78	0.80	0.79
	Sintra	0.74	0.74	0.72	0.71	0.77	0.77
	Suzhou	0.76	0.72	0.75	0.72	0.75	0.72
	Istanbul	0.76	0.76	0.77	0.76	0.76	0.75
SOC	Bogotá	0.71	0.69	0.65	0.62	0.77	0.75
	Daegu	0.82	0.80	0.80	0.78	0.81	0.80
	Helsinki	0.75	0.74	0.71	0.69	0.80	0.79
	Houston	0.72	0.71	0.68	0.66	0.78	0.77
	Manizales	0.73	0.71	0.67	0.63	0.78	0.76
	Moscow	0.76	0.75	0.72	0.71	0.80	0.79
	Ottawa	0.71	0.70	0.70	0.67	0.72	0.71
	Sintra	0.71	0.68	0.65	0.60	0.74	0.73
	Suzhou	0.75	0.71	0.71	0.67	0.77	0.75
	Istanbul	0.70	0.66	0.70	0.66	0.72	0.70
STR	Bogotá	0.66	0.66	0.63	0.63	0.69	0.69
	Daegu	0.83	0.83	0.80	0.80	0.83	0.83
	Helsinki	0.82	0.82	0.77	0.77	0.86	0.86
	Houston	0.76	0.75	0.72	0.71	0.81	0.81
	Manizales	0.68	0.68	0.64	0.63	0.72	0.72
	Moscow	0.83	0.83	0.80	0.80	0.87	0.87
	Ottawa	0.85	0.85	0.81	0.81	0.87	0.87
	Sintra	0.69	0.70	0.62	0.62	0.76	0.76
	Suzhou	0.83	0.83	0.81	0.80	0.84	0.83
	Istanbul	0.80	0.80	0.79	0.78	0.84	0.84
TOL	Bogotá	0.69	0.67	0.66	0.64	0.74	0.73
	Daegu	0.74	0.73	0.74	0.73	0.74	0.74
	Helsinki	0.76	0.76	0.73	0.72	0.80	0.80
	Houston	0.74	0.74	0.73	0.72	0.77	0.76
	Manizales	0.69	0.67	0.67	0.64	0.72	0.72
	Moscow	0.72	0.72	0.69	0.69	0.77	0.77
	Ottawa	0.76	0.76	0.74	0.74	0.79	0.79
	Sintra	0.72	0.72	0.68	0.67	0.77	0.77
	Suzhou	0.72	0.71	0.70	0.70	0.77	0.77
	Istanbul	0.75	0.75	0.72	0.72	0.81	0.80
TRU	Bogotá	0.76	0.75	0.71	0.70	0.77	0.75
	Daegu	0.78	0.78	0.74	0.73	0.82	0.82
	Helsinki	0.84	0.84	0.84	0.84	0.83	0.83
	Houston	0.79	0.79	0.77	0.76	0.80	0.80

	Manizales	0.76	0.75	0.69	0.67	0.76	0.75
	Moscow	0.83	0.82	0.80	0.80	0.83	0.83
	Ottawa	0.82	0.82	0.80	0.80	0.82	0.82
	Sintra	0.76	0.75	0.74	0.72	0.76	0.76
	Suzhou	0.85	0.85	0.84	0.83	0.85	0.85
	Istanbul	0.75	0.75	0.73	0.72	0.78	0.77
EFF	Bogotá	0.70	0.68	0.70	0.68	0.70	0.69
	Daegu	0.79	0.79	0.82	0.81	0.77	0.77
	Helsinki	0.70	0.68	0.71	0.69	0.71	0.69
	Houston	0.74	0.73	0.73	0.73	0.75	0.74
	Manizales	0.71	0.69	0.72	0.70	0.69	0.68
	Moscow	0.71	0.70	0.71	0.70	0.73	0.72
	Ottawa	0.74	0.74	0.76	0.76	0.72	0.71
	Sintra	0.67	0.65	0.70	0.67	0.64	0.62
	Suzhou	0.80	0.79	0.79	0.78	0.78	0.78
	Istanbul	0.72	0.70	0.74	0.71	0.69	0.67
MOT	Bogotá	0.74	0.74	0.73	0.72	0.73	0.72
	Daegu	0.81	0.81	0.82	0.82	0.79	0.79
	Helsinki	0.74	0.74	0.74	0.73	0.76	0.76
	Houston	0.73	0.73	0.73	0.72	0.75	0.74
	Manizales	0.75	0.74	0.73	0.72	0.74	0.74
	Moscow	0.74	0.73	0.74	0.73	0.74	0.73
	Ottawa	0.75	0.74	0.74	0.73	0.76	0.75
	Sintra	0.72	0.72	0.72	0.72	0.73	0.72
	Suzhou	0.78	0.76	0.78	0.76	0.73	0.70
	Istanbul	0.80	0.80	0.81	0.81	0.78	0.78

Table G.2. Scale reliability in the parent assessment final scales by site

Scale	Site	All		Younger		Older	
		Omega	Alpha	Omega	Alpha	Omega	Alpha
ASS	Bogotá	0.78	0.78	0.76	0.77	0.80	0.80
	Daegu	0.89	0.90	0.88	0.90	0.89	0.91
	Helsinki	0.92	0.92	0.91	0.92	0.92	0.93
	Houston	0.83	0.85	0.82	0.84	0.85	0.86
	Manizales	0.76	0.77	0.75	0.75	0.78	0.78
	Moscow	0.87	0.88	0.87	0.87	0.87	0.88
	Ottawa	0.88	0.90	0.88	0.90	0.89	0.90
	Sintra	0.88	0.89	0.87	0.89	0.89	0.90
	Suzhou	0.83	0.83	0.82	0.83	0.85	0.84
	Istanbul	0.82	0.82	0.81	0.81	0.84	0.85
COO	Bogotá	0.75	0.77	0.74	0.76	0.76	0.77
	Daegu	0.77	0.79	0.78	0.79	0.76	0.79
	Helsinki	0.83	0.83	0.83	0.83	0.83	0.83
	Houston	0.83	0.86	0.81	0.84	0.84	0.87
	Manizales	0.75	0.76	0.74	0.76	0.75	0.75
	Moscow	0.80	0.81	0.79	0.80	0.80	0.81
	Ottawa	0.80	0.82	0.79	0.83	0.81	0.82
	Sintra	0.78	0.79	0.79	0.80	0.77	0.79
	Suzhou	0.83	0.82	0.82	0.82	0.84	0.84
	Istanbul	0.74	0.74	0.74	0.74	0.75	0.76
CRE	Bogotá	0.73	0.76	0.72	0.75	0.74	0.77
	Daegu	0.83	0.84	0.83	0.84	0.82	0.83
	Helsinki	0.84	0.85	0.81	0.83	0.86	0.87
	Houston	0.79	0.80	0.76	0.76	0.84	0.84
	Manizales	0.75	0.76	0.74	0.75	0.74	0.77
	Moscow	0.81	0.81	0.80	0.80	0.81	0.82
	Ottawa	0.80	0.83	0.75	0.82	0.84	0.83
	Sintra	0.80	0.81	0.81	0.81	0.79	0.82
	Suzhou	0.81	0.81	0.81	0.80	0.82	0.81
	Istanbul	0.76	0.75	0.75	0.74	0.78	0.78
CUR	Bogotá	0.75	0.75	0.73	0.73	0.75	0.75
	Daegu	0.85	0.86	0.84	0.85	0.85	0.86
	Helsinki	0.84	0.85	0.82	0.83	0.85	0.86
	Houston	0.84	0.85	0.83	0.84	0.83	0.84
	Manizales	0.76	0.75	0.75	0.74	0.76	0.75
	Moscow	0.85	0.85	0.84	0.84	0.85	0.85
	Ottawa	0.85	0.87	0.79	0.84	0.87	0.88
	Sintra	0.83	0.85	0.83	0.83	0.83	0.85
	Suzhou	0.84	0.84	0.84	0.83	0.85	0.84
	Istanbul	0.81	0.81	0.79	0.79	0.83	0.83
EMO	Bogotá	0.76	0.78	0.75	0.76	0.78	0.80
	Daegu	0.79	0.82	0.80	0.82	0.79	0.82
	Helsinki	0.83	0.85	0.82	0.84	0.85	0.86
	Houston	0.81	0.83	0.81	0.84	0.80	0.83
	Manizales	0.76	0.77	0.73	0.74	0.78	0.80
	Moscow	0.82	0.83	0.82	0.83	0.82	0.83
	Ottawa	0.85	0.87	0.84	0.87	0.86	0.88
	Sintra	0.79	0.81	0.78	0.80	0.81	0.82
	Suzhou	0.79	0.80	0.80	0.80	0.78	0.79

	Istanbul	0.71	0.71	0.69	0.69	0.76	0.76
EMP	Bogotá	0.68	0.69	0.67	0.68	0.68	0.70
	Daegu	0.79	0.81	0.79	0.81	0.79	0.81
	Helsinki	0.83	0.84	0.83	0.84	0.83	0.84
	Houston	0.74	0.75	0.70	0.70	0.79	0.81
	Manizales	0.66	0.68	0.64	0.66	0.68	0.69
	Moscow	0.77	0.78	0.75	0.76	0.79	0.80
	Ottawa	0.82	0.83	0.81	0.83	0.82	0.83
	Sintra	0.74	0.74	0.74	0.73	0.75	0.74
	Suzhou	0.79	0.79	0.79	0.79	0.80	0.80
	Istanbul	0.74	0.74	0.73	0.73	0.77	0.76
ENE	Bogotá	0.73	0.73	0.67	0.67	0.75	0.75
	Daegu	0.77	0.75	0.72	0.70	0.76	0.75
	Helsinki	0.79	0.77	0.77	0.75	0.77	0.75
	Houston	0.76	0.76	0.71	0.71	0.80	0.80
	Manizales	0.74	0.75	0.70	0.71	0.75	0.75
	Moscow	0.76	0.75	0.74	0.72	0.76	0.75
	Ottawa	0.80	0.80	0.77	0.75	0.79	0.79
	Sintra	0.78	0.76	0.75	0.70	0.80	0.78
	Suzhou	0.69	0.68	0.67	0.66	0.71	0.70
	Istanbul	0.68	0.67	0.65	0.63	0.71	0.71
OPT	Bogotá	0.76	0.77	0.72	0.75	0.78	0.79
	Daegu	0.83	0.84	0.81	0.82	0.83	0.85
	Helsinki	0.87	0.88	0.86	0.88	0.88	0.89
	Houston	0.83	0.85	0.81	0.82	0.85	0.87
	Manizales	0.77	0.77	0.74	0.73	0.79	0.79
	Moscow	0.85	0.85	0.83	0.84	0.85	0.86
	Ottawa	0.87	0.88	0.83	0.86	0.89	0.89
	Sintra	0.78	0.77	0.78	0.75	0.78	0.78
	Suzhou	0.79	0.78	0.79	0.78	0.80	0.80
	Istanbul	0.79	0.78	0.76	0.75	0.83	0.82
PER	Bogotá	0.84	0.86	0.84	0.86	0.85	0.86
	Daegu	0.87	0.89	0.87	0.88	0.88	0.89
	Helsinki	0.91	0.92	0.91	0.91	0.92	0.93
	Houston	0.88	0.90	0.89	0.90	0.88	0.89
	Manizales	0.85	0.86	0.84	0.85	0.85	0.86
	Moscow	0.91	0.91	0.91	0.91	0.91	0.91
	Ottawa	0.91	0.93	0.90	0.92	0.92	0.93
	Sintra	0.89	0.91	0.88	0.89	0.91	0.92
	Suzhou	0.86	0.86	0.86	0.86	0.85	0.85
	Istanbul	0.85	0.85	0.84	0.84	0.87	0.87
RES	Bogotá	0.77	0.78	0.76	0.78	0.77	0.78
	Daegu	0.78	0.79	0.77	0.77	0.80	0.81
	Helsinki	0.84	0.84	0.82	0.83	0.85	0.86
	Houston	0.81	0.81	0.81	0.81	0.80	0.81
	Manizales	0.75	0.76	0.74	0.75	0.76	0.77
	Moscow	0.85	0.85	0.85	0.85	0.84	0.85
	Ottawa	0.82	0.83	0.80	0.81	0.83	0.85
	Sintra	0.82	0.84	0.80	0.82	0.84	0.86
	Suzhou	0.78	0.78	0.77	0.77	0.79	0.79
	Istanbul	0.75	0.75	0.73	0.73	0.79	0.78
SEL	Bogotá	0.79	0.80	0.79	0.81	0.78	0.79

	Daegu	0.79	0.82	0.81	0.82	0.77	0.81
	Helsinki	0.79	0.81	0.79	0.82	0.78	0.80
	Houston	0.81	0.84	0.83	0.85	0.77	0.81
	Manizales	0.79	0.80	0.78	0.79	0.80	0.81
	Moscow	0.82	0.83	0.82	0.83	0.81	0.82
	Ottawa	0.81	0.83	0.82	0.84	0.81	0.83
	Sintra	0.80	0.82	0.81	0.83	0.79	0.81
	Suzhou	0.77	0.75	0.77	0.75	0.75	0.72
	Istanbul	0.76	0.76	0.75	0.75	0.78	0.77
SOC	Bogotá	0.74	0.76	0.70	0.72	0.76	0.77
	Daegu	0.79	0.80	0.77	0.78	0.79	0.80
	Helsinki	0.82	0.82	0.78	0.78	0.85	0.84
	Houston	0.79	0.80	0.73	0.74	0.83	0.83
	Manizales	0.75	0.76	0.70	0.71	0.76	0.76
	Moscow	0.79	0.80	0.77	0.77	0.80	0.80
	Ottawa	0.80	0.80	0.74	0.74	0.82	0.82
	Sintra	0.77	0.77	0.75	0.75	0.78	0.77
	Suzhou	0.77	0.75	0.76	0.74	0.78	0.76
	Istanbul	0.71	0.69	0.70	0.68	0.71	0.69
STR	Bogotá	0.76	0.76	0.76	0.76	0.76	0.76
	Daegu	0.83	0.84	0.83	0.84	0.83	0.84
	Helsinki	0.86	0.87	0.86	0.87	0.87	0.88
	Houston	0.82	0.84	0.82	0.84	0.81	0.84
	Manizales	0.76	0.76	0.76	0.76	0.75	0.75
	Moscow	0.85	0.86	0.84	0.85	0.86	0.87
	Ottawa	0.90	0.91	0.89	0.91	0.90	0.91
	Sintra	0.78	0.79	0.75	0.77	0.81	0.81
	Suzhou	0.84	0.83	0.83	0.83	0.84	0.84
	Istanbul	0.79	0.79	0.78	0.78	0.82	0.82
TOL	Bogotá	0.70	0.66	0.68	0.63	0.71	0.68
	Daegu	0.80	0.81	0.79	0.80	0.80	0.81
	Helsinki	0.81	0.82	0.80	0.79	0.83	0.85
	Houston	0.78	0.80	0.74	0.78	0.82	0.83
	Manizales	0.70	0.69	0.69	0.66	0.71	0.70
	Moscow	0.79	0.79	0.78	0.78	0.80	0.81
	Ottawa	0.82	0.83	0.82	0.85	0.81	0.82
	Sintra	0.78	0.79	0.77	0.77	0.79	0.81
	Suzhou	0.78	0.78	0.78	0.77	0.79	0.78
	Istanbul	0.78	0.78	0.77	0.76	0.82	0.82
TRU	Bogotá	0.76	0.77	0.74	0.75	0.76	0.77
	Daegu	0.75	0.76	0.75	0.75	0.75	0.76
	Helsinki	0.82	0.84	0.81	0.84	0.83	0.84
	Houston	0.77	0.75	0.75	0.72	0.79	0.79
	Manizales	0.76	0.76	0.72	0.72	0.77	0.77
	Moscow	0.82	0.82	0.80	0.80	0.83	0.83
	Ottawa	0.79	0.81	0.78	0.79	0.79	0.80
	Sintra	0.76	0.77	0.74	0.75	0.77	0.75
	Suzhou	0.79	0.78	0.79	0.78	0.79	0.79
	Istanbul	0.78	0.78	0.77	0.77	0.79	0.79
EFF	Bogotá	0.72	0.73	0.71	0.72	0.73	0.74
	Daegu	0.76	0.78	0.75	0.76	0.78	0.80
	Helsinki	0.71	0.71	0.67	0.67	0.76	0.75

	Houston	0.82	0.82	0.81	0.82	0.83	0.84
	Manizales	0.74	0.75	0.73	0.74	0.74	0.75
	Moscow	0.77	0.77	0.75	0.75	0.79	0.79
	Ottawa	0.76	0.78	0.74	0.76	0.78	0.79
	Sintra	0.73	0.74	0.74	0.76	0.72	0.72
	Suzhou	0.81	0.81	0.80	0.80	0.82	0.82
	Istanbul	0.74	0.73	0.73	0.72	0.75	0.75
MOT	Bogotá	0.80	0.81	0.79	0.81	0.81	0.82
	Daegu	0.79	0.80	0.77	0.78	0.81	0.83
	Helsinki	0.76	0.76	0.75	0.74	0.79	0.78
	Houston	0.82	0.84	0.82	0.84	0.83	0.85
	Manizales	0.80	0.82	0.80	0.82	0.81	0.83
	Moscow	0.82	0.82	0.81	0.82	0.82	0.83
	Ottawa	0.80	0.80	0.77	0.77	0.82	0.83
	Sintra	0.82	0.83	0.80	0.81	0.84	0.86
	Suzhou	0.77	0.77	0.77	0.77	0.78	0.78
	Istanbul	0.82	0.82	0.81	0.82	0.83	0.84

Table G.3. Items factor loading in the student assessment final scales

Scale	Item ID	Item	Reverse Word	CFA Factor Loading		
				All	Younger	Older
ASS	ASS01	A leader	0	0.711	0.659	0.768
	ASS02	Want to be in charge	0	0.650	0.602	0.704
	ASS03	Know how to convince others to do what I want	0	0.413	0.385	0.446
	ASS04	Enjoy leading others	0	0.747	0.695	0.809
	ASS05	Dislike leading a team	1	0.577	0.504	0.670
	ASS06	Like to be a leader in my class	0	0.764	0.770	0.787
	ASS08	Dominant, and act as a leader	0	0.767	0.744	0.799
COO	COO01	Like to help others	0	0.653	0.621	0.685
	COO02	Get along well with others	0	0.582	0.605	0.547
	COO03	Work well with other people	0	0.577	0.592	0.548
	COO04	Start arguments with others	1	0.294	0.329	0.240
	COO06	Always willing to help my classmates	0	0.724	0.717	0.725
	COO07	Ready to help anybody	0	0.722	0.710	0.732
	COO08	Polite, courteous to others	0	0.580	0.627	0.544
CRE	CRE01	Find new ways to do things	0	0.545	0.527	0.561
	CRE02	Original, come up with new ideas	0	0.695	0.663	0.728
	CRE04	Sometimes find a solution other people don't see	0	0.559	0.543	0.589
	CRE05	Like to create things	0	0.638	0.618	0.646
	CRE06	Have a good imagination	0	0.561	0.536	0.580
	CRE07	Find it difficult to create new things	1	0.471	0.389	0.548
CUR	CUR02	Eager to learn	0	0.727	0.702	0.749
	CUR04	Like to know how things work	0	0.511	0.501	0.518
	CUR05	Like learning new things	0	0.727	0.726	0.723
	CUR06	Don't like learning	1	0.524	0.450	0.618
	CUR07	Love learning new things in school	0	0.773	0.786	0.746
	CUR08	Find science interesting	0	0.494	0.492	0.479
EMO	EMO01	Not easily upset	0	0.457	0.401	0.508
	EMO02	Keep my emotions under control	0	0.568	0.556	0.583
	EMO03	Get mad easily	1	0.732	0.705	0.756
	EMO04	Know how to control my anger	0	0.651	0.621	0.683
	EMO06	Have unpredictable emotions and moods	1	0.424	0.357	0.493
	EMO07	Stay calm even in tense situations	0	0.426	0.415	0.436
	EMO08	Often feel angry	1	0.457	0.401	0.508
EMP	EMP01	Helpful and unselfish with others	0	0.444	0.422	0.471
	EMP02	Important to me that my friends are okay	0	0.404	0.375	0.443
	EMP03	Can sense how others feel	0	0.609	0.613	0.619
	EMP04	Know how to comfort others	0	0.632	0.637	0.633
	EMP06	Understand what others want	0	0.604	0.601	0.609
	EMP07	Warm toward others	0	0.509	0.492	0.529
ENE	ENE01	Full of energy	0	0.690	0.608	0.735
	ENE02	Show a lot of enthusiasm	0	0.490	0.441	0.519
	ENE03	Less active than other people	1	0.501	0.449	0.549
	ENE04	Have less energy than my classmates	1	0.658	0.604	0.706
	ENE05	Like sports where I can run	0	0.474	0.466	0.437
	ENE06	Tire out quickly	1	0.509	0.455	0.532
	ENE08	Maintain high energy throughout the day	0	0.712	0.643	0.733

OPT	OPT01	Often feel sad	1	0.447	0.332	0.534
	OPT02	Believe good things will happen to me	0	0.583	0.518	0.636
	OPT03	Wake up happy almost every day	0	0.673	0.626	0.685
	OPT05	Enjoy life	0	0.763	0.717	0.795
	OPT06	Look at the bright side of life	0	0.727	0.692	0.749
	OPT07	A happy person	0	0.814	0.765	0.840
	OPT08	Expect bad things to happen	1	0.462	0.393	0.489
PER	PER01	Keep working on a task until it is finished	0	0.724	0.684	0.740
	PER02	Make sure that I finish tasks	0	0.724	0.689	0.739
	PER03	Give up easily	1	0.436	0.394	0.450
	PER04	Finish what I start	0	0.736	0.707	0.768
	PER05	Leave things unfinished	1	0.589	0.506	0.656
	PER07	Hate leaving tasks unfinished	0	0.529	0.481	0.591
	PER08	Finish things despite difficulties in the way	0	0.659	0.629	0.694
RES	RES01	Sometimes behave irresponsibly	1	0.454	0.402	0.497
	RES03	Often forget my duties	1	0.601	0.569	0.618
	RES04	Avoid responsibilities	1	0.564	0.513	0.616
	RES05	Keep my promises	0	0.464	0.505	0.433
	RES06	A responsible person	0	0.609	0.586	0.655
	RES08	Forget to do work I was asked to do	1	0.613	0.603	0.607
SEL	SEL01	Careful with what I say to others	0	0.501	0.490	0.516
	SEL02	Can control my actions	0	0.528	0.559	0.496
	SEL03	Think carefully before doing something	0	0.742	0.735	0.740
	SEL04	Avoid mistakes by working carefully	0	0.524	0.523	0.517
	SEL07	Stop to think before acting	0	0.682	0.649	0.723
	SEL08	Often rush into action without thinking	1	0.532	0.478	0.595
SOC	SOC01	Outgoing and sociable	0	0.548	0.463	0.646
	SOC02	Have many friends	0	0.710	0.692	0.697
	SOC03	Like to be with my friends	0	0.511	0.527	0.469
	SOC04	Like to be alone	1	0.357	0.293	0.345
	SOC05	Like talking to a lot of different people	0	0.502	0.454	0.613
	SOC07	Make friends easily	0	0.743	0.713	0.768
STR	STR01	Relaxed and handle stress well	0	0.417	0.309	0.498
	STR02	Get nervous easily	1	0.652	0.610	0.684
	STR03	Worry about many things	1	0.645	0.615	0.659
	STR04	Afraid of many things	1	0.627	0.608	0.656
	STR05	Panic easily	1	0.684	0.646	0.719
	STR07	Often worried about something	1	0.725	0.697	0.740
TOL	TOL01	Willing to be friends with people from other cultures	0	0.532	0.513	0.557
	TOL02	Ask questions about other cultures	0	0.618	0.581	0.656
	TOL03	Feel comfortable in new cultural environments	0	0.486	0.526	0.475
	TOL04	Want to travel to other countries	0	0.408	0.392	0.428
	TOL05	Like hearing about other cultures and religions	0	0.707	0.660	0.758
	TOL06	Not interested in other countries and cultures	1	0.473	0.402	0.554
	TOL07	Learn a lot from people with differing beliefs	0	0.510	0.493	0.533
TRU	TRU01	Think most of my classmates keep their promises	0	0.540	0.502	0.549
	TRU02	Believe that my friends can keep my secrets	0	0.589	0.581	0.587
	TRU05	Distrust people	1	0.568	0.466	0.630
	TRU06	Believe that other people will help me	0	0.613	0.603	0.591
	TRU07	Believe that most people are honest	0	0.749	0.722	0.744

	TRU08	Trust others	0	0.792	0.742	0.833
EFF	ASS01	A leader	0	0.377	0.333	0.465
	CRE02	Original, come up with new ideas	0	0.568	0.550	0.589
	CUR05	Like learning new things	0	0.563	0.582	0.533
	EMP06	Understand what others want	0	0.437	0.464	0.409
	OPT04	Always positive about the future	0	0.521	0.544	0.492
	PER04	Finish what I start	0	0.576	0.594	0.541
	RES02	Reliable and can always be counted on	0	0.506	0.515	0.514
	SEL02	Can control my actions	0	0.464	0.506	0.410
MOT	CUR02	Eager to learn	0	0.605	0.644	0.545
	ENE02	Show a lot of enthusiasm	0	0.468	0.483	0.423
	PER02	Make sure that I finish tasks	0	0.673	0.657	0.686
	PER08	Finish things despite difficulties in the way	0	0.657	0.613	0.701
	RES06	A responsible person	0	0.647	0.656	0.647
	SEL06	Like to make sure there are no mistakes	0	0.484	0.490	0.482

Table G.4. Items factor loading of the parent assessment final scales

Scale	Item ID	Item	Reverse Word	CFA Factor Loading		
				All	Younger	Older
ASS	ASS01	A leader	0	0.678	0.645	0.714
	ASS02	Want to be in charge	0	0.623	0.620	0.622
	ASS03	Know how to convince others to do what I want	0	0.403	0.404	0.413
	ASS04	Enjoy leading others	0	0.769	0.751	0.784
	ASS05	Dislike leading a team	1	0.471	0.445	0.489
	ASS06	Like to be a leader in my class	0	0.743	0.726	0.759
	ASS08	Dominant, and act as a leader	0	0.645	0.626	0.665
COO	COO01	Like to help others	0	0.631	0.607	0.646
	COO02	Get along well with others	0	0.628	0.633	0.621
	COO03	Work well with other people	0	0.624	0.630	0.615
	COO04	Start arguments with others	1	0.315	0.319	0.323
	COO06	Always willing to help my classmates	0	0.716	0.703	0.722
	COO07	Ready to help anybody	0	0.682	0.644	0.711
	COO08	Polite, courteous to others	0	0.613	0.626	0.603
CRE	CRE01	Find new ways to do things	0	0.599	0.600	0.604
	CRE02	Original, come up with new ideas	0	0.698	0.688	0.705
	CRE04	Sometimes find a solution other people don't see	0	0.644	0.642	0.649
	CRE05	Like to create things	0	0.702	0.701	0.697
	CRE06	Have a good imagination	0	0.642	0.640	0.632
	CRE07	Find it difficult to create new things	1	0.513	0.482	0.539
CUR	CUR01	Curious about many different things	0	0.619	0.599	0.614
	CUR02	Eager to learn	0	0.677	0.667	0.704
	CUR03	Like to ask questions	0	0.599	0.588	0.578
	CUR04	Like to know how things work	0	0.665	0.657	0.653
	CUR05	Like learning new things	0	0.729	0.719	0.728
	CUR06	Don't like learning	1	0.415	0.414	0.435
	CUR07	Love learning new things in school	0	0.704	0.686	0.708
	CUR08	Find science interesting	0	0.531	0.500	0.543
EMO	EMO02	Keep my emotions under control	0	0.547	0.553	0.544
	EMO03	Get mad easily	1	0.658	0.632	0.682
	EMO04	Know how to control my anger	0	0.620	0.606	0.634
	EMO05	Change my mood a lot	1	0.619	0.591	0.652
	EMO07	Stay calm even in tense situations	0	0.471	0.451	0.494
	EMO08	Often feel angry	1	0.680	0.655	0.704
EMP	EMP01	Helpful and unselfish with others	0	0.472	0.441	0.501
	EMP03	Can sense how others feel	0	0.626	0.612	0.641
	EMP04	Know how to comfort others	0	0.611	0.595	0.625
	EMP05	Predict the needs of others	0	0.566	0.563	0.572
	EMP06	Understand what others want	0	0.639	0.633	0.647
	EMP07	Warm toward others	0	0.493	0.473	0.512
ENE	ENE01	Full of energy	0	0.657	0.610	0.657
	ENE02	Show a lot of enthusiasm	0	0.539	0.513	0.536
	ENE03	Less active than other people	1	0.443	0.425	0.461
	ENE05	Like sports where I can run	0	0.475	0.457	0.433
	ENE06	Tire out quickly	1	0.442	0.397	0.490
	ENE07	Slow to start in the morning	1	0.445	0.420	0.495
	ENE08	Maintain high energy throughout the day	0	0.749	0.717	0.741
OPT	OPT01	Often feel sad	1	0.341	0.297	0.370
	OPT02	Believe good things will happen to me	0	0.616	0.599	0.636

	OPT03	Wake up happy almost every day	0	0.581	0.554	0.584
	OPT04	Always positive about the future	0	0.685	0.676	0.695
	OPT05	Enjoy life	0	0.687	0.656	0.698
	OPT06	Look at the bright side of life	0	0.730	0.707	0.746
	OPT07	A happy person	0	0.719	0.680	0.735
	OPT08	Expect bad things to happen	1	0.315	0.290	0.339
PER	PER01	Keep working on a task until it is finished	0	0.789	0.789	0.789
	PER02	Make sure that I finish tasks	0	0.779	0.779	0.780
	PER03	Give up easily	1	0.446	0.432	0.457
	PER04	Finish what I start	0	0.805	0.803	0.806
	PER05	Leave things unfinished	1	0.657	0.635	0.678
	PER06	Stop when work becomes too difficult	1	0.532	0.522	0.539
	PER07	Hate leaving tasks unfinished	0	0.658	0.651	0.665
	PER08	Finish things despite difficulties in the way	0	0.748	0.742	0.754
RES	RES01	Sometimes behave irresponsibly	1	0.547	0.521	0.569
	RES03	Often forget my duties	1	0.596	0.578	0.611
	RES04	Avoid responsibilities	1	0.655	0.639	0.668
	RES05	Keep my promises	0	0.622	0.614	0.632
	RES06	A responsible person	0	0.679	0.681	0.677
	RES08	Forget to do work I was asked to do	1	0.657	0.640	0.673
SEL	SEL01	Careful with what I say to others	0	0.499	0.517	0.482
	SEL02	Can control my actions	0	0.601	0.592	0.610
	SEL03	Think carefully before doing something	0	0.777	0.773	0.777
	SEL04	Avoid mistakes by working carefully	0	0.599	0.613	0.577
	SEL07	Stop to think before acting	0	0.712	0.708	0.711
	SEL08	Often rush into action without thinking	1	0.609	0.605	0.607
SOC	SOC01	Outgoing and sociable	0	0.674	0.658	0.678
	SOC02	Have many friends	0	0.692	0.661	0.707
	SOC03	Like to be with my friends	0	0.509	0.489	0.505
	SOC04	Like to be alone	1	0.352	0.326	0.321
	SOC05	Like talking to a lot of different people	0	0.575	0.542	0.611
	SOC07	Make friends easily	0	0.728	0.713	0.728
STR	STR01	Relaxed and handle stress well	0	0.356	0.332	0.382
	STR02	Get nervous easily	1	0.668	0.661	0.672
	STR03	Worry about many things	1	0.398	0.388	0.413
	STR04	Afraid of many things	1	0.655	0.645	0.661
	STR05	Panic easily	1	0.730	0.726	0.730
	STR06	Often feel nervous	1	0.733	0.731	0.742
	STR08	Get scared easily	1	0.691	0.688	0.692
TOL	TOL01	Willing to be friends with people from other cultures	0	0.489	0.464	0.508
	TOL02	Ask questions about other cultures	0	0.633	0.607	0.651
	TOL04	Want to travel to other countries	0	0.527	0.524	0.530
	TOL05	Like hearing about other cultures and religions	0	0.733	0.717	0.746
	TOL06	Not interested in other countries and cultures	1	0.357	0.329	0.386
	TOL07	Learn a lot from people with differing beliefs	0	0.387	0.363	0.412
	TOL08	Love to learn about other countries and cultures	0	0.800	0.795	0.803
TRU	TRU01	Think most of my classmates keep their promises	0	0.475	0.436	0.500
	TRU02	Believe that my friends can keep my secrets	0	0.592	0.584	0.595
	TRU04	Believe most people are kind	0	0.664	0.636	0.670
	TRU05	Distrust people	1	0.458	0.408	0.504
	TRU06	Believe that other people will help me	0	0.534	0.516	0.536
	TRU07	Believe that most people are honest	0	0.787	0.764	0.795

	TRU08	Trust others	0	0.751	0.732	0.763
EFF	ASS01	A leader	0	0.488	0.465	0.508
	CRE02	Original, come up with new ideas	0	0.600	0.581	0.616
	CUR05	Like learning new things	0	0.580	0.557	0.605
	EMP06	Understand what others want	0	0.456	0.457	0.456
	OPT04	Always positive about the future	0	0.550	0.549	0.548
	PER04	Finish what I start	0	0.599	0.610	0.597
	RES02	Reliable and can always be counted on	0	0.609	0.615	0.605
	SEL02	Can control my actions	0	0.511	0.512	0.520
MOT	CUR02	Eager to learn	0	0.674	0.664	0.687
	ENE02	Show a lot of enthusiasm	0	0.537	0.528	0.564
	PER02	Make sure that I finish tasks	0	0.743	0.744	0.740
	PER08	Finish things despite difficulties in the way	0	0.760	0.752	0.768
	RES06	A responsible person	0	0.713	0.715	0.710
	SEL06	Like to make sure there are no mistakes	0	0.621	0.621	0.625

Table G.5. CFA model-data fit indices of student assessment scales

Scale	All					Younger					Older				
	RMSEA	SRMR	CFI	TLI	BIC	RMSEA	SRMR	CFI	TLI	BIC	RMSEA	SRMR	CFI	TLI	BIC
ASS	0.042	0.021	0.979	0.969	1222790	0.038	0.020	0.981	0.972	663261	0.038	0.015	0.987	0.981	542948
COO	0.066	0.038	0.927	0.891	1023112	0.056	0.032	0.945	0.918	539061	0.079	0.047	0.902	0.852	480441
CRE	0.046	0.025	0.967	0.944	943949	0.040	0.023	0.969	0.949	505831	0.054	0.027	0.963	0.939	432543
CUR	0.051	0.025	0.968	0.946	922821	0.044	0.022	0.973	0.956	488441	0.064	0.030	0.956	0.926	430077
EMO	0.069	0.040	0.916	0.873	1232835	0.069	0.043	0.897	0.846	649334	0.073	0.038	0.925	0.888	580688
EMP	0.055	0.033	0.939	0.898	937727	0.043	0.026	0.961	0.935	504949	0.075	0.042	0.908	0.846	427145
ENE	0.069	0.038	0.919	0.878	1233121	0.069	0.043	0.888	0.832	644847	0.073	0.037	0.926	0.889	580251
OPT	0.053	0.031	0.967	0.951	1157233	0.049	0.032	0.961	0.941	607607	0.060	0.031	0.967	0.951	542310
PER	0.055	0.031	0.954	0.931	1101171	0.057	0.035	0.941	0.912	588254	0.058	0.030	0.959	0.939	504352
RES	0.067	0.037	0.923	0.871	989267	0.069	0.038	0.909	0.848	528896	0.066	0.034	0.936	0.893	455545
SEL	0.030	0.015	0.986	0.977	964500	0.029	0.016	0.986	0.977	513178	0.038	0.018	0.981	0.968	448271
SOC	0.061	0.032	0.944	0.907	1014823	0.048	0.027	0.955	0.925	525948	0.061	0.029	0.956	0.927	477331
STR	0.064	0.028	0.958	0.929	1084019	0.054	0.026	0.962	0.936	572600	0.077	0.030	0.950	0.917	505580
TOL	0.040	0.024	0.962	0.943	1174045	0.035	0.022	0.966	0.949	631168	0.050	0.029	0.953	0.929	534529
TRU	0.045	0.022	0.976	0.961	1013134	0.047	0.025	0.967	0.946	534400	0.048	0.021	0.978	0.963	471568
EFF	0.036	0.025	0.957	0.940	1285583	0.035	0.024	0.962	0.947	677067	0.041	0.027	0.946	0.924	601769
MOT	0.032	0.017	0.984	0.973	938794	0.029	0.015	0.987	0.978	494592	0.042	0.023	0.975	0.958	439235

Table G.6. CFA model-data fit indices of parent assessment scales

Scale	All					Younger					Older				
	RMSEA	SRMR	CFI	TLI	BIC	RMSEA	SRMR	CFI	TLI	BIC	RMSEA	SRMR	CFI	TLI	BIC
ASS	0.041	0.023	0.974	0.961	531425	0.039	0.024	0.972	0.958	271986	0.043	0.022	0.977	0.965	258507
COO	0.062	0.038	0.927	0.890	421226	0.058	0.038	0.923	0.885	211707	0.066	0.037	0.930	0.896	208977
CRE	0.048	0.025	0.966	0.944	393988	0.044	0.025	0.966	0.944	197952	0.049	0.025	0.970	0.949	194687
CUR	0.055	0.034	0.942	0.919	513019	0.052	0.035	0.937	0.912	252884	0.055	0.033	0.949	0.928	257066
EMO	0.088	0.051	0.884	0.807	458793	0.088	0.055	0.864	0.773	233816	0.085	0.046	0.910	0.850	224293
EMP	0.061	0.036	0.936	0.893	399308	0.055	0.035	0.935	0.892	203353	0.065	0.035	0.938	0.897	195550
ENE	0.046	0.033	0.943	0.914	541403	0.038	0.031	0.945	0.918	269143	0.055	0.034	0.938	0.907	269092
OPT	0.052	0.037	0.941	0.918	532303	0.043	0.035	0.945	0.923	265578	0.060	0.038	0.939	0.914	265158
PER	0.063	0.042	0.941	0.918	552891	0.060	0.043	0.939	0.915	284371	0.066	0.040	0.944	0.921	268194
RES	0.063	0.034	0.941	0.902	445590	0.057	0.034	0.942	0.903	227210	0.072	0.034	0.938	0.897	218346
SEL	0.035	0.018	0.984	0.973	426726	0.033	0.019	0.984	0.974	218008	0.039	0.018	0.983	0.971	208421
SOC	0.034	0.017	0.982	0.969	431192	0.031	0.019	0.979	0.965	213319	0.038	0.018	0.981	0.969	215413
STR	0.031	0.019	0.982	0.974	530628	0.027	0.019	0.984	0.977	271905	0.035	0.019	0.982	0.972	257964
TOL	0.050	0.033	0.951	0.926	499734	0.046	0.032	0.951	0.926	251736	0.056	0.034	0.949	0.924	247607
TRU	0.037	0.023	0.974	0.960	494296	0.037	0.026	0.965	0.948	247062	0.038	0.021	0.978	0.967	246036
EFF	0.048	0.032	0.935	0.909	542253	0.043	0.032	0.937	0.912	273426	0.051	0.032	0.939	0.915	267831
MOT	0.040	0.021	0.982	0.970	400625	0.036	0.020	0.984	0.974	203544	0.040	0.019	0.984	0.974	195905

Table G.7. Invariance level for student assessment group=cohort

Scale	Invariance Level	Values					Difference				
		RMSEA	SRMR	CFI	TLI	BIC	ΔRMSEA	ΔSRMR	ΔCFI	ΔTLI	ΔBIC
ASS	Configural	0.036	0.021	0.984	0.977	1105622					
	Metric	0.036	0.028	0.981	0.977	1105935	0.000	0.007	0.003	0.000	313
	Scalar	0.061	0.039	0.938	0.935	1111573	0.025	0.011	0.043	0.042	5638
COO	Configural	0.063	0.042	0.928	0.899	928278					
	Metric	0.060	0.056	0.921	0.906	928883	-0.003	0.014	0.007	-0.007	605
	Scalar	0.062	0.063	0.906	0.901	930138	0.002	0.007	0.015	0.005	1255
CRE	Configural	0.043	0.026	0.969	0.953	851657					
	Metric	0.040	0.027	0.968	0.960	851667	-0.003	0.001	0.001	-0.007	10
	Scalar	0.043	0.033	0.955	0.952	852421	0.003	0.006	0.013	0.008	754
CUR	Configural	0.047	0.026	0.969	0.953	835537					
	Metric	0.045	0.037	0.965	0.957	835831	-0.002	0.011	0.004	-0.004	294
	Scalar	0.046	0.039	0.957	0.954	836336	0.001	0.002	0.008	0.003	505
EMO	Configural	0.057	0.057	0.942	0.919	1113616					
	Metric	0.054	0.060	0.939	0.927	1113868	-0.003	0.003	0.003	-0.008	252
	Scalar	0.052	0.061	0.936	0.932	1113961	-0.002	0.001	0.003	-0.005	94
EMP	Configural	0.055	0.036	0.931	0.897	854258					
	Metric	0.050	0.040	0.929	0.912	854373	-0.005	0.004	0.002	-0.015	115
	Scalar	0.056	0.050	0.901	0.893	855700	0.006	0.010	0.028	0.019	1327
ENE	Configural	0.055	0.032	0.948	0.928	1104207					
	Metric	0.052	0.036	0.946	0.936	1104306	-0.003	0.004	0.002	-0.008	99
	Scalar	0.056	0.042	0.930	0.926	1105709	0.004	0.006	0.016	0.010	1404
OPT	Configural	0.039	0.037	0.980	0.973	1037239					
	Metric	0.038	0.042	0.978	0.974	1037439	-0.001	0.005	0.002	-0.001	200
	Scalar	0.043	0.045	0.969	0.967	1038595	0.005	0.003	0.009	0.007	1155
PER	Configural	0.047	0.038	0.966	0.953	985389					
	Metric	0.044	0.042	0.965	0.958	985453	-0.003	0.004	0.001	-0.005	64
	Scalar	0.047	0.048	0.954	0.952	986514	0.003	0.006	0.011	0.006	1061
RES	Configural	0.045	0.046	0.967	0.950	883676					
	Metric	0.045	0.058	0.959	0.950	884144	0.000	0.012	0.008	0.000	468
	Scalar	0.049	0.060	0.944	0.940	885111	0.004	0.002	0.015	0.010	968
SEL	Configural	0.047	0.059	0.965	0.948	869963					
	Metric	0.043	0.062	0.964	0.955	870027	-0.004	0.003	0.001	-0.007	63
	Scalar	0.043	0.063	0.958	0.955	870299	0.000	0.001	0.006	0.000	272
SOC	Configural	0.053	0.034	0.951	0.926	917249					
	Metric	0.051	0.044	0.945	0.932	917648	-0.002	0.010	0.006	-0.006	399
	Scalar	0.075	0.064	0.860	0.849	923344	0.024	0.020	0.085	0.083	5696
STR	Configural	0.064	0.049	0.952	0.928	978757					
	Metric	0.059	0.052	0.949	0.937	978857	-0.005	0.003	0.003	-0.009	100
	Scalar	0.061	0.057	0.938	0.933	979743	0.002	0.005	0.011	0.004	886
TOL	Configural	0.042	0.027	0.958	0.942	1064375					
	Metric	0.039	0.030	0.956	0.948	1064460	-0.003	0.003	0.002	-0.006	85
	Scalar	0.049	0.039	0.925	0.920	1066600	0.010	0.009	0.031	0.028	2140
TRU	Configural	0.051	0.035	0.964	0.947	918757					
	Metric	0.048	0.040	0.961	0.952	918927	-0.003	0.005	0.003	-0.005	170
	Scalar	0.050	0.042	0.952	0.948	919665	0.002	0.002	0.009	0.004	739
EFF	Configural	0.042	0.037	0.939	0.919	1172653					
	Metric	0.042	0.044	0.932	0.921	1173025	0.000	0.007	0.007	-0.002	372

	Scalar	0.052	0.051	0.884	0.879	1176133	0.010	0.007	0.048	0.042	3108
MOT	Configural	0.033	0.021	0.979	0.969	856490					
	Metric	0.034	0.032	0.974	0.968	856772	0.001	0.011	0.005	0.001	282
	Scalar	0.044	0.042	0.950	0.947	858236	0.010	0.010	0.024	0.021	1464

Table G.8. Invariance level for student assessment group=site

Scale	Invariance Level	Values					Difference				
		RMSEA	SRMR	CFI	TLI	BIC	ΔRMSEA	ΔSRMR	ΔCFI	ΔTLI	ΔBIC
ASS	Configural	0.054	0.033	0.962	0.947	1095136					
	Metric	0.064	0.076	0.931	0.926	1098675	0.010	0.043	0.031	0.021	3539
	Scalar	0.091	0.110	0.834	0.852	1110603	0.027	0.034	0.097	0.074	11928
COO	Configural	0.063	0.044	0.928	0.899	920487					
	Metric	0.057	0.055	0.923	0.918	920480	-0.006	0.011	0.005	-0.019	-6
	Scalar	0.071	0.071	0.857	0.873	926205	0.014	0.016	0.066	0.045	5725
CRE	Configural	0.044	0.032	0.967	0.951	848966					
	Metric	0.042	0.047	0.959	0.955	849119	-0.002	0.015	0.008	-0.004	152
	Scalar	0.059	0.064	0.899	0.911	852848	0.017	0.017	0.060	0.044	3729
CUR	Configural	0.053	0.037	0.959	0.938	814028					
	Metric	0.054	0.073	0.941	0.936	815212	0.001	0.036	0.018	0.002	1184
	Scalar	0.089	0.123	0.801	0.824	826706	0.035	0.050	0.140	0.112	11494
EMO	Configural	0.062	0.061	0.937	0.911	1102614					
	Metric	0.062	0.078	0.916	0.910	1104143	0.000	0.017	0.021	0.001	1530
	Scalar	0.072	0.084	0.865	0.880	1108423	0.010	0.006	0.051	0.030	4280
EMP	Configural	0.057	0.037	0.933	0.899	836101					
	Metric	0.053	0.058	0.919	0.912	836614	-0.004	0.021	0.014	-0.013	513
	Scalar	0.091	0.106	0.710	0.744	848427	0.038	0.048	0.209	0.168	11812
ENE	Configural	0.064	0.041	0.936	0.911	1101116					
	Metric	0.060	0.055	0.926	0.921	1101520	-0.004	0.014	0.010	-0.010	403
	Scalar	0.076	0.069	0.854	0.871	1108076	0.016	0.014	0.072	0.050	6556
OPT	Configural	0.046	0.046	0.974	0.964	1027081					
	Metric	0.049	0.072	0.961	0.958	1028284	0.003	0.026	0.013	0.006	1203
	Scalar	0.078	0.088	0.881	0.895	1038292	0.029	0.016	0.080	0.063	10008
PER	Configural	0.047	0.041	0.968	0.955	975691					
	Metric	0.054	0.072	0.946	0.942	977784	0.007	0.031	0.022	0.013	2093
	Scalar	0.074	0.090	0.876	0.890	985580	0.020	0.018	0.070	0.052	7796
RES	Configural	0.050	0.046	0.960	0.939	875860					
	Metric	0.048	0.059	0.947	0.942	876298	-0.002	0.013	0.013	-0.003	437
	Scalar	0.071	0.081	0.858	0.875	882161	0.023	0.022	0.089	0.067	5863
SEL	Configural	0.047	0.058	0.966	0.950	858815					
	Metric	0.051	0.079	0.946	0.941	859991	0.004	0.021	0.020	0.009	1176
	Scalar	0.078	0.104	0.841	0.859	867686	0.027	0.025	0.105	0.082	7695
SOC	Configural	0.066	0.045	0.932	0.897	916793					
	Metric	0.062	0.065	0.917	0.910	917448	-0.004	0.020	0.015	-0.013	654
	Scalar	0.084	0.092	0.812	0.834	924650	0.022	0.027	0.105	0.076	7202
STR	Configural	0.069	0.053	0.946	0.919	970978					
	Metric	0.071	0.073	0.922	0.915	972741	0.002	0.020	0.024	0.004	1763
	Scalar	0.091	0.107	0.842	0.861	979993	0.020	0.034	0.080	0.054	7252
TOL	Configural	0.047	0.040	0.946	0.925	1048180					
	Metric	0.051	0.072	0.918	0.913	1049720	0.004	0.032	0.028	0.012	1540
	Scalar	0.083	0.112	0.734	0.764	1062019	0.032	0.040	0.184	0.149	12298
TRU	Configural	0.052	0.033	0.966	0.949	907067					
	Metric	0.050	0.052	0.955	0.951	907587	-0.002	0.019	0.011	-0.002	520
	Scalar	0.074	0.083	0.879	0.893	913957	0.024	0.031	0.076	0.058	6370
EFF	Configural	0.047	0.040	0.930	0.906	1153939					
	Metric	0.045	0.052	0.917	0.913	1154204	-0.002	0.012	0.013	-0.007	264
	Scalar	0.084	0.101	0.661	0.698	1171833	0.039	0.049	0.256	0.215	17629
MOT	Configural	0.036	0.025	0.977	0.965	840454					
	Metric	0.046	0.072	0.946	0.941	842173	0.010	0.047	0.031	0.024	1719
	Scalar	0.092	0.152	0.738	0.769	855714	0.046	0.080	0.208	0.172	13541

Table G.9. Invariance level for student assessment group=gender

Scale	Invariance Level	Values					Difference				
		RMSEA	SRMR	CFI	TLI	BIC	ΔRMSEA	ΔSRMR	ΔCFI	ΔTLI	ΔBIC
ASS	Configural	0.040	0.022	0.978	0.970	1120945					
	Metric	0.038	0.023	0.977	0.973	1120929	-0.002	0.001	0.001	-0.003	-15
	Scalar	0.038	0.025	0.974	0.972	1121254	0.000	0.002	0.003	0.001	324
COO	Configural	0.061	0.041	0.931	0.903	930392					
	Metric	0.057	0.044	0.929	0.916	930409	-0.004	0.003	0.002	-0.013	17
	Scalar	0.057	0.049	0.920	0.916	930999	0.000	0.005	0.009	0.000	590
CRE	Configural	0.042	0.026	0.969	0.954	856537					
	Metric	0.038	0.028	0.969	0.961	856513	-0.004	0.002	0.000	-0.007	-25
	Scalar	0.037	0.027	0.966	0.964	856568	-0.001	-0.001	0.003	-0.003	55
CUR	Configural	0.043	0.024	0.974	0.961	837777					
	Metric	0.039	0.026	0.973	0.967	837796	-0.004	0.002	0.001	-0.006	19
	Scalar	0.046	0.035	0.957	0.954	839036	0.007	0.009	0.016	0.013	1240
EMO	Configural	0.054	0.055	0.946	0.925	1113914					
	Metric	0.050	0.055	0.945	0.935	1113873	-0.004	0.000	0.001	-0.010	-41
	Scalar	0.053	0.060	0.931	0.927	1115054	0.003	0.005	0.014	0.008	1181
EMP	Configural	0.051	0.034	0.936	0.904	858200					
	Metric	0.046	0.037	0.935	0.920	858216	-0.005	0.003	0.001	-0.016	16
	Scalar	0.047	0.040	0.922	0.916	858693	0.001	0.003	0.013	0.004	477
ENE	Configural	0.052	0.029	0.955	0.938	1108102					
	Metric	0.048	0.029	0.955	0.946	1108052	-0.004	0.000	0.000	-0.008	-50
	Scalar	0.058	0.042	0.924	0.920	1110893	0.010	0.013	0.031	0.026	2841
OPT	Configural	0.036	0.035	0.983	0.976	1042745					
	Metric	0.034	0.036	0.982	0.979	1042716	-0.002	0.001	0.001	-0.003	-28
	Scalar	0.036	0.037	0.978	0.976	1043238	0.002	0.001	0.004	0.003	522
PER	Configural	0.045	0.037	0.969	0.956	992375					
	Metric	0.042	0.037	0.968	0.962	992315	-0.003	0.000	0.001	-0.006	-61
	Scalar	0.044	0.040	0.961	0.959	992998	0.002	0.003	0.007	0.003	683
RES	Configural	0.043	0.046	0.968	0.953	887718					
	Metric	0.039	0.046	0.967	0.960	887678	-0.004	0.000	0.001	-0.007	-40
	Scalar	0.039	0.048	0.963	0.960	887898	0.000	0.002	0.004	0.000	220
SEL	Configural	0.044	0.057	0.969	0.953	872395					
	Metric	0.040	0.058	0.969	0.961	872357	-0.004	0.001	0.000	-0.008	-38
	Scalar	0.040	0.059	0.963	0.961	872625	0.000	0.001	0.006	0.000	269
SOC	Configural	0.056	0.035	0.944	0.916	927011					
	Metric	0.051	0.040	0.943	0.930	927048	-0.005	0.005	0.001	-0.014	37
	Scalar	0.053	0.046	0.929	0.924	927835	0.002	0.006	0.014	0.006	787
STR	Configural	0.060	0.048	0.955	0.933	980339					
	Metric	0.056	0.050	0.953	0.942	980410	-0.004	0.002	0.002	-0.009	71
	Scalar	0.057	0.054	0.942	0.938	981217	0.001	0.004	0.011	0.004	807
TOL	Configural	0.039	0.026	0.962	0.947	1069359					
	Metric	0.036	0.029	0.960	0.953	1069390	-0.003	0.003	0.002	-0.006	31
	Scalar	0.039	0.036	0.950	0.947	1069981	0.003	0.007	0.010	0.006	591
TRU	Configural	0.049	0.035	0.967	0.950	925270					
	Metric	0.045	0.036	0.966	0.958	925231	-0.004	0.001	0.001	-0.008	-39
	Scalar	0.044	0.037	0.963	0.960	925336	-0.001	0.001	0.003	-0.002	106
EFF	Configural	0.041	0.036	0.942	0.923	1177813					
	Metric	0.039	0.038	0.940	0.930	1177830	-0.002	0.002	0.002	-0.007	17
	Scalar	0.042	0.043	0.922	0.919	1178891	0.003	0.005	0.018	0.011	1061
MOT	Configural	0.030	0.018	0.983	0.975	860788					
	Metric	0.028	0.024	0.982	0.978	860819	-0.002	0.006	0.001	-0.003	31
	Scalar	0.031	0.030	0.975	0.973	861210	0.003	0.006	0.007	0.005	391

Table G.10. Invariance level for parent assessment group=cohort

Scale	Invariance Level	Values					Difference				
		RMSEA	SRMR	CFI	TLI	BIC	ΔRMSEA	ΔSRMR	ΔCFI	ΔTLI	ΔBIC
ASS	Configural	0.036	0.027	0.975	0.965	522819					
	Metric	0.034	0.029	0.974	0.969	522794	-0.002	0.002	0.001	-0.004	-25
	Scalar	0.040	0.034	0.959	0.957	523485	0.006	0.005	0.015	0.012	691
COO	Configural	0.056	0.053	0.929	0.900	411986					
	Metric	0.053	0.057	0.926	0.911	411992	-0.003	0.004	0.003	-0.011	6
	Scalar	0.052	0.060	0.918	0.913	412194	-0.001	0.003	0.008	-0.002	202
CRE	Configural	0.044	0.032	0.966	0.949	383499					
	Metric	0.040	0.034	0.965	0.957	383477	-0.004	0.002	0.001	-0.008	-22
	Scalar	0.047	0.046	0.946	0.942	384181	0.007	0.012	0.019	0.015	705
CUR	Configural	0.052	0.039	0.937	0.916	501208					
	Metric	0.049	0.043	0.935	0.925	501181	-0.003	0.004	0.002	-0.009	-27
	Scalar	0.056	0.060	0.907	0.903	502717	0.007	0.017	0.028	0.022	1536
EMO	Configural	0.047	0.064	0.963	0.945	442141					
	Metric	0.043	0.065	0.962	0.954	442121	-0.004	0.001	0.001	-0.009	-19
	Scalar	0.053	0.072	0.935	0.930	443257	0.010	0.007	0.027	0.024	1135
EMP	Configural	0.051	0.033	0.937	0.905	395664					
	Metric	0.047	0.035	0.935	0.920	395632	-0.004	0.002	0.002	-0.015	-31
	Scalar	0.048	0.043	0.921	0.914	395942	0.001	0.008	0.014	0.006	310
ENE	Configural	0.036	0.036	0.965	0.951	522592					
	Metric	0.034	0.038	0.963	0.956	522569	-0.002	0.002	0.002	-0.005	-23
	Scalar	0.039	0.050	0.943	0.940	523267	0.005	0.012	0.020	0.016	698
OPT	Configural	0.042	0.065	0.957	0.943	514773					
	Metric	0.040	0.068	0.956	0.949	514755	-0.002	0.003	0.001	-0.006	-17
	Scalar	0.043	0.074	0.944	0.942	515353	0.003	0.006	0.012	0.007	598
PER	Configural	0.037	0.039	0.978	0.971	531550					
	Metric	0.035	0.040	0.977	0.974	531497	-0.002	0.001	0.001	-0.003	-53
	Scalar	0.034	0.040	0.975	0.974	531548	-0.001	0.000	0.002	0.000	51
RES	Configural	0.044	0.048	0.967	0.950	432144					
	Metric	0.041	0.049	0.966	0.958	432103	-0.003	0.001	0.001	-0.008	-42
	Scalar	0.039	0.050	0.964	0.961	432076	-0.002	0.001	0.002	-0.003	-27
SEL	Configural	0.042	0.046	0.972	0.957	415411					
	Metric	0.039	0.048	0.971	0.964	415380	-0.003	0.002	0.001	-0.007	-31
	Scalar	0.038	0.048	0.969	0.966	415383	-0.001	0.000	0.002	-0.002	3
SOC	Configural	0.039	0.044	0.967	0.950	421570					
	Metric	0.036	0.046	0.965	0.957	421565	-0.003	0.002	0.002	-0.007	-5
	Scalar	0.050	0.057	0.925	0.919	422950	0.014	0.011	0.040	0.038	1385
STR	Configural	0.035	0.046	0.975	0.966	517606					
	Metric	0.033	0.047	0.974	0.969	517582	-0.002	0.001	0.001	-0.003	-24
	Scalar	0.037	0.048	0.964	0.962	518060	0.004	0.001	0.010	0.007	478
TOL	Configural	0.048	0.049	0.948	0.927	490229					
	Metric	0.045	0.048	0.946	0.936	490201	-0.003	-0.001	0.002	-0.009	-28
	Scalar	0.045	0.051	0.940	0.936	490383	0.000	0.003	0.006	0.000	182
TRU	Configural	0.045	0.066	0.954	0.935	484929					
	Metric	0.042	0.067	0.953	0.944	484891	-0.003	0.001	0.001	-0.009	-38
	Scalar	0.042	0.070	0.946	0.943	485118	0.000	0.003	0.007	0.001	227
EFF	Configural	0.046	0.038	0.926	0.902	535558					
	Metric	0.044	0.041	0.925	0.913	535531	-0.002	0.003	0.001	-0.011	-27
	Scalar	0.048	0.051	0.901	0.896	536405	0.004	0.010	0.024	0.017	874
MOT	Configural	0.036	0.025	0.981	0.971	395247					
	Metric	0.034	0.030	0.979	0.974	395242	-0.002	0.005	0.002	-0.003	-5
	Scalar	0.046	0.049	0.956	0.953	396300	0.012	0.019	0.023	0.021	1058

Table G.11. Invariance level for parent assessment group=site

Scale	Invariance Level	Values					Difference				
		RMSEA	SRMR	CFI	TLI	BIC	ΔRMSEA	ΔSRMR	ΔCFI	ΔTLI	ΔBIC
ASS	Configural	0.051	0.031	0.963	0.948	499749					
	Metric	0.060	0.080	0.935	0.928	501225	0.009	0.049	0.028	0.020	1475
	Scalar	0.096	0.112	0.802	0.818	508940	0.036	0.032	0.133	0.110	7715
COO	Configural	0.066	0.058	0.919	0.886	403147					
	Metric	0.061	0.073	0.910	0.902	403317	-0.005	0.015	0.009	-0.016	170
	Scalar	0.066	0.090	0.876	0.886	404619	0.005	0.017	0.034	0.016	1302
CRE	Configural	0.050	0.036	0.963	0.945	373183					
	Metric	0.047	0.050	0.956	0.951	373275	-0.003	0.014	0.007	-0.006	93
	Scalar	0.071	0.072	0.881	0.890	376173	0.024	0.022	0.075	0.061	2898
CUR	Configural	0.070	0.060	0.901	0.868	486826					
	Metric	0.065	0.071	0.894	0.886	486906	-0.005	0.011	0.007	-0.018	79
	Scalar	0.091	0.110	0.763	0.781	494604	0.026	0.039	0.131	0.105	7698
EMO	Configural	0.059	0.068	0.957	0.936	427936					
	Metric	0.056	0.079	0.949	0.943	428122	-0.003	0.011	0.008	-0.007	186
	Scalar	0.073	0.097	0.895	0.903	430496	0.017	0.018	0.054	0.040	2375
EMP	Configural	0.069	0.040	0.910	0.866	386214					
	Metric	0.064	0.062	0.895	0.883	386525	-0.005	0.022	0.015	-0.017	310
	Scalar	0.079	0.083	0.810	0.825	388997	0.015	0.021	0.085	0.058	2472
ENE	Configural	0.047	0.052	0.957	0.939	510097					
	Metric	0.048	0.065	0.942	0.936	510469	0.001	0.013	0.015	0.003	372
	Scalar	0.069	0.084	0.859	0.870	513825	0.021	0.019	0.083	0.066	3356
OPT	Configural	0.055	0.074	0.941	0.922	496336					
	Metric	0.053	0.085	0.932	0.927	496617	-0.002	0.011	0.009	-0.005	281
	Scalar	0.080	0.118	0.822	0.836	503341	0.027	0.033	0.110	0.091	6725
PER	Configural	0.050	0.063	0.968	0.957	514035					
	Metric	0.052	0.087	0.956	0.952	514857	0.002	0.024	0.012	0.005	822
	Scalar	0.071	0.104	0.906	0.913	519227	0.019	0.017	0.050	0.039	4370
RES	Configural	0.055	0.063	0.960	0.939	420580					
	Metric	0.051	0.069	0.954	0.949	420594	-0.004	0.006	0.006	-0.010	14
	Scalar	0.067	0.087	0.904	0.911	422725	0.016	0.018	0.050	0.038	2131
SEL	Configural	0.055	0.068	0.961	0.942	402024					
	Metric	0.058	0.086	0.941	0.935	402768	0.003	0.018	0.020	0.007	744
	Scalar	0.097	0.117	0.802	0.818	409216	0.039	0.031	0.139	0.117	6448
SOC	Configural	0.045	0.043	0.969	0.953	412372					
	Metric	0.042	0.050	0.963	0.959	412359	-0.003	0.007	0.006	-0.006	-13
	Scalar	0.065	0.073	0.894	0.902	414861	0.023	0.023	0.069	0.057	2502
STR	Configural	0.045	0.054	0.971	0.959	499667					
	Metric	0.055	0.082	0.944	0.939	501049	0.010	0.028	0.027	0.020	1382
	Scalar	0.070	0.100	0.893	0.902	503908	0.015	0.018	0.051	0.037	2859
TOL	Configural	0.055	0.052	0.941	0.918	468136					
	Metric	0.059	0.084	0.916	0.908	469055	0.004	0.032	0.025	0.010	919
	Scalar	0.095	0.145	0.739	0.759	476636	0.036	0.061	0.177	0.149	7582
TRU	Configural	0.045	0.055	0.963	0.948	457932					
	Metric	0.049	0.082	0.944	0.939	458601	0.004	0.027	0.019	0.009	670
	Scalar	0.076	0.122	0.842	0.855	463353	0.027	0.040	0.102	0.084	4752
EFF	Configural	0.049	0.035	0.931	0.908	521336					
	Metric	0.046	0.051	0.923	0.917	521384	-0.003	0.016	0.008	-0.009	48
	Scalar	0.074	0.082	0.775	0.792	527109	0.028	0.031	0.148	0.125	5725
MOT	Configural	0.039	0.024	0.978	0.968	383143					
	Metric	0.051	0.074	0.951	0.946	384183	0.012	0.050	0.027	0.022	1040
	Scalar	0.093	0.138	0.803	0.818	390584	0.042	0.064	0.148	0.128	6401

Table G.12. Item parameters of student assessment scales

Scale	Item	Slope (a_i)	Location	Step 1	Step 2	Step 3
ASS	ASS01	1.216	0.078	-1.517	-0.738	0.668
	ASS02	1.004	-0.186	-1.485	-0.719	0.691
	ASS03	0.445	-0.077	-0.868	-0.560	0.405
	ASS04	1.433	0.056	-1.661	-0.797	0.602
	ASS05	0.731	-0.241	-0.808	-0.575	0.425
	ASS06	1.495	-0.067	-1.654	-0.678	0.805
	ASS08	1.603	0.364	-2.077	-0.805	0.898
COO	COO01	1.463	-2.318	-0.688	-1.520	0.032
	COO02	1.063	-1.594	-0.941	-1.292	0.029
	COO03	1.006	-1.365	-1.081	-1.215	0.102
	COO04	0.334	-0.739	-0.568	-0.513	0.245
	COO06	1.767	-2.179	-1.850	-1.486	0.383
	COO07	1.676	-1.978	-1.862	-1.402	0.436
	COO08	1.060	-1.638	-0.787	-1.475	0.140
CRE	CRE01	0.890	-1.209	-1.296	-1.037	0.091
	CRE02	1.422	-1.691	-2.107	-1.233	0.616
	CRE04	0.882	-1.103	-1.182	-1.107	0.271
	CRE05	1.293	-1.900	-1.425	-1.037	0.308
	CRE06	0.860	-1.265	-0.752	-0.734	0.198
	CRE07	0.594	-0.580	-1.219	-0.461	0.399
CUR	CUR02	1.622	-1.989	-1.572	-1.463	0.442
	CUR04	0.832	-1.451	-0.802	-0.897	-0.157
	CUR05	1.953	-2.996	-1.384	-1.695	0.133
	CUR06	0.712	-0.960	-0.249	-0.691	-0.027
	CUR07	2.133	-2.482	-2.059	-1.664	0.513
	CUR08	0.586	-0.704	-0.491	-0.562	0.181
EMO	EMO01	0.533	-0.325	-1.186	-0.323	0.017
	EMO02	0.859	-0.939	-1.375	-0.727	0.130
	EMO03	1.341	-0.566	-1.383	-0.609	0.290
	EMO04	1.058	-0.854	-1.101	-0.802	0.063
	EMO06	0.433	0.079	-0.683	-0.490	0.346
	EMO07	0.481	-0.302	-0.872	-0.695	0.293
	EMO08	1.127	-0.761	-1.229	-0.594	0.275
EMP	EMP01	0.641	-0.977	-0.563	-1.090	0.035
	EMP02	0.606	-1.306	0.277	-0.850	-0.281
	EMP03	1.064	-1.090	-1.221	-0.972	0.125
	EMP04	1.160	-1.112	-1.363	-0.912	0.134
	EMP06	1.048	-0.884	-1.438	-1.217	0.405
	EMP07	0.744	-0.758	-0.627	-1.266	0.241
ENE	ENE01	1.264	-1.783	-1.635	-0.869	0.571
	ENE02	0.628	-0.939	-0.931	-1.000	0.491
	ENE03	0.550	-0.563	-1.058	-0.190	0.194
	ENE04	0.994	-1.127	-1.233	-0.536	0.317
	ENE05	0.484	-0.642	-0.320	-0.348	0.185
	ENE06	0.540	-0.405	-0.699	-0.258	0.117
	ENE08	1.315	-1.292	-1.862	-0.782	0.665
OPT	OPT01	0.446	-0.421	-0.799	-0.163	0.026
	OPT02	0.809	-1.199	-0.921	-0.799	0.216
	OPT03	1.071	-0.686	-1.372	-0.606	0.477
	OPT05	1.773	-2.589	-1.650	-1.246	0.421

	OPT06	1.417	-1.688	-1.640	-1.130	0.600
	OPT07	2.266	-2.934	-1.958	-1.510	0.723
	OPT08	0.500	-0.464	-0.865	-0.292	0.304
PER	PER01	1.583	-1.941	-2.142	-1.173	0.763
	PER02	1.611	-2.019	-2.256	-1.173	0.664
	PER03	0.515	-0.814	-0.648	-0.351	0.061
	PER04	1.698	-2.051	-2.256	-1.428	0.774
	PER05	0.894	-1.017	-1.415	-0.733	0.552
	PER07	0.695	-0.790	-0.775	-0.767	0.320
	PER08	1.267	-1.467	-1.657	-1.363	0.532
RES	RES01	0.515	-0.203	-1.783	0.154	0.393
	RES03	0.877	-0.602	-1.550	-0.363	0.252
	RES04	0.833	-0.934	-1.139	-0.623	0.308
	RES05	0.781	-1.462	-0.248	-1.407	-0.175
	RES06	1.100	-1.419	-1.112	-1.455	0.420
	RES08	0.969	-0.788	-1.623	-0.530	0.311
SEL	SEL01	0.656	-0.705	-0.880	-0.703	0.074
	SEL02	0.855	-1.239	-0.805	-1.005	-0.172
	SEL03	1.734	-1.593	-2.549	-1.320	0.680
	SEL04	0.775	-0.790	-1.344	-0.988	0.402
	SEL07	1.297	-1.054	-1.716	-1.185	0.512
	SEL08	0.670	-0.449	-1.179	-0.391	0.283
SOC	SOC01	0.702	-0.929	-0.910	-0.715	0.262
	SOC02	1.418	-1.712	-1.580	-0.540	0.251
	SOC03	0.924	-2.067	-0.145	-0.896	-0.262
	SOC04	0.325	-0.252	-0.322	-0.383	0.583
	SOC05	0.585	-0.826	-0.661	-0.602	0.093
	SOC07	1.461	-1.540	-1.607	-0.813	0.221
STR	STR01	0.456	-0.389	-0.988	-0.409	0.164
	STR02	0.924	0.083	-1.210	-0.396	0.458
	STR03	0.973	0.356	-1.476	-0.239	0.373
	STR04	0.818	-0.376	-1.039	-0.300	0.172
	STR05	1.091	-0.779	-1.298	-0.428	0.236
	STR07	1.299	-0.125	-1.983	-0.329	0.512
TOL	TOL01	0.840	-1.182	-0.562	-0.658	-0.046
	TOL02	0.957	-0.572	-1.502	-0.703	0.392
	TOL03	0.622	-0.522	-0.967	-0.874	0.504
	TOL04	0.568	-1.114	0.290	-0.263	-0.093
	TOL05	1.354	-0.985	-1.191	-1.153	0.595
	TOL06	0.598	-0.755	-0.140	-0.500	-0.041
	TOL07	0.733	-0.636	-0.836	-1.081	0.249
TRU	TRU01	0.696	-0.425	-1.461	-0.603	0.377
	TRU02	0.736	-0.697	-0.542	-0.708	0.214
	TRU05	0.724	-0.532	-1.009	-0.767	0.561
	TRU06	0.972	-1.019	-1.212	-1.052	0.167
	TRU07	1.530	-0.834	-2.089	-0.901	0.531
	TRU08	1.916	-1.354	-2.449	-1.500	0.726
EFF	ASS01	0.364	-0.030	-0.649	-0.467	0.409
	CRE02	0.925	-1.200	-1.522	-0.990	0.467
	CUR05	1.057	-1.783	-0.530	-1.261	-0.027
	EMP06	0.584	-0.703	-0.931	-1.014	0.270
	OPT04	0.701	-0.846	-0.882	-0.626	0.349

	PER04	0.954	-1.233	-1.374	-1.076	0.548
	RES02	0.749	-1.237	-0.816	-0.954	0.180
	SEL02	0.687	-1.084	-0.603	-0.928	-0.233
MOT	CUR02	1.034	-1.417	-0.984	-1.176	0.316
	ENE02	0.618	-0.838	-0.886	-0.991	0.473
	PER02	1.356	-1.717	-1.970	-1.052	0.597
	PER08	1.267	-1.417	-1.669	-1.357	0.543
	RES06	1.218	-1.588	-1.284	-1.509	0.469
	SEL06	0.636	-0.846	-0.967	-0.786	0.284

Table G.13. Item parameters of parent assessment scales

Scale	Item	Slope (Tau)	Location	Step 1	Step 2	Step 3
ASS	ASS01	1.211	-0.737	-2.343	-0.626	0.717
	ASS02	1.035	-0.810	-2.203	-0.622	0.548
	ASS03	0.485	-0.478	-1.472	-0.468	0.396
	ASS04	1.802	-0.900	-2.996	-0.665	0.696
	ASS05	0.579	-0.391	-1.387	-0.253	0.032
	ASS06	1.668	-1.123	-2.853	-0.753	0.683
	ASS08	1.066	-0.044	-2.344	-0.378	0.537
COO	COO01	1.500	-2.380	-0.801	-1.279	-0.247
	COO02	1.487	-2.502	-1.507	-1.374	-0.211
	COO03	1.429	-2.185	-2.118	-1.219	-0.024
	COO04	0.381	-0.878	-0.690	-0.008	-0.443
	COO06	2.070	-3.016	-2.482	-1.832	0.217
	COO07	1.746	-2.451	-2.324	-1.461	0.077
	COO08	1.396	-2.394	-1.298	-1.120	-0.220
CRE	CRE01	1.058	-1.363	-2.127	-0.605	0.234
	CRE02	1.622	-2.112	-2.551	-1.223	0.514
	CRE04	1.407	-1.847	-2.424	-1.338	0.360
	CRE05	1.771	-2.347	-2.877	-1.215	0.476
	CRE06	1.472	-2.303	-1.886	-1.244	0.183
	CRE07	0.767	-0.801	-1.894	-0.293	-0.059
CUR	CUR01	1.322	-2.326	-1.776	-0.962	0.064
	CUR02	1.362	-1.878	-1.884	-1.084	0.473
	CUR03	1.184	-1.874	-1.712	-0.716	-0.011
	CUR04	1.602	-2.521	-2.274	-1.262	0.049
	CUR05	2.260	-3.538	-2.105	-1.917	-0.016
	CUR06	0.496	-0.748	-0.496	-0.122	-0.556
	CUR07	1.915	-2.959	-2.256	-1.537	0.005
	CUR08	0.820	-1.133	-1.472	-0.629	0.123
EMO	EMO02	0.837	-0.700	-2.218	-0.523	0.303
	EMO03	1.060	-0.023	-1.855	-0.269	0.029
	EMO04	1.053	-0.406	-2.208	-0.828	0.242
	EMO05	0.936	-0.181	-1.680	-0.385	-0.088
	EMO07	0.643	-0.295	-1.885	-0.618	0.288
	EMO08	1.293	-0.851	-2.202	-0.429	-0.029
EMP	EMP01	0.777	-1.402	-0.770	-0.617	-0.190
	EMP03	1.268	-1.395	-2.306	-1.041	0.268
	EMP04	1.204	-1.387	-2.271	-0.900	0.225
	EMP05	0.948	-0.500	-2.471	-0.911	0.616
	EMP06	1.398	-1.304	-2.909	-1.474	0.624
	EMP07	0.877	-1.321	-1.251	-0.941	-0.196
ENE	ENE01	1.293	-2.115	-1.839	-0.876	0.385
	ENE02	0.844	-1.434	-1.711	-0.830	0.347
	ENE03	0.495	-0.608	-1.553	0.208	-0.256
	ENE05	0.542	-0.769	-0.966	0.035	-0.074
	ENE06	0.489	-0.550	-1.096	-0.009	-0.463
	ENE07	0.481	-0.467	-1.479	0.275	-0.385
	ENE08	1.942	-2.572	-3.412	-1.230	0.601
OPT	OPT01	0.369	-0.509	-1.081	0.023	-0.457
	OPT02	1.192	-1.869	-1.638	-0.932	0.054
	OPT03	0.959	-0.990	-1.963	-0.768	0.350

	OPT04	1.517	-2.257	-2.191	-0.995	0.258
	OPT05	1.717	-2.667	-2.118	-1.307	0.145
	OPT06	1.926	-2.673	-2.857	-1.379	0.407
	OPT07	2.026	-3.252	-1.976	-1.550	0.033
	OPT08	0.368	-0.692	-1.310	-0.177	-0.183
PER	PER01	2.087	-1.599	-3.038	-0.860	0.806
	PER02	2.098	-1.687	-3.390	-0.954	0.734
	PER03	0.588	-0.555	-1.291	-0.243	-0.227
	PER04	2.529	-1.999	-4.177	-1.398	1.060
	PER05	1.202	-0.720	-2.412	-0.450	0.325
	PER06	0.752	-0.370	-2.017	-0.096	0.021
	PER07	1.196	-0.621	-2.375	-0.530	0.402
	PER08	1.962	-1.577	-3.488	-1.155	0.586
RES	RES01	0.692	-0.153	-2.027	0.509	0.048
	RES03	0.901	-0.393	-2.003	-0.118	0.116
	RES04	1.219	-0.877	-2.168	-0.376	0.044
	RES05	1.255	-1.492	-1.910	-1.004	0.129
	RES06	1.546	-1.848	-1.997	-1.122	0.291
	RES08	1.170	-0.533	-2.534	-0.261	0.146
SEL	SEL01	0.638	-0.508	-1.441	-0.200	0.083
	SEL02	1.110	-1.113	-1.990	-0.764	0.032
	SEL03	2.103	-1.165	-3.880	-1.201	0.912
	SEL04	1.022	-0.662	-2.530	-0.599	0.450
	SEL07	1.669	-1.007	-3.255	-1.189	0.574
	SEL08	0.981	-0.524	-2.165	-0.309	0.075
SOC	SOC01	1.308	-1.751	-2.013	-0.767	0.364
	SOC02	1.387	-1.672	-2.629	-0.406	0.332
	SOC03	1.019	-2.131	-0.809	-0.713	-0.594
	SOC04	0.371	-0.441	-0.933	-0.134	-0.104
	SOC05	0.893	-1.120	-1.921	-0.422	0.213
	SOC07	1.798	-2.321	-2.601	-0.616	0.050
STR	STR01	0.425	-0.415	-1.504	-0.400	0.117
	STR02	1.137	0.062	-2.221	-0.425	0.206
	STR03	0.452	0.281	-1.733	-0.029	-0.026
	STR04	1.131	-0.449	-2.190	-0.361	0.078
	STR05	1.573	-0.818	-2.466	-0.468	-0.043
	STR06	1.635	-0.887	-2.742	-0.616	0.111
	STR08	1.329	-0.539	-2.563	-0.345	0.005
TOL	TOL01	0.770	-1.224	-0.951	-0.443	-0.292
	TOL02	1.202	-1.359	-1.968	-0.493	0.020
	TOL04	0.941	-1.675	-0.878	-0.431	-0.048
	TOL05	1.810	-1.838	-2.443	-1.223	0.534
	TOL06	0.398	-0.547	-0.302	-0.140	-0.621
	TOL07	0.499	-0.388	-1.588	-0.725	0.312
	TOL08	2.726	-3.254	-3.511	-1.716	0.548
TRU	TRU01	0.626	-0.377	-1.746	-0.652	0.408
	TRU02	0.912	-0.714	-1.470	-0.769	0.020
	TRU04	1.357	-1.402	-2.523	-0.579	-0.065
	TRU05	0.567	-0.376	-1.503	-0.391	0.005
	TRU06	0.913	-0.925	-1.903	-0.972	-0.068
	TRU07	2.245	-1.841	-3.648	-1.133	0.230
	TRU08	1.929	-1.471	-3.271	-1.351	0.271

EFF	ASS01	0.629	-0.483	-1.630	-0.405	0.501
	CRE02	1.149	-1.554	-2.005	-0.960	0.380
	CUR05	1.269	-2.058	-1.102	-1.349	-0.190
	EMP06	0.746	-0.911	-2.020	-1.124	0.386
	OPT04	0.991	-1.538	-1.525	-0.738	0.083
	PER04	1.057	-1.199	-2.226	-0.718	0.499
	RES02	1.298	-2.111	-1.336	-1.095	0.114
	SEL02	0.823	-1.049	-1.593	-0.626	-0.089
MOT	CUR02	1.410	-1.648	-1.997	-1.079	0.527
	ENE02	0.895	-1.187	-1.707	-0.833	0.336
	PER02	1.783	-1.567	-2.960	-0.833	0.595
	PER08	2.083	-1.726	-3.597	-1.238	0.584
	RES06	1.787	-2.124	-2.258	-1.235	0.345
	SEL06	1.137	-0.900	-2.585	-0.711	0.474

Annex H. Contributors

Introduction

SSES is a collaborative effort, bringing together experts from participating Sites.

The overall management and implementation of the Field Test and Main Study was carried out by the Australian Council for Educational Research (ACER). The OECD Secretariat worked closely with the International Project Director, Dr Sue Thomson (ACER), to coordinate all aspects of SSES implementation.

The OECD Secretariat has overall managerial responsibility for the Study, and monitors its implementation regularly.

Site Project Managers (SPMs) *(*Former SPM)*

Canada (Ottawa) – Katherine Magner, Andrea Rahim*, Praveeni Perera*

China, People's Republic of China (Suzhou) – Zhengguo Yuan, zhongjing Huang

Colombia (Bogota) – Natalia Gonzalez Gomez

Colombia (Manizales) – Santiago Isaza, Camilo Gallego Gutiérrez

Finland (Helsinki) – Satu Gripenberg*, Mette Ranta

Republic of Korea (Daegu) – Hyunjin Jinny Kim, Meesook Kim*

Portugal (Sintra) – Maria de Fátima Galvão Grilo

Russian Federation (Moscow) – Alena Kulikova, Ekaterina Orel

Turkey (Istanbul) – Sadri Sensoy, Bilge Taskireç*

United States (Houston, Texas) – Holly Heard, Jodi Moon*

OECD Secretariat

Andreas Schleicher (Special Advisor on Education Policy to the Secretary General and Director for the Directorate for Education and Skills)

Dirk Van Damme (Senior Counsellor and Acting Head of the Innovation and Measuring Progress Division to the Directorate for Education and Skills)

Rowena Phair (Project Manager)

William Thorn (Project Manager)

Marta Encinas-Martin (Project Manager)

Eva Feron (Analyst)

Milo Kankaras (Analyst)

Javier Suarez-Alvarez (Analyst)

Rose Bolognini (Project Coordinator)

Fiona Parsons (Project Coordinator)

Valentine Bekka (Project Assistant)

Mia Tuzovic (Project Assistant)

Technical Advisory Group

Kristin Anderson Moore (Child Trends Research Organisation, United States)

Filip De Fruyt (Ghent University, Belgium)

Patrick Kyllonen (Educational Testing Service, United States)

Bruno Losito (Roma Tre University, Italy)

Ricardo Primi (Universidade São Francisco, Brazil)

Beatrice Rammstedt (GESIS – Leibniz-Institute for the Social Sciences, Germany)

Brent Roberts (University of Illinois, United States)

Christopher Soto (Colby College, United States)

Fons van de Vijver (University of Tilburg, the Netherlands)

SSES lead contractors

Ohio State University (OSU), United States - 2017-18 (instrument development)

Australian Council for Education and Research (ACER) – 2018-2021

The overall management and implementation of the Field Test and Main Study was carried out by the Australian Council for Educational Research (ACER). The OECD Secretariat worked closely with the International Project Director, Dr Sue Thomson (ACER), to co-ordinate all aspects of implementation.

Sue Thomson (International Project Director)

Wei Buttress (Project Manager)

Laila Helou (Online Platform Management and Project Support)

Fei Cain (Quality Monitoring Management and Project Support)

Wolfram Schulz (Psychometrics Consultant)

Yan Bibby (Data Analysis Leader)

Greg Macaskill (Data Analyst)

Renee Kwong (Data Analyst)

Martin Murphy (Sampling Leader)

Claire Ozolins (Sampling Leader)

Jorge Fallas (Sampling)

Nina Martinus (Sampling)

Emma Camus (Sampling)

Alla Routitisky (Data Management)

Alexander Daraganov (Data Management and Student Sampling)

Nora Kovarcikova (Field Operations)

Tim Friedman (Contextual Questionnaire Data Analysis Leader)

Jennifer Hong (Contextual Questionnaire Data Analyst)

Jessica Thompson (Contextual Questionnaire Data Analyst)

Organizations working with the International Contractor ACER

RM Assessment Master (previously SoNET System)

Rakshit Shingala (Development Team Lead)

Stephen Ainley (Quality Assurance & Support Lead)

Arman Pagilagan (Developer)

Cheng Chang (Developer)

Nilupuli Lunuwila (Developer)

Stephen Birchall (Project Director)

Grigory Loskutov (Development Operations Manager)

cApStAn LQC

Shinoh Lee (Operations Lead Project Manager for phase 1)

Emel Ince (Lead Project Manager for phase 2)

Manuel Souto Pico (Lead Translation Technologist)

Adrien Mathot (Translation Technologist)

Steve Dept (Executive Sponsor)

HallStat SRL

Béatrice Halleux (Translation Referee)