TALIS Starting Strong 2018 Technical Report

TEACHING AND LEARNING INTERNATIONAL SURVEY

OECD

TALIS

TALIS Starting Strong 2018 Technical Report

Photo credits: Cover © Monkey Business Images/Shutterstock.com

Acknowledgements

The OECD Starting Strong Teaching and Learning International Survey 2018 is the result of collaboration among the participating countries, the OECD Secretariat, and an international consortium led by the International Association for the Evaluation of Educational Achievement (IEA).

The OECD recognises the significant contributions of members of the IEA Hamburg, Germany, IEA Amsterdam, the Netherlands, as well as Statistics Canada, who authored this report. The principal authors of the respective chapters were as follows:

Chapter 1: Juliane Hencke
Chapter 2: Julie Bélanger, Megan Sim, Miriam Broeks and Juliane Hencke
Chapter 3: Julie Bélanger, Megan Sim, Miriam Broeks and Juliane Hencke
Chapter 4: Jan-Philipp Wagner, David Ebbs, Sandra Dohr, Alena Becker, Juliane Hencke and Juliane Kobelt
Chapter 5: Karsten Penon and Jean Dumais
Chapter 6: Juliane Kobelt, Juliane Hencke, Alena Becker and Christine Busch
Chapter 7: Jan-Philipp Wagner, David Ebbs, Sandra Dohr, Alena Becker and Juliane Hencke
Chapter 8: Christine Busch and Alena Becker
Chapter 9: Karsten Penon and Jean Dumais
Chapter 10: Jean Dumais, Karsten Penon, Alena Becker, Jan-Philipp Wagner and David Ebbs
Chapter 11: Agnes Stancel-Piątak, Justin Wild, Minge Chen, Mojca Rozman, Plamen Mirazchiyski and Hynek Cigler
Chapter 12: Ann-Kristin Koop, Falk Brese and Agnes Stancel-Piątak

Paula Wagemaker copy-edited the report on behalf of the IEA. Elizabeth Shuey, Arno Engel and Stéphanie Jamet of the OECD Directorate for Education and Skills (EDU) provided additional editorial support. Fons van de Vijver, Timothy L. Kennel and Christian Monseur provided valuable comments during the development of some chapters. Mernie Graziotin (EDU) provided additional proofreading and managed the layout process, with support from Henri Pearson (EDU Communications team).

Table of Contents

Tables

Figures

Boxes

Abbreviations and acronyms

Abbreviation/Acronym	Description	Comment
ACER	Australian Council for Educational Research	
AV	Adaptation verification	
BRR	Balanced repeated replication	
CC	Centre co-ordinator	
CFA	Confirmatory factor analysis	
CFI	Comparative fit index	
CNO	Co-ordinated national options	
CQ	Combined centre questionnaire	Questionnaire with combined staff and leader questions; applicable if a centre consisted of only one person
CQ+	Combined centre questionnaire plus	Questionnaire with combined staff and leader questions as well as additional questions prepared for administration of the survey in Israel's ISCED level 02 centres
DME	IEA data management expert	IEA software
DPE	IEA data processing expert	IEA software
EAS	IEA eAssessment	IEA software
ECEC	Early childhood education and care	
EFA	Exploratory factor analysis	
EM	Expectation-maximisation	
EU	European Union	
FSD	Factor score determinacy	
FT	Field trial	
IEA	International Association for the Evaluation of Educational Achievement	
ID	Identification number	
IDB	International database	
IQO	International quality observer	Contracted by IEA
IRT	Item response theory	
ISC	International study centre	For TALIS Starting Strong: IEA Hamburg
ISCED	International Standard Classification of Education	Published in 2011
ISCED level 0	Centres in ISCED level 0 were defined as institutional (officially registered) settings providing ECEC programmes, i.e. formal education and care, for young children from birth up to entry into primary education	
ISCED level 02	Centres providing education designed to support early development in preparation for participation in school and society and accommodating children from the age of three to the start of primary education	Often referred to as "pre-primary education"
LQ	Leader questionnaire	
MGCFA	Multiple-group confirmatory factor analysis	
MLR	Maximum likelihood robust	
MOS	Measure of size	
MS	Main survey	
NA	Not applicable	
NADB	National adaptations database	
NAF	National adaptation form	
NDM	National data manager	

Abbreviation/Acronym	Description	Comment
NPM	National project manager	
NQO	National quality observer	Contracted by national centre
NRBA	Non-response bias analysis	
NSM	National sampling manager	
OECD	Organisation for Economic Co-operation and Development	
OSS	IEA Online Survey System	IEA software
PISA	OECD Programme for International Student Assessment	
PPS	Probability proportional to size	
PRE	IEA Participation Rate Estimator	IEA software
PSU	Primary sampling unit	
QEG	Questionnaire Expert Group	
RAND Europe	RAND Europe is a not for profit research institute that helps to improve policy and decision making through research and analysis	An international consortium partner during the TALIS Starting Strong 2018
RandA	Research and Analysis Team	Unit within IEA Hamburg
RMSEA	Root mean square error of approximation	
SAQ	Survey activities questionnaire	
SEM	Structural equation modelling	
SOP	Survey operations procedures	
SPSS	Statistical Package for the Social Sciences	Software
SQ	Staff questionnaire	
SQL	Structured query language	
SRMR	Standardised root mean square residual	
SRS	Systematic equal probability random sampling	
Statistics Canada	Canadian national statistical office	Provided the sampling referee during TALIS Starting Strong 2018
TAG	Technical advisory group	
TALIS	Teaching and Learning International Survey	
TGB	TALIS governing board	
TEDS-M	Teacher Education and Development Study in Mathematics	
TLI	Tucker-Lewis Index	
TUAC	Trade Union Advisory Committee	
U3	ECEC centres accommodating children under three years of age	
UNESCO-UIS	UNESCO Institute for Statistics	
WinW3S	IEA Within-School Sampling Software	IEA software
WLSMV	Weighted least squares mean variance	
WRMR	Weighted root mean square residual	

Three-digit operational codes of TALIS Starting Strong 2018 participants

TALIS Starting Strong 2018 participating country	Code
Chile	CHL
Denmark	DNK
Germany	DEU
Iceland	ISL
Israel	ISR
Japan	JPN
Korea	KOR
Norway	NOR
Turkey	TUR

Executive summary

This technical report details the steps, procedures, methodologies, standards and rules that the OECD Starting Strong Teaching and Learning International Survey 2018 (TALIS Starting Strong 2018) used to collect high-quality data. The primary purpose of the report is to support readers of the international and subsequent thematic reports as well as users of the public international database when interpreting results, contextualising information and using the data. A user guide complements this technical report and provides detailed guidance and examples for individuals using the TALIS Starting Strong 2018 data.

The report is structured in the form of 12 chapters followed by 10 annexes.

Chapters 1 to 12 provide an overview of the conceptual framework and questionnaire development processes, translation and translation verification, further survey operations and quality observations of the field work in the participating countries. It describes the sampling, weighting and adjudication procedures applied, as well as the preparation and structure of the international databases. Also included is an account of the scaling methodology and the results of that scaling along with information on the production and verification of the international results tables.

Annexes A to J focus on the individuals and groups who prepared and conducted the survey, the technical standards, forms and questionnaires used, the achieved participation rates of the two populations of interest, and the scaling process and its results.

The following summary highlights the key operational parameters and approaches applied in TALIS Starting Strong 2018 and further detailed in this technical report.

Governance, management and structure

TALIS Starting Strong 2018, launched and managed by the Organisation for Economic Co-operation and Development (OECD) as part of its early childhood and care (ECEC) data development strategy (OECD, 2013[1]), built on the experience of the OECD's Teaching and Learning International Survey (TALIS) in primary and secondary education.

The overall objective of TALIS Starting Strong 2018 was to provide robust international indicators and policy-relevant analysis of data pertaining to ECEC centre staff and centre leaders, their pedagogical and professional practices, and the learning and well-being environments in the centres. The intention behind this objective was to provide information that will help countries review and develop policies that promote conditions conducive to positive child development, well-being and learning.

The TALIS Starting Strong 2018 International Consortium, led by the International Association for the Evaluation of Educational Achievement (IEA), with its premises in Amsterdam, the Netherlands, and Hamburg, Germany, was responsible for implementing the survey at the international level on behalf of the OECD Secretariat. The IEA's partners were RAND Europe, based in Cambridge, the United Kingdom, and Statistics Canada, based in Ottawa, Canada.

The TALIS Starting Strong international study centre (ISC) was located at the IEA Hamburg, Germany, and was led by Juliane Hencke (IEA).

The Questionnaire Expert Group (QEG), chaired by Dr Julie Bélanger (RAND Europe), developed the ECEC centre leader and staff questionnaires and contributed (along with others) to the TALIS Starting Strong 2018 data analysis plan.

The Technical Advisory Group (TAG), chaired by Dr Fons van de Vijver, Tilburg University, the Netherlands, provided on-demand and further advice to the consortium and the OECD in relation to technical, methodological and analytical matters.

Participating countries, populations of interest and survey phases

Nine countries took part in TALIS Starting Strong. Each conducted the survey in ECEC centres belonging to level 02 of the education system as defined in the International Standard Classification of Education (ISCED) (UNESCO-UIS, 2012[2]). Level 02 centres typically cater for children from three years of age through to the time they enter primary education. In addition, four countries administered the survey in ECEC centres for children under three years of age (U3). Each participating country appointed a national project manager (NPM) who had primary responsibility for all national survey implementation work.

The survey was organised and conducted in three phases: (1) the pilot, the purpose of which was to develop and trial the content of the survey questions; (2) the field trial, designed to test and evaluate the questionnaires and item formats as well as the survey procedures and data collection modes; and (3) the main survey, which focused on collecting the full-scope TALIS Starting Strong 2018 data in the participating countries.

Framework and questionnaire development and translation

The TALIS Starting Strong 2018 conceptual framework guided the theory-driven development of the survey's instruments, thus providing an integrated theoretical and analytical underpinning to the survey.

The OECD Secretariat, together with the OECD Early Childhood and Education and Care Network (ECEC Network), defined three main policy issues as the central focus of this first cycle of TALIS Starting Strong. These were ensuring the quality of learning and well-being environments; motivating, attracting and retaining staff to the profession; and developing staff for and within the profession.

The OECD Secretariat and the ECEC Network conducted a priority rating exercise with the nine participating countries to determine the structure of the questionnaires and the themes and indicators that should be included in them. Following this exercise and further discussions with stakeholders, the OECD Secretariat and the consortium decided that TALIS Starting Strong 2018 should include at least six themes that would inform all three policy issues.

Building on the OECD Secretariat's preparatory conceptual work, the survey's Questionnaire Expert Group (QEG) developed the conceptual framework, which was based on sound theoretical principles and an in-depth understanding of relevant areas. Members of the group also considered the policy imperatives of the countries participating in the survey and the methodological demands of international large-scale surveys. Revisions to

the conceptual framework took into account the results of the pilot and field trial phases, as well as feedback received from various stakeholders.

The TALIS Starting Strong 2018 instrument development was theory driven, thus ensuring alignment with the conceptual framework and bringing a high level of quality to the technical and psychometric properties of the survey questions. Development took into consideration underpinning policy questions, key themes and country input and priorities. The QEG focused on ensuring the cross-cultural validity of survey questions and alignment between TALIS Starting Strong 2018 and TALIS 2018.

The overall task was to design and develop instruments for the two populations of interest, namely ECEC centre leaders and staff at ISCED level 02, and ECEC centre leaders and staff catering for children under the age of three (U3). The development work, which took place between March 2015 and November 2017, encompassed three stages: the pilot, which used a focus group approach; the quantitative field trial; and extensive psychometric analyses prior to the main survey. Throughout the development process, the QEG considered and incorporated into the instruments feedback that key stakeholders provided over various feedback rounds.

During the field trial, the draft staff questionnaire was administered according to a rotated questionnaire design so that as many questions as possible could be tested prior to the eventual deletion of some of them. At this stage, the instruments also included built-in experiments designed to test different question formats and wording and thereby determine which ones had the better psychometric properties.

The resulting main survey questionnaires were used for leaders and staff at ISCED level 02 centres and, with some adaptations, for leaders and staff in centres for children under the age of three. Each questionnaire (English language version) took respondents about 45 to 60 minutes to complete.

All participating countries were responsible for producing their own national translations of the survey instruments, which included a leader questionnaire, a staff questionnaire, and cover letters for online administration of the instruments for each population of interest. National study centres used the international English source version of the survey instruments as a basis for their national questionnaires and documented, on national adaptation forms (NAFs), all adaptations they made to the instruments in order to fit their content to their respective national contexts.

To ensure a high level of quality and the international comparability of the data, the TALIS Starting Strong 2018 national survey instruments underwent three verification steps: national adaptation verification, translation verification, and layout verification. In total, the nine participating countries translated the survey instruments into 11 languages. Two participating countries administered the survey in more than one language.

The survey activities questionnaire (SAQ) showed that most of the NPMs found the information provided about the instrument preparation and verification processes was clear and sufficient. They also considered the feedback from the verification processes as useful for preparing the national survey instruments.

Sampling of ECEC centres and staff

The template for a sampling design for TALIS Starting Strong 2018 was adapted from that of TALIS 2018. Two populations of interest were in scope for the survey: the ECEC centres operating at ISCED level 02, and the centres catering for children under the age of three

years (U3). A first-stage random sample of 180 centres was followed by a second-stage random sample of 8 staff (when possible) from each of the selected centres. Home-based settings were part of the international target population for ISCED level 02 and for services for children under three years of age. However, all participating countries decided to exclude home-based settings from the ISCED level 02 coverage, and some participating countries excluded home-based settings from the survey of services for children under the age of three years.

As with TALIS 2018, the expected adaptations to national conditions (e.g. number of centres or staff in the population of interest, excluded areas, analytical or reporting requirements) were implemented.

The samples of centres for each participating country were drawn internationally by the TALIS Starting Strong sampling team, while the samples of staff were drawn by the national study centres using software provided by the survey's consortium.

A designated centre co-ordinator in each participating centre was responsible for listing, on survey listing forms, each eligible staff member along with specified information about him or her. The national study centres then used the software provided by the consortium and the information provided on the listing forms to assign the leader questionnaire to the leader and the staff questionnaires to a sample of all or a minimum of eight staff members in each participating centre.

Main survey administration and quality observations

In order to provide participating countries with the ability to adjust their main survey data collection to country-specific needs, this process took place between March and July 2018 in the Northern Hemisphere countries and between July and November 2018 in the Southern Hemisphere countries.[1] Two countries extended their data collection period to improve their survey participation rates.

Five countries administered the survey online. Two countries administered TALIS Starting Strong 2018 in a mixed mode, which meant they used both online and paper instruments. Two countries opted for the paper mode. During the main survey, 48.7% of the respondents completed the survey on line and 51.3% completed it on paper.

The consortium requested three different quality observation activities: an international quality observation (IQO) programme overseen by the IEA Amsterdam (applicable during the main survey only); a national quality control programme implemented by the NPMs (conducted during the field trial and main survey); and an online survey activities questionnaire, used to gather national study centres' feedback on the different survey activities (completed after the field trial and main survey implementation).

For the international quality control programme, the IEA Amsterdam recruited independent observers for each participating country to monitor administration of the survey in a subsample of their participating centres.

Data collection, processing, weighting and adjudication

The IEA Hamburg applied a vast number of checks and cleaning routines to the data and continuously sent data to the consortium partners, the analysis teams and the OECD. They also provided countries with regular data updates for verification.

As with TALIS 2018, the IEA Hamburg adjudicated the centre or centre leader data independently from the staff data and introduced the notion of a "participating centre for centre/centre leader data". Here, a centre was deemed a "participating" centre if its leader returned his or her questionnaire with at least one valid response. The criterion determining whether a centre was participating with respect to the staff data was at least 50% of the selected staff in the centre returning their questionnaires.

The average participation rate for ISCED level 02 centre/centre leader data before replacement by another sampled centre was 75.0% (83.9% after replacement). The rates ranged from 40.9% to 99.8% before replacement. The average overall ISCED level 02 staff participation rate was 76.8%, with the range extending from 38.5% to 99.2%.

In comparison, the average for U3 centre/centre leader data before replacement was 61.6% (73.7% after replacement), and the rates ranged from 35.4% to 93.3% before replacement. The average overall U3 staff participation rate was 65.1%, with rates ranging from 37.6% to 92.8%.

Adjudication was performed to determine the overall quality of the data, that is, whether the data released to the participating countries and the OECD were "fit for use" as intended under the lead of the sampling referee for TALIS Starting Strong 2018. The design and application of the quality assurance processes drew on expert advice and opinion, on qualitative information and learned judgement, and on quantitative information.

The consortium examined each individual dataset from each participating country and then discussed and clarified any issues arising out of this process with the survey's various stakeholders as necessary.

Two versions of the international database were created: a public-use file (PUF), available on the OECD's website for secondary data users and researchers; and a restricted-use file (RUF), available only to accredited members. These members were people who were nominated by the respective national Extended ECEC Network members and who also accepted the respective confidentiality and embargo rules.

Scaling, analysis and results table production

The indices created during TALIS Starting Strong 2018 encompassed simple indices (e.g. ratios) and complex indices (i.e. scale scores).

The scaling procedure for the complex indices was conducted by the IEA Hamburg within the framework of multiple-group confirmatory factor analysis (MGCFA) and consisted of two major steps – scale evaluation and scale score computation.

Measurement invariance techniques were used to test cross-country comparability. Across the two instruments (the staff questionnaire and the leader questionnaire) and both populations of interest, the majority of scales reached the metric level of invariance allowing comparisons of correlational analyses across the participating countries. Four scales reached the scalar level of measurement invariance allowing comparisons of scale score means across the countries.

Because the main goal of TALIS Starting Strong 2018 was to generate reliable, valid and comparable population estimates based on sample data, all parameters presented in the TALIS Starting Strong 2018 international reports were weighted. Standard errors were estimated by applying Fay's variant of the balanced repeated replication (BRR) technique.

As part of the presentation of results in the international reports, the IEA Hamburg produced a large number of tables. Preparation of the tables consisted of two major steps. The first was the development, review and revision of the table shells provided by the OECD, and the second was data analysis, table production and verification. The tables contain population estimates together with their standard errors.

Notes

1. The terms Northern Hemisphere and Southern Hemisphere do not necessarily refer to the geographic location of a participating country but rather to the data collection period the country followed during the survey. The determination of the data collection window depended primarily on how each country organised its ECEC year, for example, longer closing times during summer.

References

OECD (2013), *From policy questions to new ECEC indicators: Draft background paper for the 14th meeting of the OECD ECEC Network*, OECD, Paris. [1]

UNESCO-UIS (2012), *International Standard Classification of Education (ISCED) 2011*, UNESCO Institute for Statistics, Montreal, http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf. [2]

Chapter 1. Introduction

This chapter provides an overview of the survey design, preparation and implementation of the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018, initiated and co-ordinated by the Organisation for Economic Co-operation and Development (OECD). It introduces the management structure put in place to administer this first international large-scale survey in the early childhood education and care (ECEC) sector, the populations of interest and educational levels investigated, the role of the national project managers, the key development phases of the survey, and the standardised procedures that were implemented to allow the collection of high quality data in nine countries.

1.1. TALIS Starting Strong 2018 in brief

This OECD Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 Technical Report documents the development of the survey's conceptual framework and instruments. It also describes the methods used to prepare the national survey instruments, conduct sampling, data collection, weighting, scaling and analysis, and produce tables of results. The report enables readers to review and understand these procedures and to gain insight into the rigorous quality control programme that encompassed all phases of the survey.

As part of its early childhood and care (ECEC) data development strategy (OECD, 2013[1]), and building on the experience of the Teaching and Learning International Survey (TALIS) 2018 in primary and secondary education, the OECD launched TALIS Starting Strong in 2016. The aim of this international survey of centre staff and leaders in the ECEC sector was to collect data on learning and well-being environments, with a particular emphasis on the work that staff and centre leaders carry out with children in ECEC settings. The survey also aimed to collect data on what motivates staff to join the ECEC profession, on factors affecting their career decisions, and on how staff are prepared for and trained within the profession. TALIS Starting Strong 2018 is the first survey of its kind to provide rich comparable data relevant for the delivery of quality ECEC services internationally.

This first cycle of the survey, carried out in 2018, offered participating countries the possibility of conducting the survey in two populations of interest in the ECEC sector: ECEC centres belonging to International Standard Classification of Education (ISCED) (UNESCO-UIS, 2012[2]) level 02 (usually centres catering for children from three years of age through to the time they enter primary education); and ECEC centres for children under three years of age (U3). Centres were defined as institutional (officially registered) settings providing ECEC programmes, that is, formal education and care for young children from birth up to entry into primary education (ISCED level 0). In order to be classified as a "centre", settings had to be ones providing educational activities for at least 2 hours per day, 100 days a year.

Nine countries participated in the survey at ISCED level 02, and four of these countries also administered the survey at the U3 level (Table 1.1).

Table 1.1. Countries participating in TALIS Starting Strong 2018 and their surveyed populations of interest

Participating country	U3	ISCED level 02
Chile		●
Denmark	●	●
Germany	●	●
Iceland		●
Israel	●	●
Japan		●
Korea, Republic of		●
Norway	●	●
Turkey		●

TALIS Starting Strong 2018 addressed the following 12 themes which together cover the wide range of dimensions in ECEC settings that are expected to contribute to ensuring the quality of learning and well-being environments, whether directly or indirectly (Melhuish et al., 2015[3]; National Association for the Education of Young Children, 1991[4]; Tietze et al., 2013[5]; Tietze, Roßbach and Grenner, 2005[6]; World Health Organization, 1990[7]). The themes are:

- process quality of staff–child interaction
- monitoring children's development, well-being and learning
- structural quality characteristics of ECEC centres
- pedagogical and administrative leadership
- ECEC centre climate
- ECEC centre stakeholder relations
- background and initial preparation of ECEC centre leaders and staff
- professional development of ECEC centre leaders and staff
- well-being of ECEC centre leaders, staff and children
- professional beliefs about children's development, well-being and learning
- self-efficacy of ECEC centre staff and leaders
- equity and diversity in the child group.

Readers of this document are invited to gain more insight into the policy issues, policy themes and policy indicators included in the survey from the *Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework* (Sim et al., 2019[8]), which includes detailed illustrations of the survey's conceptual model and priority areas and the mapping between these. The framework document also describes the policy aspects that initially drove TALIS Starting Strong 2018 and explores the limitations of this policy-related work. However, details on the process and methods applied to develop the conceptual framework are also documented in Chapter 2 of this technical report.

The policy themes were operationalised in two questionnaires: a leader questionnaire, completed by ECEC centre leaders, and a staff questionnaire, completed by staff members sampled from among the staff in each centre. Participating countries had the option of administering their survey instruments on paper and/or on line. Five of the nine participating countries decided to collect their data predominantly on line, two countries applied a mixed mode (on line and on paper) and two countries administered their questionnaires on paper only.

The remainder of this chapter briefly describes the management structure of TALIS Starting Strong 2018 at international and national levels and outlines the survey's three phases and its key milestones.

1.2. Link to TALIS 2018

Development of TALIS Starting Strong 2018 followed the design used for TALIS 2018. TALIS Starting Strong 2018 used the same guiding principles as TALIS 2018 and sought to maximise alignments and synergies with that survey to the greatest extent possible.

These alignments and synergies were achieved at the level of governance and implementation. Thus, for example, governing board meetings for both studies took place back-to-back at the same location, and some countries appointed one national study centre for both surveys. Overlap was also evident in the membership of the Questionnaire Expert Groups (QEGs) and the Technical Advisory Groups (TAGs) for both studies.

Commonalities furthermore existed at the conceptual and content levels, specifically with respect to the policy areas, themes and indicators that both surveys addressed. This overlap provided countries participating in the two surveys with the added analytical value of comparing indicators across the populations of interest.

At the beginning of TALIS Starting Strong 2018, the OECD Secretariat (see the next section) aimed to achieve an overlap of approximately 70% between the indicators used in TALIS 2018 and the indicators included in TALIS Starting Strong 2018, leaving about 30% of the latter set of questionnaires for items containing ECEC-specific indicators. However, this overlap at item-level had to be reduced during the questionnaire development process because a larger number of adaptations to the TALIS 2018 questions than expected was required to ensure relevance to the ECEC sector. For the main survey, the item-level overlap between TALIS Starting Strong 2018 staff questionnaire and the TALIS 2018 teacher questionnaire was approximately 29%, while the overlap between TALIS Starting Strong 2018 leader questionnaire and the TALIS 2018 principal questionnaire was approximately 48%. For more details on the overlap between TALIS 2018 and TALIS Starting Strong 2018, please refer to the *Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework* (Sim et al., 2019[8]).

1.3. Managing the survey internationally

In April 2016, the OECD awarded the International Association for the Evaluation of Educational Achievement (IEA) and its consortium partners, RAND Europe and Statistics Canada, the role of conducting the international implementation of TALIS Starting Strong 2018. Under this partnership, the OECD commissioned the IEA Hamburg office (located in Germany) to form the international study centre (ISC) tasked with co-ordinating the international implementation of TALIS Starting Strong 2018. The TALIS Starting Strong 2018 International Consortium included staff from not only the IEA office in Hamburg but also the IEA office in Amsterdam as well as personnel from RAND Europe and Statistics Canada.

The OECD Secretariat's Directorate for Education and Skills in Paris, France, was responsible for the overall supervision of the project. The Extended OECD ECEC Network on TALIS Starting Strong 2018 was led by Arno Engel, with guidance from Yuri Belfali, Miho Taguma and Stéphanie Jamet, supported by Elizabeth Shuey, Clara Barata, Victoria Liberatore, Théo Reybard, Éric Charbonnier, Maria Huerta, Katja Anger, Mernie Graziotin and Leslie Greenhow.

The team at the ISC was led by Juliane Hencke (international study director) and Steffen Knoll (deputy international study director), assisted primarily by Juliane Kobelt (international study co-ordinator) and supported by Viktoria Böhm (international study co-ordinator).

The IEA's data processing team was led by Alena Becker (international data manager) and Christine Busch (deputy international data manager). Both were responsible for supporting the countries during the survey implementation by providing survey-specific software

applications, related manuals, and training. They also had overall responsibility for the international data processing and preparation of the international database.

The IEA's sampling team was responsible for determining the survey's international sample design and national sampling plans, sample implementation and weighting, for preparing the datasets for variance estimation, and for adjudication. The team included Sabine Meinck (international sampling manager) and Karsten Penon (international sampling manager).

The analysis team at the IEA Hamburg was led by Agnes Stancel-Piątak. Minge Chen and Justin Wild were responsible for the psychometric foundation of the survey, a task involving the development and implementation of the validation and scaling procedures and the production of the scale scores. Ann-Kristin Koop was responsible for reviewing the table shells prepared by the OECD and for conducting data analyses for the survey's international report.

At the IEA Amsterdam, the IEA financial director Roel Burgers managed the financial and contractual affairs in co-operation with Christian Groth, head of accounting and controlling at the IEA Hamburg. The project team, consisting of David Ebbs, Michelle Djekic, Sandra Dohr and Jan-Philipp Wagner, was responsible for verifying the translation of the national survey questionnaires into 11 language versions and for managing the international quality observation programme in all participating countries by contracting, training and supervising independent quality observers. The IEA Amsterdam contracted cApStAn Linguistic Quality Control, specialising in semantic quality control of translations in a range of international surveys and based in Brussels, Belgium, to support the translation verification work.

Julie Bélanger at RAND Europe, United Kingdom, chaired the QEG and had overall responsibility for the group. She co-ordinated a team of internationally selected researchers responsible for developing the TALIS Starting Strong 2018 conceptual framework and questionnaires. Chapter 2 of this technical report provides a detailed list of QEG members. At RAND Europe, Megan Sim, Miriam Broeks and Katherine Stewart helped Julie Bélanger develop the conceptual framework and the questionnaires.

The survey's international study centre appointed Jean Dumais, the former chief statistician at Statistics Canada, as the survey's sampling referee, while the OECD Secretariat appointed Fons van de Vijver (Tilburg University) as chair of the joint TALIS 2018 and TALIS Starting Strong 2018 TAG.[1] van de Vijver's task was to provide expert advice and guidance on survey methods, processes and analyses.

1.4. Working with national project managers

In 2016, each participating country began the work of establishing a national study centre led by an experienced national project manager (NPM). The TALIS Starting Strong 2018 International Consortium provided guidance on the NPMs' roles and responsibilities. Each NPM had primary responsibility for preparing and co-ordinating the survey at the national level. The experience and expertise of the NPMs and their national study centre staff (e.g. involvement in other international large-scale assessments) strongly influenced how the range of national tasks was apportioned and managed. The international study centre also recommended that each national study centre should appoint a national data manager (NDM) to oversee and implement all data and technical tasks, and a national sampling manager (NSM) to help the NPM manage complex sample designs and potential national additions and extensions to the survey samples.

The number of staff members in the national study centres varied considerably from one country to the next depending on the size of the country and how it chose to organise the national data collection work. Some NPMs worked with external survey organisations to conduct scientific and/or operational tasks. These organisations were all appointed on a tendering and contracting basis.

The tasks required of the national project managers, data managers and/or sampling managers included the following:

- establishing an overall preparation and administration schedule in co-operation with the ISC
- attending NPM meetings in order to become familiar with all instruments, materials and survey procedures
- providing the consortium with an up-to-date national sampling frame of ISCED level 02 centres and, where applicable, U3 centres
- discussing national design options, such as oversampling, with the IEA's sampling expert
- performing within-centre listing, sampling and tracking
- carrying out quality control and plausibility checks of staff lists and samples to identify, for example, lists abbreviated to include just those staff who had agreed to participate or lists that were otherwise incomplete/inaccurate
- appointing experienced translator(s) to produce the national versions of the international instruments
- documenting required national adaptations of the instruments on national adaptation forms (NAFs)
- preparing for paper and/or online data collection
- identifying and training centre co-ordinators
- appointing and training national quality observers (NQOs)
- nominating possible international quality observers (IQOs) and supporting their work
- monitoring the return status of all questionnaires and projected response rates
- manually entering responses on the paper-administered questionnaires into the international database and performing consistency checks
- submitting data and documentation to the consortium and responding to data queries during data processing and analysis
- completing a survey activities questionnaire after survey administration.

Regular communication between the NPM and the ISC ensured that survey administration proceeded according to the TALIS Starting Strong 2018 technical standards (see Annex B). As an adjustment to how countries co-ordinate their ECEC calendar year, the ISC offered the participating countries two data collection periods, one for the Northern Hemisphere countries and one for the Southern Hemisphere countries. All countries met international milestones and submitted their data and documentation as required or according to deviations from the international schedule negotiated with the consortium (e.g. because of

how the country organised its ECEC calendar year or because of the need to extend data collection periods in order to improve survey response rates).

National study centres distributed paper questionnaires or login details (for the online versions) to ECEC centre leaders and staff and asked them to complete the questionnaires within a nationally defined survey window.

Because the survey was implemented on a voluntary basis in the majority of countries (i.e. sampled centres and/or staff could choose not to answer the questionnaires), implementation proved to be a challenging exercise in several cases at centre and/or staff level. Based on experience in previous TALIS cycles, the ISC provided strategies and best practice examples to help national study centres achieve the required participation rates. National study centres worked very closely with relevant unions, local, regional and state authorities and/or the responsible national ministry/ies (as applicable) to endorse the survey and enable sufficient participation. Countries also engaged in extensive public relations activities to raise awareness among ECEC centre leaders and staff before the main data collection, and many created their own survey-related websites.

In general, an international large-scale survey like TALIS Starting Strong 2018 requires high communication standards to enable smooth and transparent communication throughout its lifecycle. To achieve this aim, the TALIS Starting Strong 2018 International Consortium established different communication channels. Most communication relied on email. However, the international consortium conducted virtual meetings and five face-to-face meetings with NPMs if the survey stage and the questionnaire development process deemed this type of contact appropriate. These meetings helped the NPMs prepare for the key survey phases (see Chapter 3.6 of this report) and enabled them to exchange experiences and approaches. The international consortium used Microsoft SharePoint as the collaboration system for document and data exchange between all involved parties. To ensure confidentiality and security, the ISC operated all systems and permitted only authorised personnel to access information. The working language throughout the project was English.

1.5. Standardised procedures

The TALIS Starting Strong 2018 technical standards (see Annex B), prepared by the consortium and approved by the Extended ECEC Network, provided high-level guidance to participating countries during all stages of survey preparation, implementation and data work. The standards denote practices that experts consider to be best practice in survey research and were therefore the ones that everyone associated with TALIS Starting Strong 2018 were expected to adhere to [see, for example, Biemer and Lyberg (2003[9]); Martin, Rust and Adams (1999[10]); OECD (2014[11]); Statistics Canada (2009[12])]. Adherence to these standards was key to ensuring the validity, reliability and comparability of the survey's questionnaires and data. The consortium also developed a set of operational manuals and guidelines describing the steps that all countries needed to take to ensure successful implementation of the survey.

Because the consortium provided the international source version of the TALIS Starting Strong 2018 questionnaires in English, the participating countries needed to translate them into their local language(s). Additionally, because ECEC systems vary across these countries, TALIS Starting Strong 2018 required them to adapt questions to match their cultural contexts, where necessary. At the same time, assurances had to be built into this work so that the ECEC centre leaders and staff members across the participating countries

received questions equivalent in meaning and scope and collected data that could be compared internationally.

To achieve this aim, the consortium established standards designed to ensure that national translations were consistent with the international version of the TALIS Starting Strong 2018 instruments. The TALIS Starting Strong 2018 questionnaires were consequently subjected to a stringent translation, adaptation and verification procedure prior to data collection. The IEA supported these procedures by providing countries with its eAssessment System for adaptation and translation (see Chapter 4 of this report).

The IEA's sampling team performed all centre sampling and weighting procedures in line with established standards and guidelines. More details on the sampling procedures and on sampling weights and participation rates can be found in Chapter 5 and Chapter 9 of this report respectively.

To help participating countries implement their field operations, the consortium provided various manuals, guidelines, software and procedural training. For example, the ISC provided NPMs with the IEA Within-School Sampling Software (WinW3S)[2] to aid their adherence to standardised procedures for listing, sampling and tracking survey respondents. For more details on this matter, see Chapter 6 of this report.

The IEA Amsterdam co-ordinated the quality observations of the data collection at the international level, while the NPMs took on this responsibility at the national level. International quality observers received intensive, two-day training provided by the IEA (see Chapter 7).

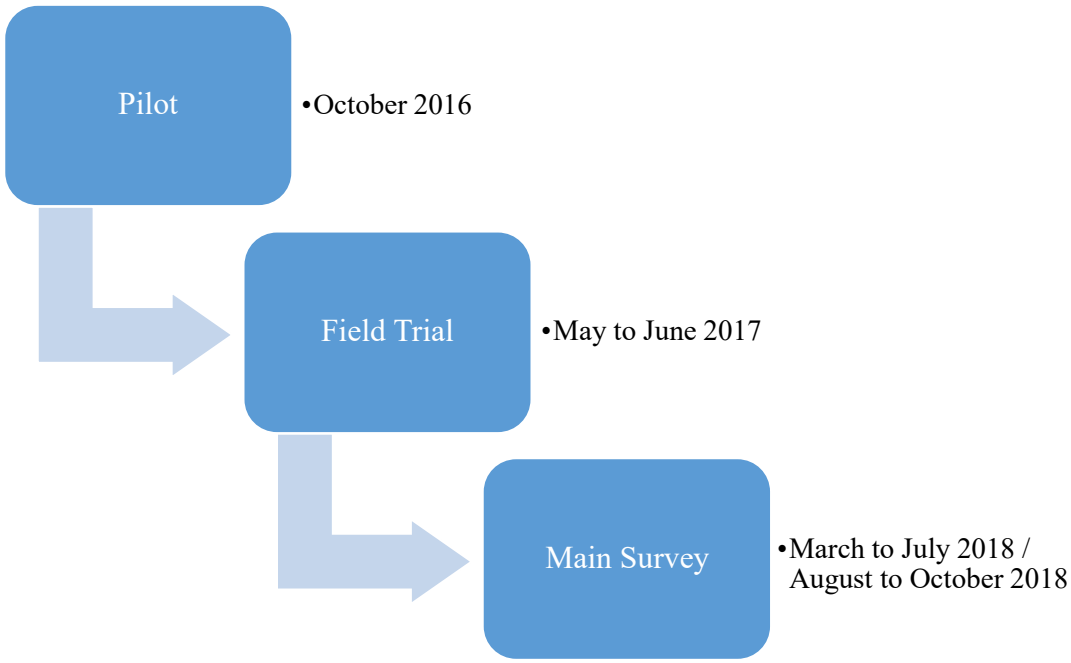
1.6. Key survey phases

The TALIS Starting Strong 2018 design included three key survey phases: a pilot study, a field trial and the main survey (see Figure 1.1). While the pilot study followed a qualitative approach, the field trial and the main survey followed a quantitative approach.

During the *pilot phase*, the consortium asked participating countries to establish at least three focus groups, one including ECEC centre leaders and the other two including ECEC staff. The purpose of these groups was to discuss the draft forms of the TALIS Starting Strong 2018 ECEC leader and staff questionnaires and to provide feedback on their functioning, cultural applicability and other aspects (see also Chapter 3).

The consortium required all participating countries to run the *field trial* according to the standardised procedures outlined in the TALIS Starting Strong 2018 technical standards (see Annex B). Countries that had opted to participate in both populations of interest had to trial the survey instruments for both of them.

Figure 1.1. Overview of key TALIS Starting Strong 2018 phases



The *main survey* was implemented according to countries' yearly education schedules. Data collection took place primarily between March and June 2018 in countries on a Northern Hemisphere schedule and between August and November 2018 in countries on a Southern Hemisphere schedule.

The consortium held the first meeting of NPMs in June 2016 in Munich, Germany, prior to administration of the pilot study. The purpose of the meeting was to present and discuss the survey's draft conceptual framework, the sampling procedures, the roadmap for instrument and materials development, the planned survey operations procedures and the overall project governance and responsibilities. NPMs were also introduced to the overall schedule for the survey, means of communication, and best practices, with the latter collated from previous TALIS cycles.

1.6.1. Pilot study

The main purpose of the pilot study was to test understanding of the questionnaire items by "regular" respondents (ECEC staff and ECEC centre leaders) in a variety of national settings. The goal was to use the outcomes of the pilot to fine-tune the questionnaires for the field trial. The pilot questionnaires included items adapted from the TALIS 2018 pilot questionnaires as well as new items developed for the ECEC context. All items were checked by focus group participants with respect to the following aspects:

- sufficient conceptual coverage from the perspectives of ECEC centre leaders and ECEC staff
- conceptual understanding and clarity
- international and cultural applicability

- feasibility of new, modified or alternative formats of new and adapted items developed by the QEG
- fine-tuning the questionnaires for the field trial based on analysis of the pilot outcomes.

In August 2016, the ISC sent the pilot study guidelines to the national study centres, and in September 2016 held a webinar on how to prepare and implement the pilot. By the end of September 2016, all participating countries had familiarised themselves with the content of the pilot study questionnaires.

The pilot study was conducted in October 2016 with ECEC centre leaders and staff within the countries who represented different locations and different cultural and language backgrounds. Countries that opted to administer both populations of interest were asked to form focus groups of participants from both populations. These groups consisted of seven to nine people who met face-to-face to discuss the pilot instruments. The ISC asked countries to translate the pilot questionnaires. Because the staff questionnaire included more items than would be used in the later survey phases and because of time constraints, participating countries had two focus groups, with each focused on only one half of the questionnaire materials. The length of the leader questionnaire was regarded as being of a length that just one focus group could readily discuss.

1.6.2. Field trial

The purpose of the field trial was two-fold. The first purpose was to gather as much quantitative data as possible to allow the instruments to be fine-tuned through use of statistical procedures and to enable substantive revisions to items in preparation for the main survey. The second purpose was to test all operational procedures in all participating countries and for both populations of interest before administration of the main survey. These procedures included, for example, within-school sampling, national instrument production, survey operations, data collection and data entry.

The second meeting of NPMs took place in January 2017 in Athens, Greece. NPMs and consortium members discussed the outcomes of the pilot study and the required changes to the field trial instruments. The NPMs also had opportunity to discuss the following with the consortium members: sampling plans, individual survey preparation schedules, and strategic plans designed to achieve high participation rates. After the NPM meeting, the consortium conducted a data management seminar to train national data managers in using the IEA software and applying related survey procedures.

The field trial questionnaires and all accompanying materials were released in February 2017 so that the national study centres could start their preparations for the field trial. Because the field trial staff questionnaires contained too many questions for any one respondent to answer within the time available, the consortium used a rotated questionnaire design during the field trial. This approach not only reduced the total response time and burden on staff members, but also enabled testing of different question formats (see Chapter 3 of this report for more details).

The minimum field trial sample size per country and population of interest was set at 30 sampled ECEC centres. However, depending on the ECEC structure of a country and the size of its centres, field trial sample sizes could reach a maximum of 80 sampled ECEC centres (see Chapter 5 for more details).

Five of the nine participating countries administered their questionnaires on line during the field trial, three countries applied a mixed-mode model of online and paper questionnaires and one country administered the survey on paper only. All countries collected their data between May and June 2017. NPMs were asked to implement a national quality observation programme to gather onsite information on the preparation and implementation of the survey (see Chapter 7).

NPMs and their teams monitored data collection and supervised manual data entry (if applicable). They then submitted their national datasets to the ISC for processing and quality checks. Extensive analyses of the field trial data by IEA's analysis team and the QEG along with consultations with the TAG, the OECD Secretariat and feedback from the Extended ECEC Network members and NPMs led, after the field trial, to key improvements in the questionnaires (refer to Chapter 3 for more details).

1.6.3. Main survey

The third NPM meeting was held in November 2017 in Reykjavik, Iceland. Its purpose was to prepare NPMs for implementation of the main survey in accordance with the outcomes of the field trial and to review and discuss the changes introduced to the draft main survey instruments. All participating countries continued to discuss sampling and survey operation procedures with the consortium; each country also engaged in individual consultations with the consortium on the field trial data.

The main survey questionnaires and associated materials were released at the end of 2017 so that the national study centres could begin their instrument adaptation and translation work. During the data collection period, national study centres implemented their national quality observation programmes. In addition, and in contrast to practice during the field trial, the consortium implemented an international quality observation programme that involved sending independent, externally appointed quality observers to a subset of participating centres (see Chapter 7).

Most of the Northern Hemisphere countries administered the survey within a self-selected period during the first and second quarters of 2018, with data submissions in June and July 2018. Two countries extended their data collection window towards October and November 2018 in order to improve their participation rates. Southern Hemisphere countries collected their main survey data between August and early November 2018 and submitted their data in November 2018. All data were then processed and cleaned at the ISC (see Chapter 8), after which the IEA's sampling and weighting team carried out weighting procedures during the second half of 2018 through to January 2019.

From the beginning of the third quarter of 2018, all available main survey data were transferred to the IEA's analysis team for analysis and table production (see Chapter 12). In September 2018, the consortium, with the OECD Secretariat in attendance, met for one day to review and adjudicate the available main survey data. A second adjudication session took place in mid-January 2019 in order to adjudicate the data for the remaining countries. (For more details on the adjudication process, please see Chapter 10.) At the end of January 2019, all data were transferred to the IEA's analysis team for scaling (see Chapter 11).

In October 2018, the fourth NPM meeting took place in Seoul, Korea. The purpose of this meeting was to finalise all data collection work, to sign-off on data and to review (in plenary sessions) the first set of draft tables containing preliminary data for the first international report of the survey's findings. Main survey data and sampling and weighting outcomes

were discussed in one-on-one consultations between each participating country and the ISC. During the meeting, the consortium asked NPMs to provide important feedback about the main survey collection experiences in their countries so that they could use that information to inform future cycles of TALIS Starting Strong 2018.

In March 2019, the NPMs and other national study centre staff received training in the correct use of the international database so that they could either replicate the analyses presented in the international report or run their own analyses for inclusion in their respective national reports on the survey's findings.

Finally, participating countries received different versions of their national data for verification purposes and, later in 2019, gained access to the draft international database (see Chapter 8 for more details) so that they could start preparing their national reports.

Notes

1. The TAG included the following experts (in alphabetical order): Pascal Bressoux (Université Pierre-Mendès-France, Grenoble), Timothy L. Kennel (Decennial Statistical Studies Division, United States Census Bureau), Paul Leseman (University of Utrecht), Bart Meuleman (Katholieke Universiteit Leuven) and Christian Monseur (University of Liège).

2. The IEA Windows Within-School Sampling Software (WinW3S) is a software program originally developed by the IEA for international large-scale assessments administered in the school context (e.g. to sample students and/or teachers within schools).

References

Biemer, P. and L. Lyberg (2003), *Introduction to Survey Quality*, John Wiley & Sons, Hoboken, NJ. [9]

Martin, M., K. Rust and R. Adams (1999), *Technical Standards for IEA Studies*, International Association for the Evaluation of Educational Achievement, Amsterdam. [10]

Melhuish, E. et al. (2015), *Curriculum Quality Analysis and Impact Review of European Early Childhood Education and Care*, European Commission, Brussels. [3]

National Association for the Education of Young Children (1991), *National Association for the Education of Young Children's Accreditation Criteria and Procedures*, NAEYC, Washington, DC. [4]

OECD (2014), *TALIS 2013 Technical Standards*, OECD Publishing, Paris. [11]

OECD (2013), *From policy questions to new ECEC indicators: draft background paper for the 14th meeting of the OECD ECEC Network*, OECD, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2013)9&docLanguage=En. [1]

Sim, M. et al. (2019), "Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework", *OECD Education Working Papers*, No. 197, OECD Publishing, Paris, https://dx.doi.org/10.1787/106b1c42-en. [8]

Statistics Canada (2009), *Quality Guidelines*, Statistics Canada, Ottawa. [12]

Tietze, W. et al. (2013), *Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)*, [National Investigation into Education, Care and Education in Early Childhood], NUBBEK, Berlin, http://www.nubbek.de/media/pdf/NUBBEK%20Broschuere.pdf. [5]

Tietze, W., H. Roßbach and K. Grenner (2005), *Kinder von 4 bis 8 Jahren: Zur Qualität der Erziehung und Bildung in Kindergarten, Grundschule und Familie*, [Children from 4 to 8 Years: The Quality of Education in Kindergarten, Elementary School and Family], Beltz Verlag, Weinheim. [6]

UNESCO-UIS (2012), *International Standard Classification of Education (ISCED) 2011*, UNESCO Institute for Statistics, Montreal, http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf. [2]

World Health Organization (1990), *WHO Child Care Facility Schedule: With User's Manual*, World Health Organization, Geneva. [7]

Chapter 2. Development of the conceptual framework

This chapter describes the process that led to the development of the conceptual framework underpinning the survey and which culminated in publication of the "Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework" (Sim et al., 2019[1]).

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

2.1. The general purpose of TALIS Starting Strong and its policy focus in 2018

TALIS Starting Strong is a large-scale international survey of ECEC staff and leaders in ECEC centres. It was initiated in line with the OECD's long-term data development strategy and a suggested data collection roadmap (OECD, 2012[2]; 2013[3]).

The overall objective of TALIS Starting Strong 2018 was to draw robust international indicators from and provide policy-relevant analysis of data pertaining to ECEC staff and centre leaders, their pedagogical and professional practices, and the learning and well-being environments in ECEC centres, in order to help countries review and develop policies that promote conditions for positive child development, well-being and learning.

The survey's particular aims were to describe how characteristics of ECEC staff and centre leaders, their pedagogical and professional practices and learning and the well-being environments in their centres vary within and across countries, and eventually over time. The learning and well-being environment and workforce indicators addressed by the survey are those that experts in the field of ECEC consider are associated with children's positive development and learning outcomes, but who also acknowledge that indicators will inevitably be influenced by cultural norms and values across countries.

The guiding principles of TALIS Starting Strong 2018 aligned with the principles underpinning the OECD Teacher and Learning International Survey (TALIS) 2018 and were as follows:

- *Policy relevance:* Clarity on key policy issues and maintaining a focus on the questions that are most relevant for participating countries need to be essential facets of the survey.
- *Value added:* International comparisons should be a significant source of the survey's benefits.
- *Indicator oriented:* The results should yield information that can be used to develop indicators.
- *Validity, reliability, comparability and rigour:* In accordance with a rigorous review of existing research, the survey should yield information that is valid, reliable and comparable across participating countries.
- *Interpretability:* Participating countries should be able to interpret the results in a meaningful way.
- *Efficiency and cost-effectiveness:* The work should be carried out in a timely and cost-effective way.

2.2. Choosing the policy focus for TALIS Starting Strong 2018

Together with the OECD Early Childhood and Education and Care Network (ECEC Network), the OECD Secretariat defined three policy issues as the central ones to examine during the first cycle of TALIS Starting Strong:

- *Policy issue 1:* ensuring the quality of learning and well-being environments
- *Policy issue 2:* motivating, attracting and retaining staff to the profession
- *Policy issue 3:* developing staff for and within the profession.

The OECD Secretariat and the ECEC Network also determined that staff- and centre-level contextual information that cuts across these different policy issues would be needed to enable rigorous interpretation of the results (Loizillon, 2016[4]).

In keeping with these determinations, the OECD Secretariat carried out a priority rating exercise in May and June 2015. The exercise was based on the preliminary framework document prepared by the OECD Secretariat, discussions during meetings of the OECD Extended Early Childhood and Education and Care Network (Extended ECEC Network),[1] consultations with countries interested in participating in the survey, and written submissions containing comments from those countries. Nine countries representing a wide variety of geographical and cultural backgrounds participated voluntarily in the exercise.[2]

The goal of the priority rating exercise was to obtain countries' preferences regarding (1) the structure of the questionnaires, that is, whether the questionnaires should cover a wide range of topics (breadth) or focus on a smaller number of topics covered in more detail (depth); and (2) the themes and indicators that should be given priority consideration for inclusion in the questionnaires.

This part of the rating exercise was organised under the three main policy areas, each of which encompassed several themes. Fifteen themes in total were presented for consideration: five for policy issue 1, six for policy issue 2, and four for policy issue 3. Each theme was accompanied by a number of possible indicators (81 in total across the themes) that could be used to inform the questions designed to collect data on the themes.

The first part of the priority rating exercise concerned the structure of the questionnaires that would be used in the survey. The OECD Secretariat invited the countries participating in the exercise to allocate, according to their preferences, 100 points across the different categories of breadth and depth and were advised that the larger the number of points apportioned to a category, the stronger the preference for that category. The secretariat also asked countries how many themes they thought the questionnaires should cover. The response categories for this question were "between 2 and 5 themes", "between 6 and 9 themes", "between 10 and 12 themes", and "between 13 and 15 themes". Countries expressed a clear preference against having fewer than six themes; the strongest preference was for six to nine themes (Table 2.1).

Table 2.1. Countries' preferences regarding the breadth vs. depth of the questionnaires (based on responses from nine countries)

Breadth and depth of coverage options	Rating points (100)
The questionnaires should cover between 2 and 5 themes	1.11
The questionnaires should cover between 6 and 9 themes	36.11
The questionnaires should cover between 10 and 12 themes	30.56
The questionnaires should cover between 13 and 15 themes	32.22
Total (should add-up to 100)	100

The aim of the second part of the priority rating exercise was to determine which themes and indicators should be given priority consideration for inclusion in the questionnaires. The OECD Secretariat invited countries to apportion 100 rating points across 15 themes and then to rank (ordinally) the indicators according to their priority within those themes. Because questions regarding staff and centre characteristics would be included in the questionnaires by default, these were not included in this part of the rating exercise.

The table in the annex of this chapter, which presents the results of this rating exercise and details all included themes and indicators, shows that countries regarded some themes as very high priority (the themes with the higher values in the annex table). These were:

- staff pedagogical practices, beliefs and self-assessment
- centre environments, staff beliefs on process quality,[3] and staff self-assessment
- working time and workload
- pre-service education and training
- in-service education and training
- leadership by centre heads/leaders
- staff professional practices
- job satisfaction.

Countries considered themes such as innovative practices and evaluation of less importance (the themes with the lower values in the annex table). Please note that in regard to the scores for indicators, a lower value corresponds to higher priority; see the column "Priority indicators" in the annex table.

Although this exercise was useful in providing guidance for development of the survey's conceptual framework and questionnaires, countries' rankings varied significantly, and the overall highest-rated themes matched the priorities of some countries more closely than they matched priorities in the other countries. In addition, not all countries implementing TALIS Starting Strong 2018 participated in the priority rating exercise because it took place before countries committed to taking part in the survey. Finally, not all countries that participated in the exercise implemented TALIS Starting Strong 2018.

Based in part on results from the priority rating exercise, and following further discussions with stakeholders, the OECD Secretariat and the ECEC Network decided that TALIS Starting Strong 2018 should include at least six themes that, in combination, would inform all three policy issues identified above. Discussions between the OECD Secretariat, the OECD Early Childhood and Education and Care Network (ECEC Network), and the Extended ECEC Network revealed the importance of several themes additional to those listed above. One of these themes concerned issues surrounding equity and diversity among children attending ECEC settings (i.e. equity and diversity in the child group). The parties involved in the discussion eventually concluded that the equity and diversity theme was encapsulated in the substance of each of the above themes, but that it was nevertheless a theme of high contemporary policy importance.

The final inclusion of themes and indicators in the TALIS Starting Strong 2018 main survey questionnaires ultimately depended on the length of the questionnaires and the results of the study's pilot phase and field trial.

2.3. Developing the conceptual framework

The TALIS Starting Strong 2018 conceptual framework provided an integrated theoretical and analytical underpinning to the survey, by articulating the survey's research foci and its links to existing knowledge and evidence and policy questions. The framework was developed by the survey's Questionnaire Expert Group (QEG) in collaboration with the TALIS Starting Strong Consortium framework development team led by Agnes Stancel-

Piątak (the IEA Hamburg) with contributions from Julie Bélanger (RAND Europe), who acted as chair of the QEG, Lynn Karoly (the RAND Corporation) and Megan Sim (RAND Europe).

The TALIS Starting Strong Consortium appointed experts to the QEG after consulting and reaching agreement with the OECD Secretariat. Having called on the OECD Secretariat and the Extended ECEC Network to nominate suitable people, the consortium determined and approved the group's final composition. Three key criteria governed selection: expertise in the main policy themes selected for the survey, experience in survey-based policy research, and expertise in relation to survey instrument development in an international context. The consortium also endeavoured to ensure the QEG experts represented the geographical and cultural contexts of the participating countries.

The QEG's overall task was to develop the conceptual framework representing the conceptual underpinning of TALIS Starting Strong 2018. The development work was based on sound theoretical principles and an understanding of the state of the art in the relevant areas. The work also took into account the policy imperatives of the countries participating in the survey and the methodological demands of large international surveys. The QEG's specific tasks included the following:

- *Developing concept notes:* These provided information about the participating countries' priorities and the TALIS Starting Strong theoretical background, justification, development directions, its overlap with TALIS 2018, and its analytical potential and indicators. Based on background documents, the concept notes also included the initial work on the framework undertaken by the OECD (Loizillon, 2016[5]), the priority rating exercise described above, and relevant empirical literature.
- *Drafting thematic sections of the conceptual framework that corresponded to the questions and items in TALIS Starting Strong 2018:* This work was also informed by the survey's theoretical underpinnings. It paid attention to existing relevant research and the overall aims for the survey and referenced the TALIS 2018 conceptual framework.
- *Revising drafts of the conceptual framework:* The revision work responded to feedback received from the OECD Secretariat, the ECEC Network and the Extended ECEC Network, the national project managers (NPMs) in the participating countries, and the Technical Advisory Group (TAG). It also took into consideration the parallel development of the TALIS 2018 conceptual framework.

The OECD Secretariat commissioned its international contractors, the International Association for the Evaluation of Educational Achievement (IEA) and RAND Europe, to establish the QEG, chair the QEG meetings, organise and manage communication amongst QEG members and ensure that the QEG produced, on time, proposals covering sample design aspects and setting out implications and needs pertaining to survey operations. The QEG's core members were:

- *(Chair) Julie Bélanger, RAND Europe, United Kingdom,* not only chaired the QEG but also had overall responsibility for the group. Furthermore, she had responsibility for the sections of the instruments covering staff and leader background and centre background. Julie Bélanger was assisted by Miriam Broeks, Megan Sim and Katherine Stewart (all from RAND Europe).

- *Alejandra Cortazar, Centro de Estudios Primera Infancia, Chile,* was responsible for the conceptual framework themes relating to the initial preparation and continuing professional development themes.
- *Edward Melhuish, University of Oxford, United Kingdom,* had joint responsibility with Henrik Zachrisson for the theme focused on the process quality of staff–child interactions. Edward Melhuish was also responsible for the following themes: monitoring and assessment of children's development, well-being and learning; professional beliefs about children's development, well-being and learning; and leader and staff self-efficacy.
- *Henrik Daae Zachrisson, University of Oslo, Norway,* was responsible, along with Edward Melhuish, for the process quality of staff–child interaction theme. He was also responsible for the themes of structural quality characteristics and equity and diversity in the child group.
- *Masatoshi Suzuki, Hyogo University of Teacher Education, Japan,* was responsible for the themes of pedagogical and administrative leadership; centre climate; stakeholder relations; and staff and leader well-being. He was assisted by Sakiko Sagawa (Nara University of Education, Japan) and Yumi Yodogawa (University of Tokyo, Japan).

In addition to these QEG experts, the following individuals served as invited experts within the QEG (i.e. the extended QEG). They provided advice relating to particular issues of interest at various stages of the framework and/or instrument development.

- *Nirmala Rao, Hong Kong University, China,* provided input into the applicability of the instruments for middle-income countries.
- *Susanne Viernickel*, *University of Leipzig*, *Germany* provided advice on the applicability of the instruments for centres catering to children under three years of age.
- *Sharon Lynn Kagan, Teachers College, Columbia University, United States,* provided feedback on the conceptual framework.
- *Pauline Slot, Utrecht University, the Netherlands,* was jointly responsible with Trude Nielsen for developing the questionnaires' situational judgement items commissioned by the OECD Secretariat.
- *Trude Nielsen, University of Oslo, Norway*, along with Dr Slot, developed the questionnaires' situational judgement items commissioned by the OECD Secretariat.

Ex officio members from the OECD Secretariat, the international contractor and its partners included:

- *Ralph Carstens, the IEA Hamburg, Germany,* as chair of the TALIS QEG, he was responsible for helping to ensure alignment between the TALIS 2018 and the TALIS Starting Strong 2018 instruments. Given his extensive involvement in previous TALIS cycles, Ralph Carstens also advised on operational and data implications.
- *TALIS Starting Strong 2018 framework development team:* The members of this team were Lynn Karoly of the RAND Corporation, Megan Sim from RAND Europe and Agnes Stancel-Piątak from the IEA Hamburg.

- *Jean Dumais, Statistics Canada,* provided advice to the group from a sampling perspective.
- *OECD Secretariat team:* As representatives from the OECD Secretariat, Clara Barata, Arno Engel and Miho Taguma ensured that the priorities and goals of the countries participating in TALIS Starting Strong 2018 were upheld.
- *Fons van de Vijver, Tilburg University, the Netherlands,* chaired the TAG and developed methods and approaches to lessen the impact of response-style bias.

The QEG also profited from additional input and feedback received from individuals and groups at various stages of the study:

- *OECD Secretariat:* The members of the secretariat team included Yuri Belfali, Montserrat Gomendio and Andreas Schleicher as well as members of the TALIS and Innovative Teaching for Effective Learning project teams.
- *Trade Union Advisory Council:* The two members from the council were Martin Henry and Anna Tornberg.
- *TALIS and TALIS Starting Strong 2018 Technical Advisory Group*: Fons van de Vijver (Tilburg University) chaired the group. He was accompanied by Pascal Bressoux (Université Grenoble Alpes), Timothy Kennel (US Census Bureau), Paul Leseman (Utrecht University), Bart Meuleman (KU Leuven) and Christian Monseur (University of Liège).
- *TALIS Starting Strong Consortium members*: In addition to RAND Europe staff and Jean Dumais, the consortium members also included Viktoria Gabriele Böhm, Juliane Hencke, Steffen Knoll , Juliane Kobelt from the TALIS Starting Strong 2018 International Study Centre at the IEA, as well as, Alena Becker, Christine Busch, Karsten Penon and Sabine Meinck all from the IEA.

Between April 2016 and June 2018, the QEG members continued to revise the conceptual framework, taking into account the ongoing development of the survey instruments, results from the pilot and field trial phases of the survey, and feedback received from the members of the consortium, the OECD Secretariat, the Extended ECEC Network, and the Extended QEG members Sharon Lynn Kagan, Nirmala Rao and Susanne Viernickel.

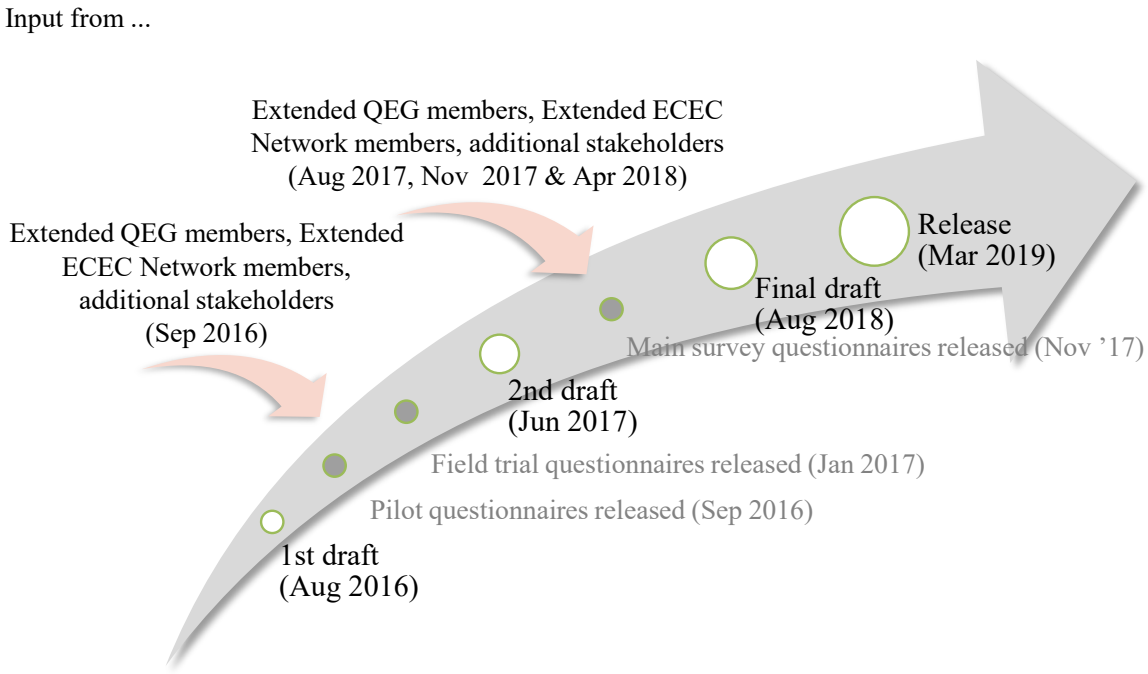
Figure 2.1 shows the development timeline for the TALIS Starting Strong 2018 conceptual framework alongside the development milestones for the survey's instruments. The figure also presents the key points when feedback was received during the development process.

Overall editing of the conceptual framework was provided by Julie Bélanger (RAND Europe and chair of the QEG), Lynn Karoly (the RAND Corporation), Megan Sim (RAND Europe) and Agnes Stancel-Piątak (the IEA Hamburg).

The final version of the conceptual framework (August 2018) has three sections:

- Introduction
- Section I: General purpose and policy relevance of TALIS Starting Strong 2018
- Section II: Knowledge relating to the themes and main indicators.

Figure 2.1. Conceptual framework development timeline



2.4. Connecting the conceptual framework with proposed analyses

Each thematic chapter in section II of the conceptual framework is based on the theoretical and empirical literature reviewed and includes a subsection titled 'Analytical potential and indicators'. The subsection describes, with reference to the literature reviewed for each theme, potential within- and cross-theme analyses that can be conducted using data collected during TALIS Starting Strong 2018. The QEG experts Alejandra Cortazar, Professor Edward Melhuish, Masatoshi Suzuki and Henrik Daae Zachrisson contributed the 'analytical potential and indicator' subsections in each of their respective chapters.

After completion of the conceptual framework, the consortium developed, in consultation with the OECD Secretariat, the QEG and the TAG, a document presenting principles to support data analysts involved in reporting the national and international findings of TALIS Starting Strong 2018. More specifically, and as proposed in the OECD's TALIS Starting Strong 2018 draft reporting plan, the analysis principles document provides high-level guidance on the content- and context-related issues that individuals need to consider when analysing the survey's data and reporting on the resultant findings. This guidance also applies to upcoming analyses conducted by the participating countries. The document furthermore aims to inform the international analysis and reporting of TALIS Starting Strong 2018 data conducted by the OECD Secretariat and any external experts contracted by the secretariat. Finally, the documents provide the NPMs with guidance on analysing and reporting their national-level data.

Annex 2.A. Results of priority rating exercise of themes and indicators

Annex Table 2.A.1. Countries' preferences regarding priority themes and indicators for inclusion in the survey instruments (based on responses from nine countries)

Themes and indicators to rate		ISCED level 0	
		Priority themes (Points, i.e. higher value corresponds to higher priority)	Priority indicators (Order, i.e. lower value corresponds to higher priority)
Total		**100**	
Policy issue 1: Ensuring quality of learning and well-being environments			
Theme 1.1 Staff's pedagogical practices, staff beliefs and self-assessment		9.67	
Indicators	Pedagogical practices in target groups, such as classroom and playtime (e.g. child-centred activity, teacher-guided activities, child–child interactions; co-constructive, instructive; free play, guided play; etc). The categorisation will be built on the 'ECEC network survey on pedagogy' with a caution on oversimplification).		3.44
	Beliefs about teaching/caring – orientation quality (e.g. educational values, attitudes with regard to the importance of different educational areas and learning goals, practitioner's definition of their professional role) (staff questionnaire and centre head questionnaire).		3.89
	Staff's actual use of curriculum in the target groups (staff questionnaire and centre head questionnaire). NOTE: we are aware that the term "curriculum" is interpreted differently in different countries. Some countries do not have a national or regional curriculum but have learning standards or curriculum/content frameworks/guidelines. Please feel free to contact the OECD Secretariat for clarification when you have questions about the meaning of curriculum.		5.22
	Types of staff practices on monitoring child development and outcomes and the staff's use of the results (staff questionnaire and centre head questionnaire).		5.22
	Use of instructional time, play time, etc. Time allocation across different activities/content areas (e.g. language and communication, outdoor play, colouring, singing, versus free time/free play) and staff beliefs on how these activities could enhance child learning and well-being.		5.67
	Staff's views regarding barriers to implementing a variety of practices, factors hindering instruction/play (centre head/leader and staff attitudinal responses) and incentives for staff to implement changes in pedagogical practices.		7.44
	Types of activities designed to promote language/literacy and mathematics that are done and how often, and how different child behaviours are viewed and handled by staff (staff questionnaire and centre head/leader questionnaire)		7.67
	Use of physical makeup of the space, i.e. classroom/playroom (e.g. indoor/outdoor) and materials (e.g. books, toys) with respect to certain pedagogical practices (e.g. play, co-operation, conflict resolution, sharing, problem-solving) (staff questionnaire and centre head/leader questionnaire).		7.78
	Individualised/differentiated teaching and caring in target groups (staff questionnaire and centre head questionnaire).		7.78
	Staff self-assessment of general pedagogical knowledge (knowledge on curriculum and how young children learn; instructional processes, student learning, formative assessment) – possible co-operation with the OECD CERI ITEL for ISCED level 0.		8.00
	Staff's use of curriculum to support parents (e.g. helping parents to support children's learning at home by using the curriculum to enhance home learning environments).		9.11
	Staff perception of child engagement in the target groups (staff questionnaire and centre head questionnaire).		9.44
	Frequency of pedagogical activities using ICT in target groups and staff beliefs on what kind of learning and well-being will be enhanced for the children (staff questionnaire and centre head/leader questionnaire).		10.33

Themes and indicators to rate		ISCED level 0	
		Priority themes (Points, i.e. higher value corresponds to higher priority)	Priority indicators (Order, i.e. lower value corresponds to higher priority)
Theme 1.2 Environments (e.g. climate and composition of classroom/playroom), staff beliefs on process quality, and staff self-assessment		9.78	
Indicators	Climate in target group (e.g. composition, group/class size, staff–child ratio, group size and group organisation).		1.89
	Staff's beliefs about important aspects to create a supportive environment for learning (e.g. quality and frequency of child–staff interactions, child–child interactions) and staff's beliefs about how to improve these aspects for better process quality (both staff questionnaire and centre head/leader questionnaire).		2.44
	Time distribution in target groups.		4.78
	Classroom or group management in target class.		4.89
	Staff's perceptions of the impact of group/class size, staff–child ratio, group size on staff's ability to provide effective teaching and caring.		5.11
	Staff's self-assessment of classroom/playroom management skills.		6.00
	Staff's perceptions of the effective use of resources (e.g. "If you had an additional 1000 euros, would you spend it to pay staff better, make group sizes smaller, buy better learning materials, build a nicer infrastructure, etc?).		6.44
	Staff self-assessment of non-cognitive skills/patience/motivation.		6.67
	Staff's views regarding barriers to classroom/playtime management.		6.78
Theme 1.3 Staff's professional practices		7.11	
Indicators	Collaboration among staff (staff questionnaire and centre head/leader perceptions).		1.44
	Communication/collaboration with parents, types of information staff communicate with parents, types of activities parents participate in, and staff/provider beliefs on what role parents should play (staff questionnaire and centre head/leader questionnaire).		2.33
	Collaboration with persons from other sectors (e.g. social workers, health specialists) and/or with a wider community (e.g. community leaders, museums, libraries, local industries) (staff questionnaire and centre head/leader questionnaire).		2.67
	Staff's participation in decision making at the centre (staff questionnaire and centre head/leader questionnaire).		3.00
	Participation in ECEC policy making and improving quality of ECEC (staff questionnaire and centre head/leader questionnaire).		3.89
Theme 1.4 Innovative practices and evaluation		2.11	
	Staff's perception of their pedagogical autonomy for innovation.		0.67
	Incentives for staff to adopt innovative pedagogical practices.		0.78
	Staff evaluation and dissemination of innovative practices in centres.		1.00
	Types of innovation in your centre in past year (technological, pedagogical, R&D, organisational/administrative, etc).		1.22
	Types of innovation in the target groups in the current or past school year.		1.33
Theme 1.5 Leadership by centre heads/leaders		7.33	
Indicators	Role and function of the centre leader (administrative/managerial/ organisational leadership and pedagogical leadership).		3.11
	Training and development of centre leaders/heads.		3.89
	Staff perception of centre leadership (staff responses).		4.22
	Staff satisfaction with centre policies (staff responses).		4.78
	Centre leaders' workload.		5.22
	Centre leaders' job satisfaction.		5.44
	Centre ethos (e.g. goal driven, high aspirations, community engagement).		5.56
	Centre autonomy in key areas (hiring and dismissing teachers, career ladders, pay, funding, etc).		5.67

Themes and indicators to rate		ISCED level 0	
		Priority themes (Points, i.e. higher value corresponds to higher priority)	Priority indicators (Order, i.e. lower value corresponds to higher priority)
	Centre leaders' perceptions of challenges in funding and resource management.		6.11
	Leaders' working hours.		7.22
	Attracting effective centre leaders.		7.44
Policy issue 2: Motivating, attracting and retaining staff to the profession			
Theme 2.1 Working time and workload (both staff and centre heads)		9.67	
Indicators	Types of activities which time is spent on (e.g. teaching/caring practice, planning time, time for reflection, working in team).		1.89
	General working time use (e.g. overall working time compared with time spent on teaching/caring during a typical week).		2.33
	Workload and work duties (including e.g. teaching/caring in indoor/ outdoor, administrative and extra-curricular duties).		2.89
	Group or class demographics of children in target groups (e.g. gender, language background, children with disabilities) and types of support in place for staff to work with various populations (e.g. children with special needs, children with varying language backgrounds).		2.89
Theme 2.2 Job satisfaction		6.67	
Indicators	Overall job satisfaction with ECEC as a profession (compared with general job satisfaction with different job profiles in PIAAC).		1.33
	Job satisfaction with current ECEC pedagogical environment in target groups (e.g. staff–child ratio for pedagogical purpose, staff support, resources).		2.33
	Job satisfaction with current work environment in target groups (e.g. staff–child ratio as an element of work environment, contract types, wages and non-financial benefits).		2.56
	Perception of the value of the profession (centre head/leader and staff responses).		3.44
	Likelihood of leaving the profession.		3.67
Theme 2.3 Recognition, reward and evaluation of staff		5.22	
Indicators	Centre policies that recognise, reward and evaluate staff – frequency, criteria, sources of data, outcomes (centre heads/leaders provide centre-level data and staff provide personal responses).		2.22
	Career progression and prospects of staff (centre head/leader and staff personal responses) (including aspirations to leadership positions, local or national government officer positions, glass-ceiling effects).		2.44
	Expectations on the types of recognition (e.g. non-financial measures, salary increase) (by staff response and centre head/leader perceptions).		2.56
	Unmet needs for appraisal and feedback.		2.89
	Perceptions of the impact of policies that recognise, reward and evaluate staff (centre head/leader and staff perceptions).		3.11
	Recognition for being innovative in pedagogical practices.		3.11
Theme 2.4 ECEC workforce supply and demand and recruitment strategies		5.56	
Indicators	Levels and distribution of staff shortages (centre head/leader responses).		0.89
	Criteria of recruitment for qualified staff, e.g. types of knowledge, skills and behaviour needed for the profession (centre head/leader and staff responses).		1.22
	Effectiveness and satisfaction with recruitment procedures (centre head/leader and staff responses).		1.56
	Decision-making responsibilities in recruitment (centre heads/leaders).		1.89

Themes and indicators to rate		ISCED level 0	
		Priority themes (Points, i.e. higher value corresponds to higher priority)	Priority indicators (Order, i.e. lower value corresponds to higher priority)
Theme 2.5 Attracting good students into ECEC study programmes/the ECEC profession		3.22	
Indicators	Appreciation of ECEC profession by students, parents, principals and society (staff and centre head point of view).		1.11
	Staff satisfaction with their career choice (including whether they would recommend the profession to their students or children).		1.22
	Alternative pathways to the ECEC profession.		2.33
Theme 2.6 Staff attrition and turnover rates		5.67	
Indicators	Staff turnover and attrition (centre leader responses).		1.11
	Causes of staff turnover and attrition, as well as career paths (years of experience, past job histories/career changes, if any) (staff responses).		1.56
Policy issue 3: Developing staff for and within the profession			
Theme 3.1 Pre-service education and training		10.22	
Indicators	Characteristics of pre-service education and training (e.g. provider and level, length/duration).		1.67
	Pedagogy and contents of pre-service education and training (e.g. lecture, seminar, workshop, practice; ECEC-specific contents, child development, communication with parents, different set of pedagogies).		1.78
	Length/duration of practical training and other requirements upon certification/qualification		2.56
Theme 3.2 Satisfaction and effectiveness of pre-service education and training		4.67	
Indicators	Perceived impact of pre-service education and training.		1.00
	Development needs and unsatisfied demand.		1.33
	Staff's rationale for choices of ECEC as a study programme.		1.67
Theme 3.3 In-service education and training		8.11	
Indicators	Types and content areas of formal in-service education and training (e.g. curriculum-focused, new pedagogy-focused, behaviour and health-focused, communication with parents).		1.44
	Types of staff support.		2.11
	Types and content areas of informal in-service education and training (e.g. practitioner-initiated networks, online learning, including MOOCs, etc).		2.44
Theme 3.4 Satisfaction, take-up and effectiveness of in-service education and training		5.00	
Indicators	Perceived impact of in-service education and training.		1.89
	Development needs and unsatisfied demand.		2.33
	Personal and centre objectives of in-service education and training.		2.44
	School priorities for staff's professional development.		2.44
	Barriers for further engagement in in-service education and training and views about how to overcome such barriers/incentives (staff as well as centre heads/leaders employers to support staff participation).		2.56
Total		100	

Notes

1. The Network on Early Childhood Education and Care aims at assisting the Education Policy Committee in supporting Members and Partners in the development of effective policies and practices in the field of early childhood education and care to promote better social, cultural, educational and economic outcomes for children. The network is composed of experts with responsibility for ECEC policies and/or policy research in that area, depending on national governance structures. A sub-group of the network, established specifically for TALIS Starting Strong and named the Extended ECEC Network on the TALIS Starting Strong, brought together participating and interested countries for more detailed and technical discussions on this specific project. Because some of the participants in the two networks overlapped, the subgroup held, whenever possible, meetings back-to-back with the meetings of the TALIS Governing Board, which oversaw the TALIS 2018 project.

2. Countries that provided ratings were Germany, Israel, Japan, Kazakhstan, Korea, Luxembourg, Norway, Turkey and the United States.

3. Process quality refers to the process quality of staff–child interactions. The TALIS Starting Strong Conceptual Model differentiates conceptually and analytically between staff–child interaction, ECEC centre characteristics, and staff and leader characteristics. Factors related to the direct interaction between ECEC staff and children concern the process quality of the interaction, as well as monitoring child development, well-being and learning (Sim et al., 2019[1]).

References

Loizillon, A. (2016), *Towards a conceptual framework for an international ECEC staff survey: evidence paper*, https://one.oecd.org/document/EDU/EDPC/ECEC/RD(2016)2/ANN1/en/pdf. [4]

Loizillon, A. (2016), *Towards a conceptual framework for an international ECEC staff survey: evidence paper*, http://dx.doi.org/EDU/EDPC/ECEC/RD(2016)2/ANN1. [5]

OECD (2013), "From policy questions to new ECEC indicators: draft background paper for the 14th meeting of the OECD Early Childhood Education and Care Network", *EDU/EDPC/ECEC(2013)9*, OECD Publishing, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2013)9&docLanguage=En. [3]

OECD (2012), "Revised project proposal of new policy output on early learning and development, Annex 2: Existing data for international comparison", *EDU/EDPC/ECEC(2012)3/REV1/ANN2*, OECD Publishing, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/EDPC/ECEC(2012)3/REV1/ANN2&docLanguage=En. [2]

Sim, M. et al. (2019), "Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework", *OECD Education Working Papers*, No. 197, OECD Publishing, Paris, https://dx.doi.org/10.1787/106b1c42-en. [1]

Chapter 3. Development of the staff, leader and combined questionnaires

This chapter describes the aims, principles, individuals, timeline and key deliberations and decisions regarding the development of the staff, leader and combined questionnaires of the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018. The process commenced in early 2016 and culminated in the release of the final versions of the questionnaires in November 2017.

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

3.1. General aims, principles and timeline

As with any survey seeking to meet high scientific standards, the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 had to be theoretically sound. The process involved in meeting this requirement was not only critical to the success of the study but also meant identifying and documenting the conceptual knowledge and theoretical underpinnings of the survey's key themes, research questions and desired indicators. The conceptual framework, described in detail in Chapter 2, outlines and explains these concepts as well as the key design aspects of TALIS Starting Strong 2018.

This chapter describes the aims, principles, individuals, timeline and key deliberations and decisions with respect to the instrument development work, which commenced in early 2016 and continued until the release of the main survey instruments to the national study centres in November 2017.

The primary aims of the instrument development process were to create instruments and survey questions that aligned well with the conceptual framework, were of high technical and psychometric quality and had the potential to generate data, measures and scales that addressed TALIS Starting Strong 2018's underpinning policy questions, key themes, priorities and indicators. At the same time, because the instrument development overlapped with the instrument development for TALIS 2018, efforts were made to ensure alignment between the two surveys where appropriate to ensure some level of comparability of data (especially between ISCED level 02 and ISCED level 1) for those countries implementing both surveys.

Furthermore, the issue of item comparability between TALIS 2018 and TALIS Starting Strong 2018 was deemed important for guaranteeing a strong link between the studies, thereby enhancing their policy relevance. The items refer to the sub-questions attached to each question stem. In practice, there were instances of the same question stems being used for both surveys, but with context-specific variations in the question items.

The OECD Secretariat and the TALIS Starting Strong Consortium initially estimated that approximately 70% of the content of the TALIS Starting Strong 2018 questionnaires could be taken from TALIS 2018 with only minimal adaptations. However, after discussions with members of the survey's Questionnaire Expert Group (QEG), other experts and the participating countries, the OECD Secretariat and the TALIS Starting Strong Consortium realised that the questionnaires needed considerably stronger tailoring to the early childhood education and care (ECEC) context. The QEG accordingly developed a larger number of TALIS Starting Strong 2018-specific items or significantly adapted items from the TALIS 2018 instruments to include in the pilot questionnaires.

Table 3.1 outlines the number of TALIS Starting Strong 2018 items categorised by level of overlap with the TALIS 2018 items during the different phases of questionnaire development. The last row of the table indicates the extent of the overlap determined through a calculation that involved counting those items in TALIS Starting Strong 2018 that were the "same" as those in TALIS 2018 as well as those items that required only "minimal adaptations".

Table 3.1. Overlap between TALIS Starting Strong 2018 questionnaire items and TALIS 2018 items

Overlap type	Pilot (Oct 2016)		Field trial (Jun 2017)		Main survey (Nov 2017)	
	Leader questionnaire	Staff questionnaire	Leader questionnaire	Staff questionnaire	Leader questionnaire	Staff questionnaire
Same item	63	59	43	40	28	23
Minimal adaptation	135	147	114	108	67	57
Major adaptation	11	39	45	18	15	26
New	157	342	263	155	85	157
Total	**366**	**587**	**465**	**321**	**195**	**263**
% TALIS overlap	*54*	*35*	*46*	*34*	*49*	*30*

A particular target for the QEG was to develop international source questionnaires that would take an average of 45 minutes or fewer to complete (English language version) yet also exhibit the following characteristics: relevance, validity, consistency, comparability and clarity. Therefore, throughout each phase of the questionnaire development (pilot, field trial, main survey), the QEG worked to reduce the length of the instruments.

At the beginning of the development process, the questionnaires were drafted intentionally longer so that the validity and international comparability of the items could be tested. To achieve the desired target length of the questionnaires, the QEG reduced the questionnaire length for the field trial and then reduced it further for the main survey, considering, as they did so, stakeholder feedback, thematic relevance of each item, and statistical and psychometric properties based on the pilot and field trial data. This process meant that everyone involved in developing the TALIS Starting Strong 2018 instruments endeavoured to apply the following set of key principles to that process to the greatest extent possible:

- a theory-driven development wherein a conceptual framework guided the inclusion of questions
- maximum country input (at the policy as well as the operational level) into the development and selection of constructs, questions and items
- extensive consultation with experts in the corresponding thematic areas as well as experts in questionnaire and sample design
- use of selected constructs, variables and measures from TALIS 2018 to allow analyses of comparisons between both surveys (while allowing for improvements to questions and items taken from TALIS 2018 as deemed appropriate)
- descriptions of the cross-cultural validity of measures and procedures to establish this validity.

The remainder of this chapter documents how these principles guided the development of the survey instruments. Chapter 11 provides a detailed report of the results of this investigation into the instruments' cross-cultural validity/invariance.

As described in Chapter 1, TALIS Starting Strong 2018 was implemented in three phases – pilot, field trial and main survey. This three-phase approach was especially important for ensuring the success of the instrument development work.

The pilot study was conducted in all participating countries, albeit in an ad hoc way for one country that joined the survey after completion of the pilot survey. The purpose of the pilot was to validate the quality, clarity, appropriateness and relevance of questionnaire content. The QEG then substantially modified questionnaires in accordance with the qualitative results and feedback collected during this phase.

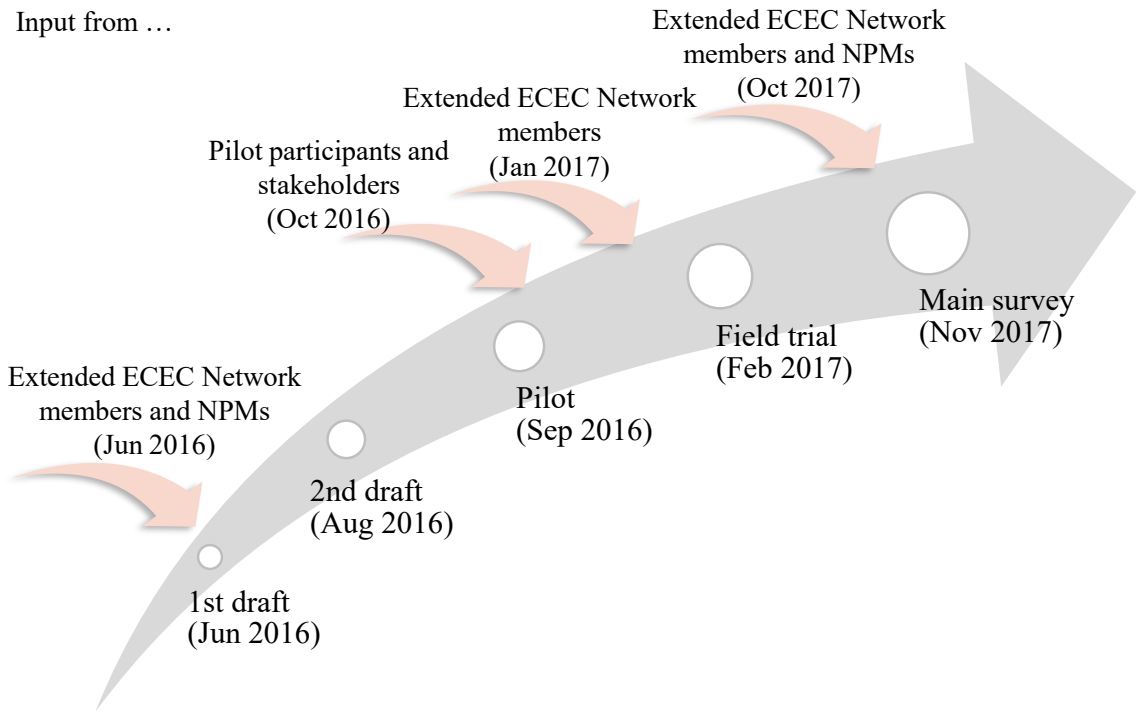
One of the main goals during the field trial stage was to collect quantitative information about the statistical and psychometric properties of the survey instruments, questions and measures. Qualitative feedback (e.g. from key stakeholders and from centre leaders and staff relayed by national study centres) was also collected at this time and used to inform revisions of the main survey instruments. These changes were usually less extensive than those made after the pilot phase.

The main survey stage focused primarily on supporting the people in the participating countries responsible for translating the international (English) versions of the instruments into local languages and contexts by helping them adapt specific terminology in the questionnaires to fit those contexts. Translators also received instrument errata and any related guidance requested during this stage.

The development of TALIS Starting Strong 2018 consisted of an iterative process of rounds of feedback provided by stakeholders and revisions made by the QEG. The following sections of this chapter provide a detailed account of the different stages of the development of the instruments.

Figure 3.1 illustrates the development stages that began with the first release of the questionnaire to the members of the OECD Extended Early Childhood and Education Network (Extended ECEC Network) in June 2016 and continued through to the release of the main survey questionnaires in November 2017. The graphic also demonstrates the various feedback rounds from key stakeholders of the project, including Extended ECEC Network members, national project managers (NPMs), participants in the pilot focus group discussions, and other external stakeholders (not included in Figure 3.1) such as union representatives, national steering committees, and (extended) QEG and Technical Advisory Group (TAG) members.

Figure 3.1. Instrument development progress



The following list includes the key instrument-development dates and periods. Additional consultation and revisions occurred at numerous points between these key dates through written exchange or virtual meetings.

3.1.1. Pilot phase

- *21 March 2015:* First Extended ECEC Network meeting. Countries discuss priority areas for the overlap between the questions and their associated items in the TALIS Starting Strong 2018 instruments and the questions and items in the TALIS 2018 instruments. Throughout their discussions, countries draw on the results of the priority rating exercise for TALIS Starting Strong 2018 and the draft teacher and principal questionnaires for TALIS 2018.
- *March 2016:* QEG (and Extended QEG) established.
- *20–21 April 2016:* First QEG meeting (Hamburg, Germany). Members discuss the conceptual underpinnings of the instruments, establish the QEG working processes and reach agreement on plans and responsibilities for instrument development.
- *17–18 May 2016:* First meeting of the joint Technical Advisory Group (TAG) of TALIS Starting Strong 2018 and TALIS 2018. Members discuss preparation of the TALIS Starting Strong 2018 technical standards, sampling and pilot study.
- *20 May 2016:* Virtual meeting of the QEG. Members discuss the first collated draft of the centre staff and the centre leader pilot questionnaires

- *6 June 2016:* TALIS Starting Strong Consortium shares the first draft of the pilot instruments with the Extended ECEC Network and requests written feedback by 23 June.
- *20 June 2016:* OECD Secretariat and the TALIS Starting Strong Consortium conduct webinar with the participating countries to discuss the draft pilot questionnaires.
- *June 2016:* Extended QEG review of instruments focuses on applicability of the instruments for middle-income countries and economies, and for ECEC staff and centre leaders working with children under three years of age.
- *13–15 June 2016:* First NPM meeting (Munich, Germany). NPMs share their feedback on the draft pilot questionnaires.
- *18–20 July 2016:* Second QEG meeting (Cambridge, United Kingdom). Further development of the pilot instruments.
- *16 September 2016:* Virtual meeting of the QEG. Members discuss and make final revisions to the pilot questionnaires.
- *20 September 2016:* External experts Pauline Slot and Trude Nilsen, with support from the QEG, discuss draft versions of the situational judgement items.
- *30 September 2016:* Pilot questionnaires (seven versions: staff questionnaire ISCED level 02 (SQ); leader questionnaire ISCED level 02 (LQ); staff questionnaire U3 (SQ U3); leader questionnaire U3 (LQ U3); combined centre questionnaire ISCED level 02 (CQ); combined centre questionnaire U3 (CQ U3); combined centre questionnaire plus ISCED level 02 (for Israel) (CQ+)) and glossary and focus group guidelines released to countries. The instruments include the five situational judgement items developed by Pauline Slot and Trude Nilsen with support from the QEG.
- *10 October to 28 October 2016:* Pilot conducted (focus group approach).

3.1.2. Field trial phase

- *7 November 2016:* QEG members receive the pilot data, thus starting the process of revising the instruments for the field trial.
- *10 November 2016:* Virtual meeting of the TAG. Members discuss the TALIS Starting Strong 2018 questionnaire development, cross-study comparability and plans for the field trial data analysis and scaling, among other issues.
- *14–16 November 2016:* Third QEG meeting (Paris, France). Members review and discuss the pilot data and feedback.
- *7 December 2016:* Webinar for experts, that is, the members of the TAG, (Extended) QEG, TALIS Starting Strong Consortium and the OECD Secretariat. The webinar participants discuss the tests of the situational judgement items in the TALIS and TALIS Starting Strong 2018 instruments.
- *15–16 December 2016:* Webinar (16 December). The TALIS Starting Strong Consortium shares and discusses the first version of the field trial instruments with Extended ECEC Network members and other stakeholders, along with a proposal to reduce the length of the instruments. Draft instruments also shared with the TAG

for their feedback, and with Extended QEG members Susanne Viernickel and Nirmala Rao, for their input into the instruments' applicability to ECEC centres for children under the age of three and to middle-income countries, respectively.

- *10 January 2017:* Participating countries and stakeholders submit their written feedback on the draft instruments to the QEG in preparation for further revisions of the instruments.
- *12 January 2017:* Webinar for countries to review and approve reduced versions of the questionnaires.
- *16 January 2017:* TALIS Starting Strong Consortium shares the second version of the field trial instruments with NPMs with the expectation that the NPMs will discuss these at the second NPM meeting to be held in Athens, Greece.
- *23–26 January 2017:* Second NPM meeting (Athens, Greece). NPMs discuss the content of the questionnaires.
- *7 February 2017:* TALIS Starting Strong Consortium releases the field trial questionnaires,[1] glossary, adaptation guidelines and adaptation forms to NPMs.
- *May to June 2017:* Field trial conducted in all countries.
- *1–2 June 2017:* Second TAG meeting (Paris, France). Field trial outcomes, non-response bias analyses plans, and quality control questions reviewed.

3.1.3. Main survey phase

- *July to August 2017:* Field trial data from all countries processed and analysed.
- *July 2017:* Short online questionnaire sent to NPMs, who are asked to gain support from national ECEC experts for analysis of the situational judgement items. The aim of this exercise is to draw conclusions that will complement information from the descriptive statistics and thereby help the QEG make meaningful decisions on how well these items worked in the field trial.
- *7–21 August 2017:* Initial and updated field trial data (without scales) shared with QEG so that they can begin analysis directed towards revising the instruments for the main survey.
- *September 2017:* Written consultation by the QEG and the TALIS Starting Strong Consortium with TAG on social desirability scales and cross-country applicability of the scales in the field trial, as well as on measures needed to produce higher response rates during implementation of the main survey.
- *1 September 2017:* Field trial data analyses (including scales and item evaluation), logs and process data completed and delivered to the QEG so as to help them revise the instruments. This material includes the results of the experiments that tested alternative versions of items.
- *13–15 September 2017:* Fourth QEG meeting (Oslo, Norway). Members discuss field trial results and preparation of the main survey instruments.
- *29 September 2017:* Virtual meeting of the TAG and QEG. Members discuss revisions to the situational judgement item questions, the target group[2] section of the questionnaires, and the technical standards.

- *6 and 9 October 2017:* TALIS Starting Strong Consortium shares draft versions of the main survey instruments with the Extended ECEC Network.
- *25–27 October 2017:* Extended ECEC Network meeting (Paris, France). Members discuss instruments with NPMs, participants from the Trade Union Advisory Committee to the OECD, UNESCO and members of the TALIS Starting Strong Consortium.
- *8–10 November 2017:* Third NPM meeting (Reykjavik, Iceland). TALIS Starting Strong Consortium and OECD Secretariat discuss main survey instruments with NPMs.
- *15 November:* QEG submits final revisions of the instruments to the TALIS Starting Strong 2018 International Study Centre (ISC) for final formatting. Revisions are based on feedback from the Extended ECEC Network meeting and the third NPM meeting.
- *30 November 2017:* TALIS Starting Strong Consortium releases main survey materials (instruments,[3] glossary, adaptation guidelines and adaptation forms) to NPMs.
- *November 2018:* Third TAG meeting (Paris, France). Members review the implications of survey-related issues for the reported results and for future cycles of the survey.

Many individual-level consultations and numerous virtual meetings were held among the members of the QEG, invited experts, country representatives and other stakeholders before, after and in between the above-stated key activities, meetings and milestones. In addition, the ISC carried out operational work relating to copy-editing, instrument layout and finalisation of guidelines and glossaries before releasing the questionnaires and accompanying materials to NPMs for translation and adaptation.

In summary, one of the main goals of the instrument development process was to engage various stakeholders at key development stages in order to increase the quality and applicability of the questionnaires. While any process faces time constraints and although the instrument development work could have benefitted from additional time for stakeholders to reflect on feedback and data, all stakeholders and the individuals contributing to this process were satisfied with the products and results.

Throughout the development process, the OECD Secretariat and the TALIS Starting Strong Consortium were committed to reassuring the Extended ECEC Network members that the themes and questions included in the questionnaires matched the policy priorities of the participating countries. The consortium also included reviews from NPMs with the aim of increasing the cultural appropriateness of the instruments. When seeking technical advice during the development stages, the OECD Secretariat and the consortium additionally invited TAG members to regularly share their feedback. The consortium furthermore considered feedback from several additional stakeholders (e.g. the Trade Union Advisory Committee to the OECD, Extended QEG members).

3.2. Questionnaire Expert Group (QEG) role, membership and collaboration

The QEG's overall task was to design and develop instruments for the two populations of interest, that is, ECEC centre leaders and staff at ISCED level 02, and ECEC centre leaders

and staff catering to children under three years of age (U3). The QEG's specific tasks included:

- reviewing TALIS questions and items from 2008, 2013 and 2018 in order to decide on their applicability for TALIS Starting Strong 2018
- proposing new indicators for consideration to meet the stated policy objectives of TALIS Starting Strong 2018, while ensuring that these were complete, coherent and formed a robust basis for analysis
- drafting suitable questions addressing the indicators and research questions
- reviewing and revising questions and items in accordance with stakeholder feedback and pilot and field trial evidence and results, especially with respect to functioning and response burden
- advising on the field trial and main survey analysis and reporting plans
- guiding the interpretation of the psychometric analysis conducted for the field trial, especially when cross-cultural validity could not be asserted.

The QEG's key responsibility was to develop proposed questionnaire content for review by the Extended ECEC Network and, by extension, the NPMs. The Extended ECEC Network's primary role was to advise on the political relevance of the questionnaire content and its adherence to the TALIS Starting Strong 2018 goals as well as the applicability of the anticipated data in both national and cross-national contexts. As an extension of this role, the Extended ECEC Network had final approval of the questionnaires used in the three major phases of validation and implementation (i.e. the pilot, field trial and main survey). The NPMs' main role was to advise on the validity of questions, foreseeable issues with respect to translation, and the clarity of the wording and sequencing of the questions for the defined populations of interest.

Chapter 2 of this document lists the QEG's core members and their roles and responsibilities.

The breadth and depth of expertise as well as the availability of contributions and feedback from the vast range of stakeholders and perspectives were instrumental to developing survey instruments that the Extended ECEC Network could endorse.

The QEG's responsibilities regarding the instruments ended after finalisation of the international master instruments. The subsequent adaptation and translation stage was completed prior to the main survey by the ISC. The analysis and reporting of the main survey data was the responsibility of the OECD Secretariat.

In terms of languages, the QEG's primary responsibility was to produce master instruments in international English (UK spelling conventions). NPMs were responsible for adaptation and translation at the national level (see Chapter 4).

3.3. Pilot phase

The main goals at this initial stage were to:

- discuss and agree on draft forms of the TALIS Starting Strong 2018 ECEC staff and leader questionnaires

- consider possible overlaps between the TALIS Starting Strong 2018 draft items and the items in the TALIS 2018 draft
- consider whether to have separate questionnaires for ISCED level 02 and for centres catering to children under three years of age (U3)
- consider whether to have separate questionnaires for small centres with only one staff member or for home-based centres
- produce proposed instruments and related recommendations for the Extended ECEC Network and subsequent collection of pilot data
- elaborate the conceptual framework document.

Compared to teachers participating in TALIS, ECEC staff exhibit considerable diversity in terms of the positions they hold and in their pre-service training. Any sample of ECEC staff is likely to include, for instance, unqualified staff working as assistants, educators who have completed vocational education and training, and kindergarten teachers holding Master's degrees. This variety was a key consideration in the development of the items in the questionnaires.

The remainder of this section outlines the development process and key considerations and directions taken before application of the pilot-phase instruments in the field.

3.3.1. Development of proposal for the questionnaires

Using the Extended ECEC Network's priorities as a starting point, the QEG reviewed and discussed the themes, constructs, questions and items in the TALIS 2018 draft principal and teacher questionnaires that could potentially be used in the TALIS Starting Strong 2018 instruments. The QEG used the information generated by this review to discuss the framework development goals and the input required from QEG members.

During this time, the QEG used the ISCED level 02 questionnaire as the starting point while also paying close attention to the additional TALIS Starting Strong 2018 population of interest (U3) and the corresponding sampling implications. QEG members also considered issues and constraints with respect to data consistency and reliability across countries and the two populations of interest, and in regard to analysis of and comparisons with the upcoming TALIS 2018 results. The criteria for reviewing the TALIS 2018 questions and items included, among others, the Extended ECEC Network's priorities, the suitability and/or applicability of the items for the ECEC context, use of data during reporting of the survey's results, and relevance of the questions and items with respect to enabling comparisons of the data from the two surveys.

The QEG then designed and drafted, on the basis of the Extended ECEC Network's broad specifications, the ECEC staff and leader questionnaires for both populations of interest, ISCED level 02 and U3.

The QEG based their subsequent work on the proposed core TALIS 2018 ISCED level 2 instruments and on the extensive review of the first draft of the instruments. The review incorporated feedback from fellow QEG members and other invited experts, the Extended ECEC Network, NPMs (during and after their first meeting), steering groups in the participating countries, the OECD Secretariat, the TAG, and other stakeholders, especially the Trade Union Advisory Committee to the OECD.

The above reviews informed the drafting and rewording of the second draft of the pilot instruments as well as the questions, items and response categories to include in them. The

guiding priority for the QEG was to shorten the instruments significantly and offer advice on improving the instruments' capacity to collect the required information. The review process included all questions and items and focused on a number of aspects, such as the rationale for and value of inclusion, reading load for respondents, repetitive and long wording, inconsistent or unclear terminology, and so on.

Two specific examples and areas of discussion illustrate the development process:

- The QEG, the TAG and the OECD Secretariat discussed concerns about possible response bias occurring when ECEC staff were asked to self-report on the pedagogical practices they were using in their respective centres. Because TALIS 2018 was testing a new item format called "situational judgement items" at this time, the three stakeholders agreed to develop and test similar items in the pilot and field trial for TALIS Starting Strong 2018, and to maintain a strong focus on process quality. The OECD Secretariat commissioned an external expert involved in developing these items for TALIS 2018 and an expert on process quality in ECEC to work alongside the QEG to draft items adapted to the ECEC context.
- The QEG considered two options for the practices section of the ECEC staff questionnaire. The first was to use a "target group" approach to contextualising the questions. The second option, a generic one, did not refer to a particular target group. Because of TALIS 2018's use of the "target class" approach in previous and current cycles and because of the Extended ECEC Network's interest in practices and the prevailing opinion that context influences these practices, the QEG decided to pursue the target-group option for the pilot.

3.3.2. Evaluating the suitability for U3

The QEG considered the analytical consequences of using (1) universal instruments across both populations of interest (namely ISCED level 02 and U3) or (2) instruments focused on the levels offered by a centre and/or taught by the individual ECEC staff. While ECEC centres in many countries were allocated to only one level in the main survey, some or all ECEC centres in other participating countries were offering provision at both the ISCED 02 and U3 levels.

TALIS Starting Strong's interest lay in comparing characteristics between and across these two levels. The recommended solution was to randomly assign a centre in which more than one level was being offered to one of the populations of interest during the sample selection and then to request all participating staff within the selected centre to respond in accordance with the level to which their centre had been assigned (see also Chapter 5 and Chapter 6).

The QEG also agreed that producing instruments that were otherwise "universal" across the two populations of interest was possible because this process would require only minor adaptations to the wording and reference context for a limited number of questions and items. This approach meant a more inclusive database could be built – one that would enable broader analysis and comparisons of data across the populations.

The QEG invited an expert in the provision of ECEC at ISCED level 01 to review the applicability of the instruments to the U3 setting in the early and later drafts of the pilot questionnaires. This person reviewed the instruments question by question and recommended, after also taking country feedback into account, some minor modifications (e.g. rewording or deleting or adding a small number of items).

The U3 version of the staff questionnaire included five unique items, while the ISCED level 02 version contained 22 unique items. In both versions of the staff questionnaire, one of the unique items had wording specific to either U3 or ISCED level 02. The leader questionnaire contained one item unique to the U3 version, and 18 items unique to ISCED level 02.

3.3.3. Development of the combined centre questionnaires

Early in the development process, NPMs raised concerns about the applicability of the ECEC staff and leader questionnaires in centres with just one person or in centres with only one main teacher and assisting staff. Within the ECEC context, some sampled ECEC centres may consist of only one person who cares for a small number of children. Representatives from countries where this type of centre is found considered the leader and staff questionnaires developed for TALIS Starting Strong 2018 to be unsuitable for such centres. As a result, a combined centre questionnaire was developed that contained (combined) suitable questions from the staff questionnaire and selected questions from the leader questionnaire. This combined version featured the staff questionnaire as a base, with non-applicable items (e.g. regarding perceptions of the centre leadership) removed and questions from the leader questionnaire added where deemed appropriate (these questions typically concerned centre characteristics).

A longer version of this combined questionnaire was also drafted to accommodate the specific case of the Israel ISCED level 02 population, where centres consist of a main ECEC staff member who acts as both lead pedagogical staff and leader but is supported by usually one or two additional teacher assistants who help with some tasks. Although the main ECEC staff fulfil the TALIS Starting Strong 2018 centre leader definition, they spend a large proportion of their time working directly with children. If these main staff had completed only the leader questionnaire, the responses to the staff questionnaire would have been collected solely from teaching assistants, an occurrence that would not have represented the reality in these centres.

In response, the QEG re-introduced additional questions about collaboration among staff and about centre leadership to the combined questionnaire that would be administered in the Israeli level 02 centres. In summary, the majority of the questions in this questionnaire came from the staff questionnaire, while the questions most relevant to leadership in these centres came from the leader questionnaire. These versions of the questionnaires were by default longer than the individual staff and leader questionnaires, but they retained the question wording and ordering to ensure that analysts could compare the resultant data with data from other centres.

3.3.4. Focus group operations

Because 2018 marked the first implementation of TALIS Starting Strong and because the ECEC context is known to be diverse both across and within countries, the TALIS Starting Strong Consortium required all participating countries to conduct the pilot study. This approach ensured that both educational levels (ISCED level 02 and U3) and all regions and language contexts across the participating countries were represented in the pilot.

The terms of reference for the survey requested a qualitative approach to the pilot, in line with the approach taken in TALIS 2013 and TALIS 2018. The key reason for this request was that the OECD Secretariat considered focused, structured group discussions to be the most effective way of obtaining meaningful feedback and suggestions from ECEC centre staff and leaders. This proposed approach required one or more groups of ECEC staff and

leaders in ISCED level 02 centres and in centres for children under the age of three to complete the questionnaires and then discuss them in a structured session.

The aim of the pilot was to collect information on a number of instrument-related aspects. These included applicability of concepts, level of complexity, overall organisation of topics and questions, applicability of questions and items across the ISCED level 02 and U3 centres, wording and definition of terms, appropriateness and cultural relevance, mandatory national adaptations, foreseen adaptation and translation issues, flow of questions (overall and specifically with respect to "skip" instructions), overall length and overall "feel" of the questions and questionnaires.

To facilitate this process, the QEG members provided the TALIS Starting Strong Consortium with key "probing" questions designed to evaluate the quality of the respective sections of instrument content. The consortium incorporated these key questions into detailed annotated questionnaires that group discussion moderators could use during their sessions. The consortium also developed easy to implement guidelines for the pilot implementation.

All participating countries agreed to participate in the pilot. However, the country that joined the study later than the other countries implemented its focus group in an ad hoc manner on a slightly delayed schedule. During this stage of the survey, national teams in the participating countries collected rich feedback and comments from staff and centre leaders at ISCED level 02 and, where applicable, level U3.

3.4. Field trial phase

The main tasks of this phase of instrument development were based on the rich set of feedback and comments collected from respondents during the pilot phase. These tasks included revising and finalising the instrument design and content for the field trial, advising on possibilities for linking the TALIS 2018 and TALIS Starting Strong 2018 data, planning for translation and verification procedures, and analysing the field trial data. Time was also put aside for reflecting on issues related to socially desirable responding and other response style issues, and on possible ways of addressing these. The QEG also discussed the situational judgement items.

The third in-person QEG meeting focused on discussing the pilot data and accompanying detailed written feedback from the wider stakeholder groups, including from the Extended ECEC Network. The main goals of this meeting were to propose deletions, consider priorities for TALIS 2018 overlap questions, propose improvements to questions and propose additional questions if required. Further feedback and virtual meetings took place after this meeting within the QEG and with other experts and stakeholders until the final field trial questionnaires were released in February 2017.

Throughout the process, QEG members individually reviewed (according to their expertise) the questionnaire sections and glossary items and submitted their proposed revisions to the group. QEG members also followed the general guidelines outlined below when facing decisions about revising questions and items in the questionnaires or deleting them from or adding them to the questionnaires:

- pilot or field trial data to supersede stakeholder written feedback
- consistency in change requests across stakeholders to receive more weight than change requests from single countries or a relatively lower number of countries

(however, the possibility for national adaptations could be considered in some cases)

- consideration to be given to the analytical potential of items (e.g. if the QEG expected that a group of items would form a scale, they tried not to delete those items)
- consideration to be given to technical aspects (consistency across items, item length and complexity, response formats) to help reduce the reading burden on respondents
- OECD priorities for minimal adaptation from TALIS 2018 to be seen as important so as to ensure comparability of results with TALIS 2018 data where possible
- tension between the length of questionnaires and stated priorities to be considered and discussed with the OECD
- QEG content expertise and past experience with other international surveys such as TALIS to be taken into account.

The QEG members jointly discussed the changes indicated in the submitted reviews. During this work, general methodological issues emerged that required QEG consideration. These issues included instructions related to the "target group" section in the staff questionnaires, a possible instrument modularisation/rotation and the inclusion of measures of pro-social responding. The following sections provide more details about these issues, along with a description of the key areas of content revision.

3.4.1. Rotated field trial questionnaires

As indicated in the first section of this chapter, the length of the questionnaires was gradually reduced over the survey phases, with the field trial questionnaires being shorter than the pilot questionnaires and the main survey questionnaires being shorter than the field trial questionnaires. In order to test more questions and to experiment with alternative question formats, the ISC suggested that a rotated questionnaire design should be implemented for the field trial version of the staff questionnaire. To enable this approach, the ISC developed three forms of the staff questionnaire with expected response times of about 60 minutes each. The QEG and the TAG approved the final versions of these three forms.

At the point of administering the survey, some of the questions overlapped across the three forms, while other sets of questions were only included in two of the three forms. If the three forms had been administered as part of one questionnaire, this would have resulted in a theoretical total response time of 75 minutes. Table 3.2 illustrates the applied rotated questionnaire design. As is evident, three sections in the staff questionnaire were included in all three forms of the questionnaire for analytical purposes ("Your background information", "Your professional practices" and "Personal attitudes"). All other questionnaire sections were included in two of the staff questionnaire forms.

Table 3.2. Rotated field trial staff questionnaire design

Questionnaire section	Form A	Form B	Form C
Your background information		common	
Your professional development (part 1)	x	x	
Your professional development (part 2)	x		x
Your work week		x[a]	x[b]
Your professional beliefs	x	x	
Your professional practices	x[b]	x[b]	x[a]
Practices used in the <ECEC centre>	x[a]		x[b]
Your work with the <target group>		x[a]	x[b]
Your work climate and job satisfaction (part 1)		x	x
Your work climate and job satisfaction (part 2)	x	x	
Personal attitudes		common	
Total estimated time	*56 min*	*58 min*	*56 min*

Note: [a] and [b] indicate the inclusion of a section in a specific form, but with experimental questions included, meaning the format of questions asked in one of these forms differed from the format used in the other form (see Section 3.4.2).

No rotation was implemented for the leader questionnaire or the combined centre questionnaire. However, because the combined centre questionnaire combined questions from the two core questionnaires, it was longer than the staff questionnaire. To reduce the length of this questionnaire, the QEG based the field trial questionnaire on the shortened form of the staff questionnaire, while other, analytically less relevant questions were excluded from administration of the field trial survey.

3.4.2. Embedding experiments and their subsequent evaluation

As suggested by the TALIS Starting Strong Consortium and also recommended by the TAG, the field trial instruments included built-in experiments designed to address context effects so that the comparability of items across surveys could be empirically addressed. The field trial questionnaires consequently included multiple versions of some questions that could be trialled with different groups of respondents (see Table 3.3).

Table 3.3. Summary of main study questions that were tested during the field trial through use of multiple versions

	Staff questionnaire
SQ20	In the field trial, one version of this question asked for number of hours, while the other provided response categories with a range of hours.
SQ25 and SQ26	In the field trial, two versions of each situational judgement item question were tested, one with four response categories, thus allowing for each response to be rated in terms of likelihood, and one asking the respondent to choose their first choice from among the categories. (Note: only two of the three situational judgement item questions used in the field trial were included in the main study.)
SQ29 to SQ33	For each process quality question, two versions were tested: "To what extent do <ECEC staff> at this <ECEC> centre encourage children to do the following activities" and "To what extent do the following activities take place in this <ECEC centre>". This approach helped determine whether these items could be used in small centres. (Note: only four of the six process quality questions in the field trial were later included in the main study.)
SQ38	One version of this question used response categories for estimating proportions of children in the target group and the other version asked respondents to specify the number of children.
---	One version of a question on time spent asked for number of hours, while the other provided response categories with a range of hours. (Note: this question was not included in the main study.)

Note: The question numbers used in this table are from the *final* main study versions of the questionnaires.

3.4.3. Inclusion of a scale measuring social desirability

"Social desirability" (sometimes called "pro-social" responding) is a response style assumed to function as a filter during respondents' interaction with questionnaires.[4] Because the TAG expressed concerns about the validity, reliability and comparability of the TALIS Starting Strong 2018 data, the QEG discussed with them the possibility of including a social desirability scale similar to the one used in the TALIS 2013 field trial to measure the extent to which respondents were influenced by a social desirability bias in their responses to the survey. The scale used in TALIS 2013 was based on the original Marlowe-Crowne Social Desirability Scale (Crowne, D. P. and D. Marlowe, 1960[1]) and it provided data that could be used to help describe or even adjust for responses affected by socially desirable styling. TALIS 2018 had likewise assessed response styles, such as acquiescent, midpoint, and extreme responding (e.g. the proportion of extreme responses across all Likert items).

The QEG and the Extended ECEC Network agreed that use of the social desirability scale in TALIS Starting Strong 2018 should be limited to the field trial, as the data from this phase would not be made publicly available. However, it would still be possible to use the data to estimate the impact of social desirability at the country level and thereby inform the analysis. The QEG asked the TAG to help them adapt the items from TALIS 2013 for the ECEC context. This work led to the inclusion of a separate, dedicated set of (brief) items in the TALIS Starting Strong 2018 staff questionnaire and thus an implicit way of gauging other response styles such as extremity and acquiescence. The measure focused on both positive and negative impression management.[5]

3.5. Main survey phase

3.5.1. Analysing the field trial response and process data

After the field trial, and consistent with and following the approaches used in all TALIS cycles, the ISC implemented a full quantitative application of all instruments in all applicable contexts, populations and languages. In addition to testing the survey operations in preparation for the main survey, the field trial provided valuable data that enabled the QEG to examine the functioning of the items and scales in the questionnaires. Also, as suggested by the TALIS Starting Strong Consortium and agreed to by the TAG, the field trial included built-in experiments that made it possible to test different formats and wordings of questions to determine which ones had the best psychometric properties.

The TALIS Starting Strong Consortium shared the field trial data analyses, along with written feedback on the field trial questionnaires received from country stakeholders, with the QEG so that they could use this information to inform the development of the main survey questionnaires. The consortium asked the QEG to review the feedback and data analyses so that they could prepare proposals on what to revise and what to delete in the questionnaires before the final in-person QEG meeting between 13 and 15 September 2017. Throughout the questionnaire revision process, the QEG members adhered to the following general guidelines when making decisions about revising, adding or deleting questions and items:

- *Field trial data to supersede stakeholder written feedback:* The data analysis included response time, non-response, distribution of responses across response categories, summary statistics of continuous variables and scale analyses.

- *Consistency in change requests across stakeholders to receive more weight than change requests from a single country or just a few countries:* The QEG could, however, consider possible national adaptations in some cases.
- *The analytical potential of items to be considered:* If, for example, a group of items formed a scale, the QEG needed to try to avoid deleting those items.
- *Technical considerations to be taken into account to help reduce the reading burden of respondents:* These considerations included consistencies across items, item length and complexity, and response formats.
- *Consideration to be given to OECD priorities to minimally adapt relevant TALIS 2018 questions and items:* This matter was deemed important in terms of ensuring that the results from TALIS Starting Strong 2018 could be compared with the TALIS 2018 data where possible.
- *Tension between the length of questionnaires and stated priorities to be taken into account*, with the QEG to discuss any such tensions with the OECD Secretariat.
- *QEG members to draw on their own content expertise and past experience with other international surveys*, such as TALIS.

The TALIS Starting Strong Consortium documented the rationales for changes made (or not made) to the questionnaires during these phases so that they would be available during future consultation, if needed.

Finally, as recommended by the TAG, the TALIS Starting Strong Consortium assessed response styles in the field trial questionnaires. TALIS 2013 had found sizable cross-cultural differences in social desirability, but the 2013 survey also found that score corrections did not affect the size or patterning of the cross-national differences. The TAG's analysis of the TALIS Starting Strong 2018 field trial data also found large cross-cultural differences, but the correlation between the scores corrected and not corrected for social desirability remained high. The TAG consequently concurred with the QEG's recommendation to exclude the social desirability items from the main survey, given that scoring of extremity and modesty responding (response styles) would probably provide an adequate approximation of social desirability.

3.5.2. Key areas of instrument revision and further development

The QEG based their questionnaire development work on the substantial data and feedback received and on corresponding advice on making the wording of question stems, items, response categories, instructions and explanatory notes clear and consistent. More details on the conceptual and theoretical underpinnings of each theme can be found in the *Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework* (Sim et al., 2019[2]).

Given the volume of exchange and communication between QEG members individually and as a group (with these activities often supported through virtual meetings), detailing the genesis of specific instrument-development aspects is not possible. However, the section that follows provides insight into the development of a number of key themes and areas as described in the conceptual framework.

The review process paid attention to, among other matters, aligning terms used in TALIS Starting Strong 2018 with terms used in the OECD's *TALIS*, *PISA* and *Education at a*

Glance publications and in other UNESCO, OECD, IEA and European Commission publications and manuals.

When possible, questions were combined in order to decrease the reading and response burden on respondents. However, explanatory notes were added when necessary to facilitate consistency in comprehension and responses.

Themes mainly concerned with ECEC staff beliefs and staff–child interactions

Process quality of staff–child interactions

The questions addressing this theme were central to TALIS Starting Strong 2018. As such, the QEG allocated a considerable amount of development work to them. The QEG and NPMs expected indicators related to the process quality of staff–child interactions to be not only key dependent variables but also a highly likely outcome of the analyses of the majority of constructs measured in the survey. Because process quality is most commonly measured by trained raters observing interactions in the classroom/playgroup, efforts to adapt these measures to a survey platform formed the basis of this theme's development.

The QEG first created a very long list of items representing different aspects of process quality in ECEC centres. The list was based on an existing observational tool, the Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale (Siraj, I., Kingston, D. and E. Melhuish, 2015[3]). Multiple rounds of feedback and pilot and field trial data collection resulted in the retention of a reduced number of items. However, that number was deemed sufficient to ensure an adequate coverage of relevant aspects of process quality in the staff questionnaire. The QEG also revised and retained two questions framed as situational judgement items as a result of the feedback and data analyses.

Approaches to pedagogical practices

Process quality is linked to pedagogical practices. A distinction is often drawn between child-centred instruction (in which activities are child-initiated and children engage in problem solving and inquiry-oriented learning) and didactic instruction (staff-directed, with planned tasks focusing on acquiring and practising academic skills). Some approaches to ECEC pedagogy stress the importance of staff-directed transmission of skills related to the curriculum. This emphasis results in a didactic approach even with very young children, where teachers use direct instruction and rewards to reinforce learning processes with the aim of preparing children for primary school. The current consensus view can be characterised as social-constructivist. While this view stresses the importance of children's intrinsically motivated activity and initiative as the driver of development, it acknowledges that development does not take place in a cultural void.

The QEG included seven questions on pedagogical practices in the "Your work with the target group" section of the staff questionnaire. They also introduced the concept of "target group" to elicit understanding of staff practices in relation to a group of children they had recently worked with (as defined by the questionnaire, the first group of children whom staff had worked with on their last day before completing the survey). However, countries highlighted issues with this concept because some education systems exhibit significant variation in the extent to which staff engage in group-related work from one day to the next. To address this issue, the QEG included a question after the field trial designed to provide insight into whether the target group that staff used as a reference when answering these questions was representative or not.

Furthermore, between the version of the questionnaire used in the field trial and the one used for the main study, the QEG endeavoured to ensure that process quality and structural quality indicators in the target-group-section would be designed in a way that made it possible to explore associations between them.

Monitoring and assessment of children's development, well-being and learning

Monitoring and assessment refer to how early childhood professionals gain understanding of children's development and learning. Monitoring and assessing children's development, well-being and learning outcomes therefore help ECEC staff identify children's needs and thus support their development. The QEG developed several questions for inclusion in the survey with the aim of aiding examination of the within- and cross-country heterogeneity of ECEC staff's monitoring and assessment practices. However, due to questionnaire length constraints, the items examining this theme were reduced somewhat in line with analysis of the pilot and field trial data and stakeholder feedback.

Themes mainly concerned with the ECEC centre context

Structural quality characteristics

Structural quality characteristics include the composition of the child group, group/class size, and staff–child ratio, in addition to staff qualifications. The QEG developed centre background questions to allow exploration of such indicators as centre enrolment, number of staff, and staff shortages and attrition across the participating countries, as well as examination of the associations between centre and staff size and reported pedagogical practices and process quality indicators.

Pedagogical and administrative leadership

The questions on centre leadership included in both the staff and leader questionnaires provided some overlap with the questions included in TALIS 2018 and featured a focus on distributed leadership. However, the number of questions exploring pedagogical and administrative leadership was reduced as a result of the pilot and field trial data and feedback. The retained items included aspects of barriers to and facilitators of effective leadership, appraisal and feedback and centre evaluation.

Climate

Centre climate comprises both playroom/classroom climate and organisational climate, both of which can be captured through the process and structural elements of ECEC quality. Working conditions refer to a set of structural characteristics that can influence the extent to which ECEC staff are motivated by and satisfied with their chosen profession. The TALIS Starting Strong 2018 questionnaires included items designed to measure both these aspects and included indicators such as staff engagement, shared culture, climate for staff learning, working hours, and sources of work stress. The field trial phase focused on reducing the length of these questions and response options, as well as on increasing clarity and coherence across the items and between the staff and leader questionnaires.

Stakeholder relations

ECEC learning and well-being environments do not operate in isolation, but instead often reflect work with various stakeholders to enhance children's development, well-being and learning. Both the staff and leader questionnaires included questions exploring the extent

and importance of stakeholder relations in ECEC centres. However, this theme was another one that experienced important reductions in length (especially in the staff questionnaire) after scrutiny of the pilot and field trial data. Nevertheless, questions on parental engagement, relationships and outreach to other stakeholders, including primary schools for transition purposes, were retained.

Themes mainly concerned with leader and staff backgrounds

Background and initial preparation

In addition to collecting data on pre-service education and training, TALIS Starting Strong 2018 followed TALIS practice by collecting key elements about ECEC staff backgrounds. The relevant questions asked ECEC staff and leaders about their personal characteristics (e.g. gender, age, employment status, work experience). These basic characteristics are ones likely to be of interest in terms of their relationship to other indicators, and they may also be of value in providing descriptive information about ECEC centres and systems. Although many TALIS 2018 items were adapted to the ECEC context, some effort was made to retain as much overlap between the two surveys as possible during the TALIS Starting Strong 2018 questionnaire development phase.

Professional development

Both the staff and leader questionnaires included questions on the various activities that promote ECEC staff skills and knowledge and that advance staff's ability to work effectively with children once they secure employment in ECEC centres. Given the high priority given to this theme, the QEG retained and improved a large number of questions after the revision phases of the development work. These questions included indicators of participation in induction and professional development activities, incentives and barriers to participation and perceived needs for further training.

Well-being

During TALIS 2018, stakeholders identified well-being as an important theme for TALIS Starting Strong 2018. As in previous cycles of TALIS, well-being was seen as a construct that includes elements of job satisfaction and perceptions of the value of the profession, as well as career aspirations and work stress. Because many of these indicators were newly developed for TALIS Starting Strong 2018, the QEG initially drafted a large number of questions for testing and then refined them and reduced their number during the field trial phase in order to create clear and coherent indicators for the main survey.

Professional beliefs about children's development, well-being and learning

The focus for this section of the questionnaires was to reduce their length and ensure alignment between beliefs about children's development, well-being and learning and questions on the process quality of staff–child interactions. The particular aim here was to allow for analytical links between these sections. Ultimately, the section contained questions targeting respondents' views not only on the importance of developing specific abilities and skills in children to prepare them for the future but also on priorities for government spending on ECEC. Further refinement of the questions and their response options took place in accordance with the review of the data and feedback.

Self-efficacy

With the concept of self-efficacy having been measured in the TALIS cycles, the QEG worked to adapt these items for TALIS Starting Strong 2018. In accordance with this previous work, ECEC leader and staff self-efficacy was conceptualised as beliefs regarding capability to enact certain behaviours that may influence children's outcomes, such as achievement, interest and motivation. The QEG made substantial adaptations to the original questions used in TALIS 2018 to provide a more relevant measure of staff self-efficacy in the ECEC context. Because TALIS Starting Strong 2018 used TALIS 2018 as a starting point for these questions, a similar question format and intent to measure how well staff felt they could accomplish key tasks remained. However, the questions' individual items were tailored to the ECEC context. After the pilot finished, the QEG reduced the number of items. The final indicators and dimensions concerned with self-efficacy included self-efficacy relating to the following: equity and diversity practices, the process quality of staff–child interactions, monitoring and assessment of children, and shortage of resources.

Themes intersecting with other themes

Equity and diversity

TALIS Starting Strong 2018 addressed two aspects of equity and diversity among children: socio-economic and cultural. The QEG developed several questions to allow exploration of approaches to diversity across countries and examination of the relationship between reported approaches and pedagogical practices and the composition of children in the centres in terms of their socio-economic and cultural backgrounds. Development of these items adopted the dual emphasis on socio-economic equity and multiculturalism taken by TALIS 2018 by adapting some questions from that study to ECEC settings and adding new questions based on the Early Childhood Rating Scale-Revised (ECERS-E) (Mathers, S. et al., 2013[4]). The indicators included composition of the children in the centre and approaches to diversity and self-efficacy in centres' equity and diversity practices.

Table 3.4 presents the themes that were partially or fully covered in each version of the questionnaires. Each theme containing all available items for that theme out of the pool of all available items for the questionnaire type (staff or leader) is marked as fully covered. Themes listed as partly covered typically contain many items relevant to the theme but not all of them because as soon as one item was not included in the questionnaire during the development work, the theme was listed as partly covered. For example, some items across the different themes in the ISCED level 02 questionnaire were not included in the U3 staff questionnaire because they were not applicable for that age group.

Table 3.4. Themes partly (p) or fully (f) covered in each version of the main survey questionnaires

Themes	ISCED level 02				U3		
	Staff questionnaire	Leader questionnaire	Combined centre questionnaire	Combined centre questionnaire plus	Staff questionnaire	Leader questionnaire	Combined centre questionnaire
1. Process quality of staff–child interactions	f	f	p	f	p	p	p
2. Monitoring children's D, WB, & L	f	p	p	f	f	p	p
3. Structural quality characteristics	f	f	p	f	p	f	p
4. Pedagogical and administrative leadership	f	f	p	p	p	f	p
5. Climate	f	f	p	f	f	f	p
6. Stakeholder relations	f	f	p	f	f	p	p
7. Background and initial preparation	f	f	p	p	p	f	p
8. Professional development	f	f	p	p	p	p	p
9. Well-being	f	f	p	p	p	p	f
10. Professional beliefs about D, WB, & L	f	p	f	f	p	p	f
11. Self-efficacy	f	f	p	f	p	f	p
12. Equity and diversity in the child group	f	f	p	p	p	p	p

Notes: D, WB, & L = development, well-being and learning, respectively. "Fully" means that the questionnaire version contained all available items relevant to the theme during the main study; "partly" means that the questionnaire version contained some but not all available items relevant to the theme during the main study.

3.5.3. Revising and choosing content for the main survey in collaboration with the Extended ECEC Network

During the in-person QEG meeting held from 13 to 15 September 2017, which was also attended by representatives of the OECD Secretariat, the TALIS 2018 International Consortium, the Trade Union Advisory Committee to the OECD, and the TAG, the QEG members discussed the proposals for revisions and deletions made by each QEG member in relation to the questions they were responsible for. Based on the data available at the time of the meeting (seven out of nine countries for ISCED level 02, three out of four for U3), the average length of time needed to answer the field trial instruments was 55 minutes for the staff questionnaire and 77 minutes for the leader questionnaire. However, the time taken varied considerably across the countries.

The target length of the international version of the main survey questionnaires had been earlier set at 45 minutes each. Consequently, during the meeting, the QEG discussed and prepared draft main survey questionnaires that met this criterion. The process involved in selecting the questions, items and scales to be deleted encompassed three stages, outlined in order below. The group's discussions were also informed by the draft TALIS Starting Strong 2018 reporting plan. During each of the following stages, the QEG also took into consideration the priorities identified by the participating countries and the OECD Secretariat.

1. *Quality of the field trial data:* The QEG examined the properties of the scales and items, including the fit of the data to the theoretically proposed scales, and the

variation in items. A proposed scale on support for parent/guardian engagement and participation in ECEC centres, for example, was found to have poor fit with no clear factor structure. The QEG therefore agreed not to include this scale in the analysis of the main survey data but to retain the items on parent engagement in the questionnaire.

2. *Conceptual reasons:* QEG members made sure they retained items, scales and questions that added face validity to the questionnaires (e.g. ensuring that each dimension of process quality was represented) and were not present in other parts of the questionnaires.
3. *Questionnaire length:* Finally, the QEG eliminated low-priority questions and items that took considerable response time in each section.

The final versions of the main survey staff and leader questionnaires that resulted from this process each had an estimated response time of approximately 45 minutes.

The OECD Secretariat organised an additional webinar on 29 September 2017 to discuss the questions pertaining to the "target group" of children in the centres as well as the situational judgement items. Webinar participants also included the QEG, TAG, OECD Secretariat and the TALIS Starting Strong Consortium. The webinar provided important input into the revisions made to the questions relating to the target group. The webinar also provided important input into revision of the situational judgement items included in the TALIS Starting Strong 2018 main survey questionnaires.

The QEG and the TALIS Starting Strong Consortium completed and shared the first draft of the main survey questionnaires with participating countries on 6 October 2017 in preparation for webinar discussions between the countries and the Extended ECEC Network on 11 October 2017. During the webinar, countries were informed of any changes that were made to the main survey questionnaires as well as any other developments related to them since the field trial. The QEG members asked the participating countries and stakeholders to provide additional written feedback on these draft main survey instruments by 18 October 2017 so they could prepare the second draft of the main survey questionnaires. The QEG then invited the Extended ECEC Network to approve the main survey questionnaires during the sixth Extended ECEC Network meeting on TALIS Starting Strong 2018, held from 25 to 27 October 2017.

Table 3.5 provides an overview of the adaptations made at each phase of questionnaire development. It summarises the number of questions that were deleted, added or modified during each phase or that were left the same as in the previous questionnaire version.

Table 3.5. Overview of the number of questions modified at each questionnaire development phase

Question compared to previous version	Pilot		Field trial		Main survey	
	Leader questionnaire	Staff questionnaire	Leader questionnaire	Staff questionnaire	Leader questionnaire	Staff questionnaire
Same	2	3	6	11	8	6
Modified	38	46	36	45	29	39
Added	15	27	3	5	0	1
Deleted	10	26	10	16	8	16
No. of questions in questionnaire	***55***	***77***	***45***	***61***	***37***	***46***

Note: The starting points for the pilot staff and leader questionnaires were the TALIS 2018 pilot teacher questionnaire and the TALIS 2018 pilot principal questionnaire, respectively.

As explained in Section 3.3, TALIS Starting Strong 2018 developed a number of instruments so that the diverse ECEC contexts across the participating countries could be addressed. Table 3.6 presents an overview of the questionnaires used during the main survey data collection in the participating countries.

Table 3.6. Overview of TALIS Starting Strong 2018 instruments administered during the main survey

Participating country	ISCED level 02			U3		
	Leader questionnaire	Staff questionnaire	Combined centre questionnaire *or* Combined centre questionnaire plus	Leader questionnaire	Staff questionnaire	Combined centre questionnaire
Chile	X	X				
Denmark	X	X		X	X	
Germany	X	X		X	X	X
Israel		X	X+	X	X	
Iceland	X	X				
Japan	X	X				
Korea	X	X	X*			
Norway	X	X		X	X	
Turkey	X	X				

Notes:
+ Israel used the combined centre questionnaire plus.
* Korea used the combined centre questionnaire in case a single-person setting was sampled.

As described in Section 3.4.2, the field trial questionnaires included built-in experiments designed to address context effects, thereby making it possible to empirically address the comparability of items across surveys. The data on these experiments obtained from the field trial formed the basis of the decision regarding which options to retain, if any, in the main survey questionnaires.

In particular, the data distribution for the items in the staff questionnaire about work-hour distributions across tasks indicated that asking respondents to give the number of hours they spent on each specified activity was problematic and led to unreliable data. The QEG therefore decided to depart from the approach taken in TALIS 2018 and to use response categories containing ranges of hours. QEG members took a similar approach to the

question about proportions of children in the target group by including response categories with ranges instead of the response option asking for exact numbers.

On finding that the data from the situational judgement items suggested asking respondents to select their first and second choices was more difficult to implement than asking them to rate each item, the QEG chose this second option for the final version of the questionnaires.

The experiments with regard to the wording of the process-quality questions indicated that, in most cases, the original wording ("To what extent do the following activities take place in this <ECEC centre>") performed better, which led to the decision to use this wording consistently across all relevant questions in the staff questionnaire section "Practices used in this <ECEC centre>". The QEG deleted the question related to the last experiment (i.e. concerning staff time) from the questionnaires in order to shorten them.

3.5.4. Co-ordinated national options

Mindful of the target length of the main survey questionnaires, yet recognising the individual priorities of participating countries, the OECD Secretariat agreed that the TALIS Starting Strong 2018 questionnaires could include a number of co-ordinated national options (CNOs). Table 3.7 provides an overview of the CNOs included in the questionnaires.

Table 3.7. Overview of co-ordinated national options

Leader questionnaire	Staff questionnaire
LQ-07: "Please indicate at how many <ECEC centres> you currently work as an <ECEC centre leader>."	SQ-03: "In what country were you born?'
LQ-16: "Is this <ECEC centre> a for-profit or a non-profit organisation?"	SQ-12P: (Stem: Were the following elements included in your formal <education or training> programme?) "Working with <dual/second language learners>"
LQ-34J: (Stem: To what extent is this <ECEC centre's> capacity to provide a quality environment for development, well-being and learning currently hindered by any of the following issues?) "Shortage of <ECEC staff> with competence in working with children speaking another language than the language(s) used in the <ECEC centre>"	SP-14P: (Stem: Were any of the topics listed below included in your professional development activities during the last 12 months?) "Working with <dual/second language learners>"
	SP-16P: (Stem: For each of the areas listed below, please indicate the extent to which you currently need professional development.) "Working with <dual/second language learners>"
	SQ-27: "Do you work with children who are <dual/second language learners>?"
	SQ-28: "When interacting with children who are <dual/ second language learners>, how often do you engage in the following activities?"

As the name CNOs suggests, countries could choose whether or not to include these in their national questionnaires. The CNOs underwent the same verification processes and international data processing as the international questions, a process that ensured the data collected through these options would be comparable at the international level for those countries that decided to include them.

Notes

1. The following versions were released: rotated versions (Forms A, B, C) of the staff questionnaire ISCED level 02 (SQ); leader questionnaire ISCED level 02 (LQ); rotated versions (Forms A, B, C) of the staff questionnaire U3 (SQ U3); leader questionnaire U3 (LQ U3); combined centre questionnaire ISCED level 02 (CQ); combined centre questionnaire U3 (CQ U3); rotated versions (Forms A, B, C) of the combined centre questionnaire plus ISCED level 02 (for Israel) (CQ+) rotated versions of the instruments were used to test alternative formats or versions of items.

2. In the TALIS Starting Strong 2018 questionnaires, some questions ask for information on respondents' teaching of one particular group of children. In order to randomise the selection of the group of children, the question asks ECEC staff to think of one specific group. This group of children is referred to as the "target group".

3. The following versions were released: staff questionnaire ISCED level 02 (SQ); leader questionnaire ISCED level 02 (LQ); staff questionnaire U3 (SQ U3); leader questionnaire U3 (LQ U3); combined centre questionnaire ISCED level 02 (CQ); combined centre questionnaire U3 (CQ U3); combined centre questionnaire plus ISCED level 02 (for Israel) (CQ+).

4. The rationale for the use of the social desirability scale used in TALIS 2013 was that social desirability and other response styles were frequently mentioned in TAG discussions as potential threats to the validity of the TALIS findings. However, no evidence was available to evaluate the impact of response styles as potential confounders of project findings. The question of the potential impact of corrections for the patterning of country differences was considered particularly important. The scale that was designed for assessing social desirability was derived from the frequently employed Marlowe-Crowne Social Desirability Scale that measures positive and negative impression management.

5. Impression management is the "the processes by which people control how they are perceived by others" (Leary, 2001[5]).

References

Crowne, D. P. and D. Marlowe (1960), "A new scale of social desirability independent of psychopathology", *Journal of Consulting Psychology*, Vol. 24/4, pp. 349–354. [1]

Leary, M. (2001), *International Encyclopedia of the Social & Behavioral Sciences, 2001*, https://www.sciencedirect.com/topics/psychology/impression-management (accessed on 19 July 2019). [5]

Mathers, S. et al. (2013), *Mapping the Early Childhood Environment Rating Scale to the Early Years Foundation Stage (EYFS)*, A+ Education Ltd., Manchester, http://www.ecersuk.org/resources/ECERS-EYFS+mapping+Final+Jan+2013.pdf. [4]

Sim, M. et al. (2019), "Starting Strong Teaching and Learning International Survey 2018 Conceptual Framework", *OECD Education Working Papers*, No. 197, OECD Publishing, Paris, https://dx.doi.org/10.1787/106b1c42-en. [2]

Siraj, I., Kingston, D. and E. Melhuish (2015), *Assessing Quality in Early Childhood Education and Care: Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale for 2-5-Year-Olds Provision*, Trentham Books, London. [3]

Chapter 4. Preparation of national survey instruments

Administration of Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 in the participating countries required national study centres to prepare national survey instruments by adapting and translating the English source version of the survey instruments. In keeping with the survey's strong emphasis on quality control, the TALIS Starting Strong International Consortium developed a stringent verification process designed to ensure the high quality and international comparability of the survey data.

This chapter provides an overview of the preparation of the national survey instruments for TALIS Starting Strong 2018. It summarises the procedures applied to translate and adapt the international source version of the instruments into the identified target languages and outlines the processes of adaptation, translation, and layout verification. The chapter also reflects on the outcomes of these procedures and processes by discussing feedback from the national study centres.

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

4.1. Overview

This chapter describes the preparation of national survey instruments for the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018. The TALIS Starting Strong International Consortium released the international source version of the survey instruments in English to the national study centres of the participating countries at three different stages of the project: pilot, field trial and main survey, as described in Chapter 1 of this report. National centres received these instruments at the beginning of each stage and used them to prepare the survey instruments in their respective country's target language or languages.

The first stage, the *pilot study*, conducted in October 2016, assessed the targeted respondents' comprehension of the items and subsequently fine-tuned the survey instruments in preparation for the field trial (second stage). The instruments for the pilot study were derived from the TALIS 2018 pilot study and adapted to the early childhood education and care (ECEC) context (see Chapter 3 for a detailed description of instrument development). Because the aim of the pilot study was to gather qualitative information rather than quantitative, internationally comparable data, the piloted survey instruments did not undergo the formal processes of translation, adaptation and layout verification.

The second stage, the *field trial* in May and June 2017, incorporated feedback from the pilot study into the survey instruments. Unlike the pilot study, the field trial was designed to gather quantitative and internationally comparable data, albeit from small sample sizes and with limited weighting. The consortium consequently employed a stringent four-part verification process designed to ensure the TALIS Starting Strong instruments were of high quality.

1. *Translation and adaptation of survey instruments and national translation review:* The national study centres translated the international source version of the questionnaires into the respective target language(s) and adapted them to suit the respective national ECEC contexts. Centres also used national adaptation forms (NAFs) to document the desired adaptations as well as national additions of items, response options, and questions.
2. *International adaptation verification:* The international study centre (ISC) at the IEA Hamburg verified the national adaptations documented by the national study centres on the NAFs and approved those that improved the national survey instruments without diminishing their ability to collect internationally comparable data. Approved national adaptations were subsequently implemented by the ISC.
3. *International translation verification:* The IEA engaged native-speaking linguistic experts at cApStAn Linguistic Quality Control to verify the equivalence of the national versions of the survey instruments provided by the national study centres by assessing those versions against the international source version.
4. *International layout verification:* The ISC ensured that the national survey instruments were assembled correctly and fulfilled the standardised layout requirements.

This four-part verification method resulted in all participating countries producing national survey instruments efficiently and of high quality. The instrument preparation and verification work thus ensured that the national TALIS Starting Strong instruments yielded data that can be used to make reliable inferences about and conduct reliable cross-national

comparisons of the early childhood education and care environments in the participating countries.

In order to standardise the verification process, the international consortium defined a number of procedural requirements for translating and adapting the international source version and for verifying national adaptations and translations as well as the layout of the final national instruments. These requirements, defined in the TALIS Starting Strong technical standards 4.1–4.27 (see Annex B of this report) and additional survey manuals, provided the national project managers (NPMs) at the national study centres with the information they needed to prepare and review the national survey instruments.

During the third stage of TALIS Starting Strong 2018, that is, the *main survey*, which took place from April to December 2018, the consortium provided the national study centres with the international source version of the TALIS Starting Strong instruments. Preparation of the subsequent main survey national instruments built on the processes of adaptation, translation, and layout verification evident in the earlier stages and incorporated the feedback and lessons learned from the field trial into the instruments.

Throughout the stages of TALIS Starting Strong 2018, the consortium strived to simplify and structure the verification processes. To that end, the consortium and the national study centres used software provided by the IEA. The software included the following:

- the IEA eAssessment Designer, hereafter referred to as "the Designer", used to implement national adaptations
- the IEA eAssessment Translation System, hereafter referred to as "the Translation System", used to facilitate the creation, verification and revision of national translations
- the IEA Online Survey System (OSS), which provided the platform for the online administration of the survey
- the IEA Data Management Expert, used to enable manual data entry by countries that administered the TALIS Starting Strong questionnaires in paper form.

Focusing predominantly on the main survey, the remainder of this chapter provides further information about the preparation of the national survey instruments by describing the consortium's involvement in the different parts of the verification process (adaptation verification, translation verification, and layout verification).

4.2. Instruments requiring adaptation, translation, and layout verification

As described in Chapter 1, the scope of TALIS Starting Strong 2018 encompassed two populations of interest: ECEC staff and centre leaders working at centres providing services for children at ISCED level 02 (hereafter referred to as "ISCED level 02"); and ECEC staff and centre leaders working at centres providing services for children under the age of three years (hereafter referred to as "U3"). For each population of interest, the consortium provided the national study centres with the international source version of the following survey instruments:

- leader questionnaire (LQ) and cover letter[1]
- staff questionnaire (SQ) and cover letter
- combined centre questionnaire (CQ) and cover letter (to be used by sampled ECEC centres that consisted of only one person)

- combined centre questionnaire plus (CQ+)[2].

All participating countries administered the survey instruments in ISCED level 02 centres and therefore prepared national versions of the ISCED level 02 questionnaires in the languages in which the questionnaires would be administered. Because Denmark, Germany, Israel and Norway also used the questionnaires for U3, these countries' national study centres also prepared national versions of the U3 questionnaires (see Table 4.8). All national versions of the survey instruments were subject to the aforementioned processes of adaptation, translation, and layout verification.

Participating countries could also choose to adapt the survey instruments to fit their survey administration mode(s) and the organisational setting(s) of the participating ECEC centres. When administering the survey instruments, countries could administer them on line, in paper form or in both modes (see Chapter 1). Most countries opted for online administration only. Israel and Japan issued the questionnaires in paper form only; Chile and Germany used both modes. Those countries that opted for online administration prepared national versions of the cover letters accompanying the instruments. The cover letters were also subject to adaptation, translation, and layout verification.

Although the international source version of the survey instruments was designed to fit both centre-based and home-based centres, countries had the option of compiling two separate versions of the questionnaires so as to align them more closely to the organisational setting of the sampled centres. Of the participating countries, only Denmark and Germany created a second version of the questionnaires, both of which were tailored to better fit the home-based ECEC centres. Because the ISC anticipated that differences between the two versions would be minimal (e.g. adaptations encompassing setting-specific terminology), they decided that they did not need to verify the home-based centre versions. Instead, they asked the NPMs to carefully adapt the verified centre-based questionnaires to the context of the home-based centres.

Table 4.8 shows the survey instruments that were prepared for each participating country and that subsequently underwent adaptation, translation, and layout verification. The table also indicates which countries prepared cover letters. The notes under the table provide information about the adaptations to the home-based settings.

Table 4.8. Prepared survey instruments by country and population of interest

Participating country	Leader questionnaire (LQ)	Staff questionnaire (SQ)	Combined centre questionnaire (CQ)	Combined centre questionnaire plus (CQ+)
ISCED level 02				
Chile	X+	X+		
Denmark	X+	X+		
Germany	X+	X+		
Iceland	X+	X+		
Israel[1]		X		X
Japan	X	X		
Korea	X+	X+	X+	
Norway	X+	X+		
Turkey	X+	X+		
Under 3				
Denmark[2]	X+	X+		
Germany[3]	X+	X+		X+
Israel[1]	X	X		
Norway	X+	X+		

Notes:
X denotes that the country used the respective questionnaire; X+ indicates that the country used a cover letter to facilitate online administration of the respective questionnaire.
1 Israel had two independently operating national study centres, one per population of interest.
2 Denmark adapted the U3 staff questionnaire for administration in home-based centres.
3 Germany used the U3 CQ for administration of the survey in home-based centres.

4.3. Identifying, documenting, approving and implementing national adaptations

Because the objective of the adaptation process was to adjust the consortium-provided English international source version to fit the ECEC contexts in the participating countries, part of the NPMs' responsibility was to adapt this version to those contexts. The NPMs in the national centres, along with their team members, began this process by identifying the desired adjustments and documenting them on the national adaptation forms (NAFs) provided by the consortium. After reviewing and approving the adaptations, the ISC used the Designer to implement them in the national instruments. ISC members also discussed any matters of concern with the OECD Secretariat and RAND Europe.

Throughout the adaptation process, the consortium emphasised that centres needed to keep their national adaptations to a minimum to ensure not only equivalence of content across the national survey instruments but also comparability of the data collected by them. However, some adaptations, such as references to aspects of national education systems or place names, were necessary to ensure that respondents in all participating countries received survey questions that were equivalent in meaning. In short, the adaptation process encompassed four types of change:

1. *Structural adaptations:* These adaptations referred to additions to or splitting of categories and/or additions of national questions or dimensions. They also encompassed decisions not to administer categories, questions or dimensions, and changes such as width and range definitions for certain questions.
2. *Non-structural adaptations:* These adaptations were defined as any adaptation that did not alter the structure of the international questionnaires, but resulted in the meaning of the international question changed to fit the local context.

These adaptations differed from translations because they replaced terms or phrases in the source version with specific national terms or phrases.

3. *Required adaptations (localisations):* Localisations were adaptations that were necessary to maintain equivalent meaning across the national survey instruments and ensure the comparability of the collected data. These adaptations were often names of people or places, and they were made because of the need to maintain their essential characteristics (e.g. easy to read, common, denoting a particular meaning, such as "centre management team"), or they were terms relating to educational systems within countries (e.g. the equivalent term for "ICT" or "ISCED level 02").
4. *Optional adaptations (co-ordinated national options):* Co-ordinated national options were optional questions whose implementation was co-ordinated between several participating countries in order to ensure responses to the questions could be compared across those countries. The co-ordinated national options underwent the same verification process used for the corresponding international questions.

The ISC took care to emphasise in the NAFs (sent to the national centres) all required adaptations (localisations) and optional adaptations (co-ordinated national options) in order to draw the NPMs' attention to them.

The national study centres were responsible for documenting all of their adaptations in the NAFs. The ISC then subjected these adaptations to a thorough verification based on judgements of necessity, appropriateness and cross-national equivalence. The ISC's next step was to use the Designer to implement all approved adaptations in the national survey instruments, after which the national study centres reviewed the survey instruments to verify that the adaptations had been correctly implemented. This stage of the adaptation process also allowed for a final round of revisions that the ISC could subsequently implement. The adaptation process ended when the NPMs informed the ISC that they had completed their review of all implemented adaptations in the TALIS Starting Strong instruments.[3]

4.4. Identifying target languages

As stipulated in the TALIS Starting Strong 2018 technical standards 4.1–4.3 (see Annex B), participating countries administered the survey in a common (state) language. The TALIS Starting Strong International Consortium encouraged national study centres in countries with two or more commonly spoken languages to issue the survey instruments in all of those languages.

In compliance with the standards, most of the participating countries decided to administer the questionnaires in one language. Of the nine participating countries, only Israel and Norway administered the survey in two languages. All language sets of the survey instruments underwent two rounds of verification, once in preparation for the field trial and once in preparation for the main survey. Table 4.2 lists the language(s) of the survey instruments used in each participating country.

Table 4.9. Languages of administration

Participating country	Language(s) of administration
Chile	Spanish
Denmark	Danish
Germany	German
Iceland	Icelandic
Israel	Arabic and Hebrew
Japan	Japanese
Korea	Korean
Norway	Bokmål and Nynorsk
Turkey	Turkish

4.5. Engaging translators

Another of the NPMs' responsibilities was that of initiating and co-ordinating translation of the survey instruments into the identified language(s) of administration. The consortium advised NPMs to engage at least one translator and one translation reviewer for each target language in order to ensure high-quality translations. More specifically, the consortium advised NPMs to recruit translators who fulfilled the following requirements:

- an excellent knowledge of English
- an excellent knowledge of the target language (mother tongue)
- familiarity with survey instruments
- experience in translating electronic text.

The consortium also advised that the translation reviewers should be people who had a professional background as ECEC staff and possessed the following qualities:

- an excellent knowledge of English
- an excellent knowledge of the target language (mother tongue)
- familiarity with the country's current ECEC context.

The translators were responsible for creating national translations of the international source version of the survey instruments, while the translation reviewers were responsible for judging the appropriateness of the translations in light of the national ECEC context applicable to the specific population(s) of interest. The NPMs subsequently reviewed the translations and the reviewers' comments and made changes where they deemed appropriate. By involving three individuals who independently compared the national translations against the international source version in English, this tripartite method ensured that the translations produced by the national study centres were of high quality.

4.6. Producing translations

The national research centres used the international source version provided in English by the consortium to produce national translations. The production of national translations of the survey instruments had two components, the first being the adaptation of terminology and classifications to fit national standards and requirements (see also Chapter 3), and the second being the translation of the questions into the language(s) of administration.

Although 2018 marked the first cycle of TALIS Starting Strong, some of the items in the questionnaires were based on items that had been translated and administered during previous TALIS cycles. Because the first round of translation and translation verification of the full complement of TALIS Starting Strong questionnaires took place before the survey's field trial in May and June 2017, most of the items in the main survey questionnaires had already been translated and verified. Exceptions included the revisions made to the items after the field trial.

As mentioned above, the national study centres used the Translation System to translate and verify the translations, regardless of the mode of survey administration (paper or online). The system thus represented the main platform from which NPMs, translators, translation reviewers and translation verifiers worked to create high-quality national survey instruments.

For questions that remained unchanged or were only minimally revised after the field trial, the ISC prepopulated the Translation System with the translations used in the field trial. This procedure simplified the translation process for the national study centres and ensured consistency between the field trial and main survey translations of these questions. For diverse reasons, the consortium, with approval from OECD, agreed that Israel did not have to use the Translation System but could submit their main survey questionnaires as Microsoft Word documents.

Due to the complexity and idiosyncrasy of languages, the consortium did not provide explicit instructions for translating the English source version into the target languages. Instead, the survey manuals drew the national study centres' attention to several issues likely to arise during the adaptation and translation processes. The manuals also listed the key actions that centres needed to take during the translation process. These were:

- identifying and minimising cultural/contextual differences
- finding words, terms and phrases in the target language that were equivalent to those in the international version
- making sure that the essential meaning of the text and the reading level did not change
- ensuring that when national adaptations were required, they were appropriately implemented and documented
- remaining aware of and minimising changes in layout due to translation.

The consortium furthermore required centres to adhere to the following guidelines when assessing the linguistic quality of their translations:

- the translated texts should have the same register (language level and degree of formality) as the source text
- grammar and usage (e.g. subject/verb agreement, prepositions, verb tenses) in the translated texts should be correct
- unless given permission to do so by the ISC, centres should not try to clarify text from the source version in the translated versions, nor should they omit text from the source version or add more information to the translated version
- qualifiers and modifiers in the translated version should be equivalent to those used in the source version and also appropriate for the target language

- social, political and historical terminology in the translated texts should be equivalent to the terminology used in the source version and also appropriate for the target language and level of education
- idiomatic expressions should be translated appropriately and not word by word
- spelling, punctuation and capitalisation in the translated texts should be appropriate for the target language(s) and the country's national contexts.

Where relevant, the consortium also emphasised the importance of consistent translations across all languages of administration. Consequently, NPMs were advised to ensure that translators working on different language versions collaborated with one another during the production of several national survey instruments.

At the end of the translation process, the translation reviewers recruited by the NPMs reviewed the national version(s) of the survey instruments to ensure that the translations were appropriate, consistent with the instruments used in the field trial, and complied with the stipulated translation requirements. After this review, the translators integrated, where they thought appropriate, the reviewers' suggested changes into the instruments. If disagreements between translators and translation reviewers arose, the NPMs acted as arbitrators and had the right to make the final decision.

4.7. Process of international translation verification

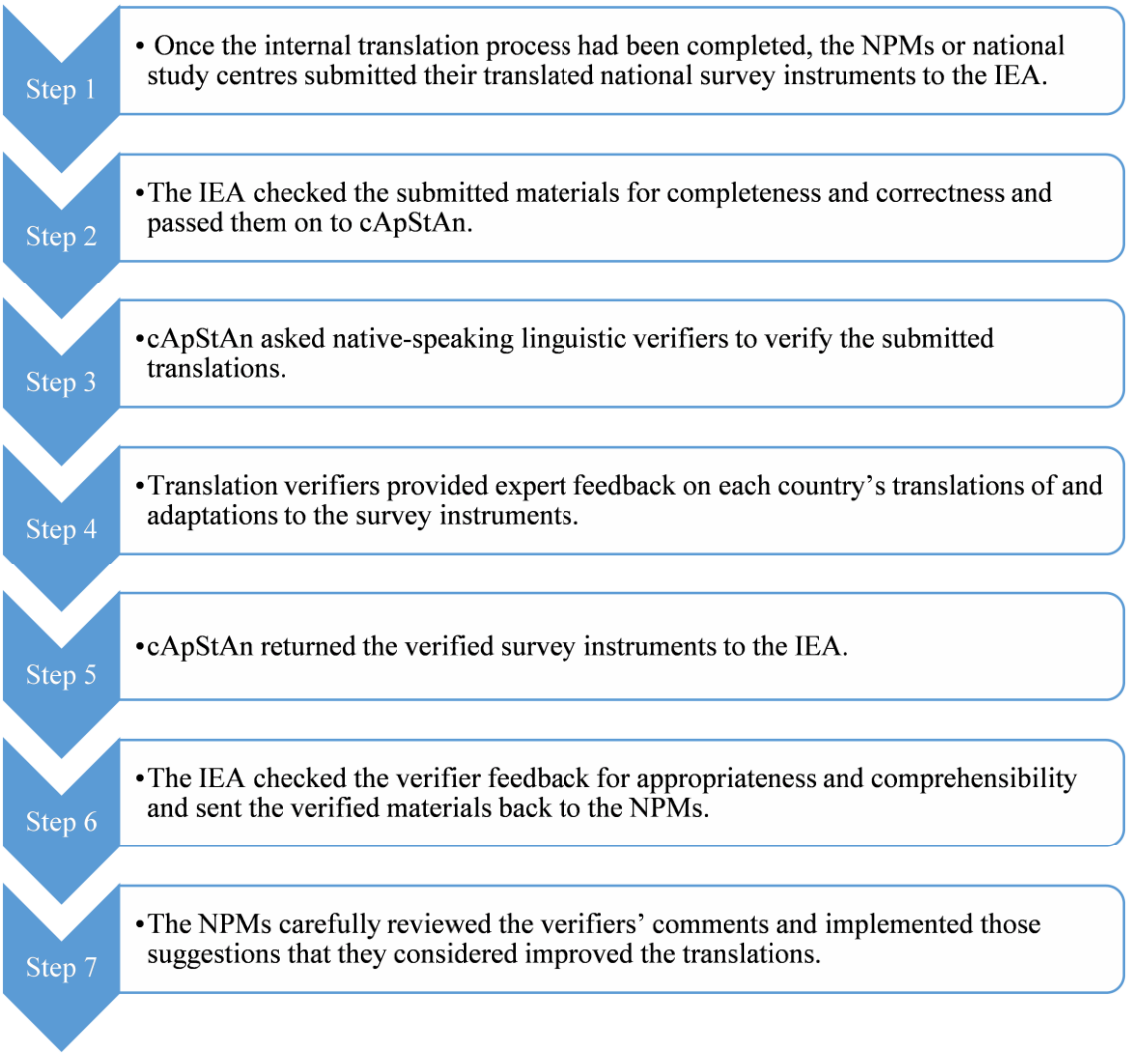
As stated in the TALIS Starting Strong technical standards (see Annex B), international quality control was an essential part of all stages of instrument preparation. Consequently, the consortium also implemented a rigorous translation verification process overseen by the IEA. Once again, the aim of this step was to ensure the high quality of the national survey instruments and their equivalence in meaning to the international source version. The IEA achieved this purpose by engaging the services of native-speaking linguistic verifiers through cApStAn Linguistic Quality Control, a leading provider of translation and translation verification services based in Brussels, Belgium.

All national centres, except those in Israel, used the IEA-provided Translation System for the entire translation verification process. Israel gained the ISC's permission to use Microsoft Word (specifically, the track changes and comment functions) for their translation verification process.

The Translation System was designed to make the process as smooth and straight-forward as possible by giving its users access to the latest technological solutions. It provided the NPMs, national study centres, cApStAn, the verifiers and the IEA with the tools necessary to translate, review, verify and revise the national versions of the survey instruments.

The translation verification process for each language version of the survey instruments consisted of the steps shown in Figure 4.2.

Figure 4.2: The different steps of the translation verification process



Having taken care to balance the cultural and national appropriateness of the target versions with faithfulness to the source version, the translation verifiers provided expert feedback on the adaptations and translations (step 4). The verifiers commented on any issues they found in the translated materials and used the following coding system to indicate the severity of those issues:

- *Code 1. Major change or error:* Examples included mistranslations, translations that changed the meaning of a question, omitted or added information, incorrect order of questions, incorrect order of response options, and incorrectly implemented national adaptations.
- *Code 2. Minor change or error:* Examples included spelling errors, grammatical errors and syntax errors that did not affect the comprehensibility of the question.
- *Code 3. Suggestion for alternatives:* Used by verifiers to suggest an alternative wording for an otherwise appropriate translation.
- *Code 4. Acceptable change:* Used by verifiers to indicate that a national adaptation had been correctly documented and implemented.

- *Code 1?* Used by verifiers whenever they were in doubt about the severity of an issue or unsure of how to correct a possible error.

After translation verification, the IEA sent the verified survey instruments back to the respective countries, where the NPMs carefully reviewed all verifier comments (Step 7). The Translation System allowed NPMs to react to each verifier comment in one of three ways:

- *Accept:* If NPMs completely agreed with the verifier's suggestion, they accepted the change and no further action was needed.
- *Modify:* If the NPMs partially agreed with the verifier's suggestion, they had the option of modifying the suggested change and implementing the modified text in the survey instruments.
- *Reject:* If the NPMs disagreed with the verifier's suggestion, they could reject the suggested change in the Translation System. NPMs were also required to provide an explanation of why they rejected the verifier's suggestion.

Because Israel did not use the Translation System, its NPMs used the track changes function of Microsoft Word to accept or reject the verifier's comments and suggestions directly in the Word documents.

Reacting to verifier comments was the last step of the translation verification process. The procedure outlined above allowed the NPMs to conclude the translation verification process in a way that meant the final national survey instruments reflected their final say.

4.8. Layout verification: paper and online data collection

The last step of the national instrument production process occurred after completion of the translation verification process. During this step, the ISC verified the layout of the paper versions and created the online versions of the instruments. The ISC carefully checked the national versions of the instruments against the international source version in English and the documentation in the corresponding NAFs. The aim of the layout verification was to ensure that the national versions of the TALIS Starting Strong instruments looked, as much as possible, like the international source version and that the paper and online versions were therefore, to the greatest extent possible, equivalent.

The process used to verify the layout of the paper version was similar to the adaptation approval process. NPMs submitted all ISCED level 02 and U3 questionnaires and, if applicable, the cover letters for online administration in Microsoft Word, to the ISC for approval. The ISC then compared the ISCED level 02 and U3 questionnaires, including the staff and leader questionnaires, the combined centre questionnaire and the combined centre questionnaire plus, against the international source version.

ISC team members checked each questionnaire for font size, font changes, adjustment of cells, response options, blank pages, word emphases, tracked changes, page breaks, and comments. If ISC team members found deviations from the documentation in a questionnaire's associated NAF, they adjusted the paper version (in Microsoft Word) and asked the respective NPM to verify the correction of the mistake.[4]

After approving the paper-based layout, the ISC simultaneously implemented the instruments for all populations of interest in the IEA Online Survey System (OSS) and then checked each online version of an instrument against its approved paper version. This practice helped ensure that the instruments within one country were the same,

regardless of whether they were administered on paper or on line. The ISC also ran visual checks, employing the same standards and procedures used to verify the paper layout. After finalising each country's complement of online files, the ISC asked the country to thoroughly check the files and to report any mismatches to the ISC. In general, at least two rounds of checking and verifying the language versions of the instruments were needed before a country's NPM finally approved them.

4.9. Feedback from NPMs on the preparation of the national survey instruments

To gain insight into NPMs' perspectives on the verification processes, the consortium designed a survey activities questionnaire (SAQ) and distributed it to the national study centres after the main survey data collection. Nine of the 10 national study centres[5] provided their answers in the SAQ. This section provides a brief overview of the NPMs' feedback.

NPMs reported that the adaptation procedure was clear and that the information the consortium provided about the adaptation process was also clear and sufficient. Overall, national study centres said the adaptation process had been a positive experience for them, and they had only a few minor suggestions for improvement.

NPMs' responses to the question in the SAQ about engaging the services of translators and translation reviewers showed that nearly all of the national study centres were able to employ individuals who fulfilled the requirements set out in the related survey manuals. NPMs' feedback on the translation verification process was mostly positive, and they considered the verifiers' feedback useful. Most national study centres corrected errors identified by the verifiers and adopted verifiers' suggestions for improvement. The only criticism related to the functioning of the Translation System. For instance, one NPM reported that using the Translation System was time-consuming, while another national study centre said they experienced long loading times and difficulty handling different versions of identical items in different survey instruments.

All NPMs said they found the instructions for layout verification provided by the ISC helpful for preparing their national survey instruments. However, in two cases, NPMs detected some minor inconsistencies in questions in the paper instruments after the layout verification had been finalised. One of these cases could be corrected before printing the paper questionnaires; the other case was corrected by including a sticker with the correct question item. The verification process used for the online questionnaires was also seen as relatively positive. Criticism was minor and related to the display of the survey instruments in the IEA Online Survey System (OSS).

In general, analysis of the SAQ showed that NPMs were mostly satisfied with the verification processes. However, countries administering instruments to both populations of interest reported that the verification steps were time-consuming, which was a particular problem for them because their NPMs were operating under a tight schedule. The minor suggestions for improvement provide not only reasons but also a starting point for further development of the processes of adaptation, translation, and layout verification of paper and online questionnaires.

Notes

1. The cover letters used during the online administration introduced the survey and provided individual participants with the information they needed to access the questionnaires.

2. The CQ+ was prepared for administration in Israel's ISCED level 02 centres only.

3. The national team responsible for Israel U3 implemented the following changes to the following questions without seeking adaptation verification:

LQ-04A: item was set to not administered

- LQ-17: part of the question instruction was deleted
- LQ17E: item was set to not administered
- SQ-23C: item was set to not administered
- SQ-35: item was set to not administered
- SQ-35E: item was set to not administered
- SQ-36: option 4 was not administered.

4. Item LQ-8F was identified as missing in the German centre leader questionnaire for ISCED level 02 after it had been printed. The national study centre corrected this omission by inserting a sticker featuring the full item before sending out the questionnaires to centre leaders.

5. Because Israel had two national study centres (one responsible for ISCED level 02 and one for U3), the SAQ was distributed to 10 national study centres in total, despite the number of participating countries being nine.

Chapter 5. Sample design

The Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 administered its instruments to two target populations. The ISCED level 02 target population consisted of early childhood education centres providing ISCED level 02 education and care, as defined by the International Standard Classification of Education 2011 (UNESCO-UIS (Institute for Statistics), 2012[1]), as well as their leaders and staff. The U3 target population consisted of services for children under the age of three years and their centre leaders and staff. This chapter describes the sample design prepared for and implemented in the countries participating in the survey. It reviews all design features such as sampling methods, stratification, and the sample sizes of centres and staff.

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

5.1. Overview

This chapter covers the sample design prepared for and implemented in the countries that participated in the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018. It details the sampling strategies used, describes all steps of the sample selection process, and documents the sizes of the samples in each country. Chapter 9 and Annex E of this report complement this current chapter. Chapter 9 details the strategies used to estimate the characteristics of the populations of interest and their sampling errors, while accounting for the design features described here. Annex E provides the characteristics of each national sampling plan.

The populations of interest comprised centres, along with their leaders and staff, providing either education and care for ISCED level 02 children, or education and care for children under three years of age (U3). TALIS Starting Strong defined staff members as all persons working regularly in a pedagogical manner with ISCED level 02 children or children under three years of age within institutional (officially registered) settings providing early education and care.

The international sampling plan prepared for each of the TALIS Starting Strong 2018 populations was, in essence, a stratified two-stage probability sampling design. This meant that centre staff (second-stage units or secondary sampling units, abbreviated as SSUs) were randomly selected from the list of in-scope staff for each of the randomly selected centres (first-stage or primary sampling units, abbreviated as PSUs). The centre leader in each selected centre was the person selected to complete the leader questionnaire.

5.2. Target population and survey population

TALIS Starting Strong aimed to cover all staff providing early childhood education and care (ECEC) for (a) ISCED level 02 children, and (b) U3 children in each participating country. Staff at ISCED level 02 were those who, as part of their regular duties in a target centre, were providing learning opportunities in programmes at that ISCED level. Staff for children under the age of three years were those who, as part of their regular duties in a target centre, were providing learning opportunities for children in this age bracket. Staff who were providing learning opportunities for both target populations in a sampled centre were also included in each TALIS Starting Strong population of interest. This consideration applied no matter how much or how little work these staff were engaged in.

TALIS Starting Strong's *international target population* restricted the survey to (a) staff who were working in institutional (officially registered) settings providing ECEC, and (b) the leaders in those settings. These settings could be home-based (within the rooms of the respective staff) or centre-based (within rooms specifically dedicated to ECEC). Staff working with children with special needs were in scope if they were working in "regular centres". However, centres catering exclusively for children with special education needs were out of scope. Other staff who were out of scope included short-term substitute educators (to replace staff on sick leave); nannies or other such people involved in informal arrangements; volunteers who occasionally came in to provide a special activity; auxiliary staff (e.g. cleaners, cooking staff) who did not interact regularly in a pedagogical manner with the children; and medical and therapeutic staff (e.g. speech therapists, occupational therapists) whose work was primarily non-pedagogical.

A centre leader was defined as the person with the most responsibility for the administrative, managerial and/or pedagogical leadership in the ECEC centre. In line with

their leadership role, centre leaders could be responsible for monitoring children, supervising other staff, maintaining contact with parents and guardians, and planning, preparing and carrying out the centre's pedagogical work. Centre leaders may have also spent part of their time working with the children.

For national reasons, participating countries could choose to restrict the coverage of their national implementation of TALIS Starting Strong to a specified part or parts of the country. For example, countries could remove from the international target population settings deemed difficult to access or where the instruments required major modification to work well, and thereby create a *national survey population*. Participating countries were invited to restrict these exclusions to the greatest extent possible to maintain international comparability of the survey data. A 5% threshold was adopted for TALIS Starting Strong as an upper limit for the overall exclusion of staff from the survey population. However, because of limitations in the availability of comprehensive and reliable national statistics on enumerations of staff, validating adherence to this rule proved to be difficult for some participating centres.

Annex D of this report presents the national definitions of the TALIS Starting Strong 2018 target and survey populations for ISCED level 02 and services for children under three years of age (U3). The annex also details those centres and their staff that were included in or excluded from the survey. Ultimately, the samples of centres and staff were selected from the national survey population. Figure 5.1 illustrates how the international target population and the national survey population related to each other.

Figure 5.1. Establishing the national survey population

<table>
<tr><td rowspan="3">TALIS Starting Strong 2018 out-of-scope</td><td colspan="3">TALIS Starting Strong 2018
international target population
=
All centres at ISCED level 02 or for children
under three years of age</td></tr>
<tr><td>NATIONAL exclusions</td><td colspan="2">NATIONAL survey population</td></tr>
<tr><td>Defined by the country</td><td>Not sampled</td><td>In sample</td></tr>
<tr><td>• Centres exclusively for children with special needs
• Short-time substitute educators
• Nannies or other informal or unregulated arrangements
• Volunteer persons who would occasionally come in for a special activity
• Auxiliary staff who did not interact regularly with children in a pedagogical manner
• Medical and therapeutic staff whose work was primarily non-pedagogical</td><td>Not more than 5% of staff</td><td colspan="2">At least 95% of staff</td></tr>
</table>

Although home-based settings were part of the international target population for ISCED level 02 and services for children under the age of three years, all participating countries decided to exclude home-based settings from the ISCED level 02 population, while some countries excluded home-based settings providing services for children under three years of age.

The sampling team informed the TALIS Starting Strong national project managers (NPMs) that centre co-ordinators should list all staff including those also working in other centres. Selected staff were asked whether they were working in any other centre and, if so, in how many centres. This information was used during the weighting process to correct for the chance of multiple selection (see "weight adjustments" in Chapter 9). Although the following possibility was very unlikely, a staff member could be selected from more than

one centre. In these instances, staff were asked to complete a questionnaire for each centre they were selected from.

5.3. Sample size requirements

To allow for reliable estimation and modelling while permitting some amount of non-response, TALIS Starting Strong 2018 set the minimum number of centres per country for each population of interest (ISCED level 02 and services for children under the age of three years) at 180. Within each centre, the minimum number of staff members to be selected were eight. If there were fewer than eight staff members in a centre, then all staff members were selected.

TALIS Starting Strong recognised that staff from one centre would be more likely to share opinions and behave in similar ways than would staff across different centres, at least for some of the variables. This tendency for staff in the same centre to be "more alike" than staff from different centres is called a "clustering effect". In essence, as the clustering effect becomes larger, the need to select more centres becomes greater in order to maintain the precision of the estimates evident in a non-clustered sampling design. The clustering effect, together with other features of the sampling design (e.g. stratification, unequal weights), contributes to the design effect. The design effect measures the ratio of the variance within the given sampling design to the variance expected from a simple random sample of the same size (Kish, 1965[2]).

As a goal for this survey, an effective sample of 400 staff was set. Those engaged in the preparation work for TALIS Starting Strong 2018 used, as a working hypothesis, a design effect of two. Working to this assumption, they took design effects from earlier studies with adult clustered populations (mainly from earlier cycles of TALIS and the Teacher Education and Development Study in Mathematics (TEDS-M)) and compared them for various variables. They then averaged any found design effects and adjusted them for typical cluster sizes in ECEC. Note, however, that design effects vary greatly depending on country, population of interest, and variables, a situation that was indeed evident in the large variance in design effects apparent in TALIS Starting Strong 2018 (see Chapter 9).

The design effect and the assumed non-response in TALIS Starting Strong increased the needed sample size. To achieve an effective sample, the sampling team needed to draw a sample of 180 centres, as depicted in Table 5.1.

Table 5.1. Establishing the effective sample size

Effective sample size	a	400
Design effect	b	2
Actual sample size	c = a x b	800
Staff per centre	d	8
Minimum staff response	e	75%
Minimum response per centre	f = d x e	6
Number of centres required	g = c / f	135
Minimum centre non-response	h	75%
Sample size for centres	i = g / h	180

The precision expected from the sample of 8 staff in 180 centres is equivalent to that of a simple random sample of 400 staff selected from the (often unavailable) national list of staff. An effective sample size of 400 staff results in the following 95% confidence limits for sample estimates of population means, percentages, and correlation coefficients.

- Means: m ± 0.1s (where m is the mean estimate and s is the estimated standard deviation)
- Percentages: p ± 4.9% (where p is a percentage estimate)
- Correlations: r ± 0.1 (where r is a correlation estimate).

Participating countries could choose to augment their national sample by selecting more centres, or by selecting more staff within each selected centre, or by increasing both.

In some cases, because the average number of staff in the centres was lower than the number expected in the international plan, the sampling team maintained a minimum total number of participating staff by increasing the number of centres to be sampled. Annex E presents the number of selected centres per participating country and population of interest.

5.4. National sampling strategies

Participating countries could suggest variations to or adaptations of the international sampling plan to better suit their national needs or conditions. All changes to the international sampling plan had to be reviewed and approved by the sampling team.

5.4.1. Centre sampling frames

Participating countries were asked to provide the IEA Hamburg with a current and complete list of centres providing ECEC for the respective populations of interest. This list constituted the centre sampling frame for TALIS Starting Strong and was expected to correspond to the survey population as defined and described on the sampling forms.

The sampling frame had to contain certain key variables. These were a national centre identifier, a measure of size (preferably the number of staff for the population of interest), and values for variables to be used for stratification.

5.4.2. Stratification

Stratification is part of many sampling designs and entails the grouping of sampling frame units by common characteristics. Examples for such groups of units (centres in the case of TALIS Starting Strong) are different types of centres, geographic regions, urbanisation level, or source of funding. TALIS Starting Strong generally used stratification to improve the efficiency of the sample design. However, the international sampling plan did not require any stratification of the centres or of the staff within the selected centres.

The sampling team invited participating countries to discuss their respective stratification strategies with them. If a country required estimates with high precision levels for specific subgroups of interest, the sampling team could select more centres from within specific strata than a proportional sample allocation would have required. Therefore, the participating country and the sampling team together determined how many centres should be selected per stratum. A proportional sample allocation would have reflected the proportions of the population regarding stratum size in the sample. The use of sampling weights makes it possible to implement disproportional sample allocations without biasing population estimates.

Stratification could be done explicitly (whereby a fixed portion of the total sample was allocated to the stratum) or implicitly (whereby the variable was used to sort the sampling frame before sample selection, thus giving, on average, a proportional representation of the

implicit strata in the sample). Samples for the different explicit strata were completely independent from one another.

In most cases, stratification resulted in a combination of some or all of the details relating to geography, urbanisation, source of financing, type of settings, and language. Annex E provides stratification details for each participating country and each population of interest in which it participated.

5.4.3. Sample selection

The sampling team at the IEA selected all the samples of centres.

The usual method the team used to select the centre samples was systematic random sampling with probability proportional to size (PPS) within explicit strata. In cases where there was no measure of size (MOS) available or where most centres had fewer than eight staff, the team used systematic equal probability random sampling to select the centres. When they used implicit stratification, they sorted the centres in the explicit strata by implicit strata and MOS prior to sampling. The team always used MOS to sort the sampling frames before sampling, whether or not stratification was applied. Sorting by MOS was done in a serpentine manner, which meant alternating increasing order and decreasing order so that adjacent schools would be of similar sizes, even across implicit strata. This method is useful when creating replication zones for estimation of sampling error (see Chapter 9).

Replacement centres

Whenever possible, TALIS Starting Strong 2018 sampling team assigned two replacement centres for each sampled centre. These centres were the centre just above and the centre just below the selected centre on the sampling frame sorted by MOS, when available. The replacement centres also had to come from the same explicit stratum as the sampled centre. It was expected that this strategy would help maintain the sample size and minimise non-response biases by using centres with characteristics similar to those of the non-responding centres. Centres selected for the original sample could not be simultaneously selected as a replacement centre.

Systematic random sampling with probability proportional to size (PPS)

The mechanics of systematic random sampling with PPS can be described as follows. Let M be the sum of all measures of size (MOS) in an explicit stratum, let m_i be the MOS for centre i in the explicit stratum and M_i be the cumulative sum of the centre sizes up to and including centre i, and let n be the number of centres to be sampled from that explicit stratum. From here, the sampling interval k is computed as M / n, and a random starting point d (in each stratum) is drawn at random from the interval $1 \leq d < k + 1$. The sample is selected by walking steps of fixed length k along the (ordered) sampling frame. Where the step lands points to the centre to be added to the sample. The pointer for the j^{th} centre selected is calculated by $\text{int}(d + (j - 1) \times k)$. This process was completed separately for each explicit stratum.

Table 5.2 illustrates how systematic random sampling with PPS can be implemented using an ordinary spreadsheet. In this illustration, explicit stratum "A" consists of 12 centres, and a sample of $n = 3$ centres is needed from this stratum. The sampling interval $k = 209 / 3 = 69.7$. Suppose that the random start is $d = 49$. Here, for the first selection, $j = 1$ and the pointer is $\text{int}(49 + (1 - 1) \times 69.7) = 49$. In other words, the selected centre is the one containing the 49th staff member from a hypothetical comprehensive staff list. If $j = 2$, the

pointer is at int(49 + (2 - 1) × 69.7) = int(118.7) = 118. Finally, the pointer is at int(49 + (3 – 1) × 69.7) = int(188.4) = 188. The centres that are automatically selected as replacements are the centres immediately before and after a selected centre, if available. If the selected centre is at the beginning or at the end of the explicit stratum, the replacements might be different ones. In this example, the second replacement for centre 12 is the one above the first replacement.

Table 5.2. Illustration of systematic random sampling with PPS

National centre ID	Explicit stratum	Implicit stratum	MOS	Cumulative MOS	Centre selected	Pointer	Selections and replacements
1	A	1	10	10			
2	A	1	12	22			
3	A	1	15	37			R1
4	A	1	17	54	1	49	S
5	A	2	20	74			R2
6	A	2	18	92			
7	A	2	16	108			R1
8	A	2	16	124	2	118	S
9	A	3	15	139			R2
10	A	3	17	156			R2
11	A	3	26	182			R1
12	A	3	27	209	3	188	S

Note: S = sampled centre; R1 = first replacement; R2 = second replacement.

Systematic equal probability random sampling (SRS)

Systematic equal probability random sampling requires a similar approach, but it does not use an MOS. The pointer directly indicates which centre to sample.

Table 5.3 shows another example of a sample selection with systematic equal probability random sampling (SRS). The explicit stratum B has 12 centres, and three of them should be selected. The sampling interval k is 12 / 3 = 4. In this case, the random start is $d = 3$. Therefore, the three pointers are int(3) = 3, int(3 + 4) = 7, and int(3 + 2 × 4) = 11.

Table 5.3. Illustration of systematic random sampling with SRS

National centre ID	Explicit stratum	Implicit stratum	Centre selected	Pointer	Selections and replacements
1	B	1			
2	B	1			R1
3	B	1	1	3	S
4	B	1			R2
5	B	2			
6	B	2			R1
7	B	2	2	7	S
8	B	2			R2
9	B	3			
10	B	3			R1
11	B	3	3	11	S
12	B	3			R2

Note: S = sampled centre; R1 = first replacement; R2 = second replacement.

Within-centre sampling in the participating countries

At the end of the centre selection process, the sampling team sent each participating country an MS Excel file that contained the selected centres for the original sample, the replacement centres, and a standardised TALIS Starting Strong 2018 centre identification number.

The IEA Hamburg provided each participating country with the IEA Windows Within-School Sampling Software (WinW3S) for selecting the staff. WinW3S facilitated the creation of the staff sampling frames and the random selection of staff, thus ensuring compliance with the sample design and the need to furnish complete documentation.

Annex E presents the sizes of the centres and staff samples for each participating country.

5.4.4. Sampling for the field trial

During May and June 2017 and prior to the main data collection, the TALIS Starting Strong instruments were field trialled in each participating country. The sample of 30 centres needed for the field trial in each country was selected simultaneously with the main survey sample. In practical terms, the first sampling step involved selecting enough centres for both stages (210 centres at minimum). During the second step, centres were subsampled from this sample for the field trial. The third and final step was to assign the replacement centres, one for the field trial and two for the main survey.

The simultaneous selection of the centre samples for the field trial and the main survey made it possible to control sample overlap; it also helped reduce the response burden on participating centres. Avoiding overlap was possible if the populations were big enough, that is, had more than 210 centres eligible for the survey. Overlap occurred only in Iceland, where all centres had to be selected for the main survey.

In some cases, circumstances required that a new sample be selected for the main survey after the field trial.

5.4.5. Samples by participating country

The country reports in Annex E provide more details on each set of plans, including population statistics, exclusions, the sampling design, and the realised sample.

References

Kish, L. (1965), *Survey Sampling*, Wiley, New York. [2]

UNESCO-UIS (Institute for Statistics) (2012), *International Standard Classification of Education: ISCED 2011*, UNESCO-UIS, Montreal, http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf. [1]

Chapter 6. Field operation procedures

This chapter summarises the procedures and outcomes of the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 field operations. It focuses on the work within national study centres that included contacting centres, performing the within-centre sampling, and monitoring the collection of data for the two populations of interest in the TALIS Starting Strong 2018 main survey. The chapter also references the materials and software that the international study centre (ISC) provided to all national study centres. The ISC asked all participating countries to follow the specified standards and procedures and emphasised that deviations from these would be allowed only in certain cases and if discussed and agreed to with the TALIS Starting Strong 2018 International Consortium and the OECD. The overall administration of TALIS Starting Strong 2018 in all participating countries was successful. No mayor obstacles were encountered, and the field operation procedures, as defined and communicated by the ISC, were met.

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

6.1. Technical standards, manuals and software

During all phases of TALIS Starting Strong 2018, the national study centres adhered to the standardised procedures prepared by the international study centre (ISC) and its consortium partners. The ISC outlined these procedures in the following documents, which they released to national study centres prior to the field trial. The ISC also released updated versions of these documents before the main survey.

- *TALIS Starting Strong 2018 Technical Standards:* The ISC prepared several versions of this document during implementation of the study, each time considering feedback from diverse stakeholders such as the Extended ECEC Network members, the national project managers (NPMs), the OECD and the survey's Technical Advisory Group (TAG). The ISC released the final version of this document just before the main survey so that its content accommodated guidance arising out of implementation of the field trial. Release of the document was also contingent on the Extended ECEC Network members approving the standards, which they did during their sixth meeting in October 2017 (see Annex B).
- *TALIS Starting Strong 2018 National Project Managers Manual:* The ISC prepared this manual to provide NPMs with an overview of TALIS Starting Strong 2018, details of the tasks NPMs and national study centres would carry out, and information about key milestones and deliverables.
- *Survey Operations Procedures Unit 1: Sampling Centres:* Prepared by the IEA's sampling team, this document defined the target populations of staff and leaders in both populations of interest (ISCED level 02 and U3). The manual also provided descriptions of how to prepare and implement a national sampling plan, how to prepare the sampling frame and how to select the centre sample.
- *Survey Operations Procedures Unit 2: Contacting Sampled Centres:* As indicated by its title, this document, provided by the ISC, presented guidelines for obtaining and increasing co-operation with sampled centres, with particular emphasis given to the importance of public relations strategies. The document also presented centres with descriptions of how to adapt the centre co-ordinator manual to the national context as well as details of instrument shipping and quality control measures.
- *Survey Operations Procedures Unit 3: Instrument Preparation:* The ISC released this document to provide national study centres with instructions on how to produce internationally comparable national versions of the released international survey instruments. The document also contained detailed descriptions of the different verification steps (adaptation, translation, layout) included in the process.
- *Survey Operations Procedures Unit 4: Data Collection and Quality Control:* The ISC used this document to inform national study centres on how to prepare for and support data collection in the sampled centres. The information included detailed explanations of the listing and within-centre sampling of staff members, assigning and administering the instruments, and tracking and monitoring the quality of the data collection.
- *Survey Operations Procedures Unit 5: Data Capture Procedures:* The ISC used this document to show the national study centres how to use the IEA Data

Management Expert (DME) software. The purpose of this software was to capture and verify data collected by the paper versions of the survey questionnaires.

- *Centre Co-ordinator Manual:* This manual, written by the ISC, was intended for centre co-ordinators. The centre co-ordinator in each participating centre served as the main contact person for the national study centre. The person fulfilling the role of a centre co-ordinator could be a staff member or the centre leader of the sampled centre. The manual described, in detail, the steps for listing and tracking staff members and for organising administration of the survey on site. NPMs were responsible for translating the manual into the language(s) in which they would administer the survey and for adding national information where necessary. Responsibility for translations and adaptations rested solely with the NPMs. The ISC asked international quality observers (IQOs) to make sure the NPMs used the correct template. Chapter 7 of this technical report provides more details on quality assurance.
- *National Quality Observer Manual:* Prepared by the IEA Amsterdam, this manual showed NPMs how to implement a national quality observation programme. This programme related closely to the programme carried out by the international observers. However, NPMs were free to adapt the manual and procedures according to their country-specific needs.
- *International Quality Observer Manual:* The IEA Amsterdam prepared this document and delivered it directly to the international quality observers, all of whom were contracted by the IEA. The manual outlined the tasks the observers needed to complete in order to check the quality of the survey operation procedures within the participating countries. The observers interviewed each NPM and national team and visited a subset of participating centres to interview the centre co-ordinators. Observers documented the results of their visits in centre visit records.
- *Note on the Investigation of Non-response Bias:* Prepared by the IEA's sampling team, this document explained how to investigate the extent of agreement between the characteristics of the sample (at the centre and staff level) and independently available population statistics. Whether participating countries had to carry out this analysis depended on their participation rate. The consortium asked national study centres to provide this additional information only in those instances where the data collection had yielded less than the minimally required 75% rate for centre participation *after* replacement but had still encompassed at least 50% of the original sample of centres.

In addition to providing the manuals and documents just described, the ISC gave NPMs and their respective national study centres the following main software packages to facilitate their data collection:

- The *IEA Windows Within-School Sampling Software (WinW3S):* This software package helped the national data managers prepare the survey listing forms, qualify and randomly sample staff members in the selected centres, and produce tracking forms for the sampled individuals. The software stored all tracking data in a single database so that this information could be used later to verify the integrity of the sampling procedures, to verify the completeness of the response data and to compute sampling weights and participation rates.

- The *IEA eAssessment:*[1] Centres used this second software package to adapt and verify the translation of the international version of the survey instruments into national contexts and languages. More specifically, centres used the IEA eAssessment Designer to implement national adaptations and the IEA eAssessment Translation System to facilitate the creation, verification and revision of national translations (see Chapter 4 of this report for more details).
- The *IEA Online Survey System (OSS):* This third software package helped the ISC prepare the national questionnaires for online administration. During this process, the ISC asked the participating countries to review the prepared online questionnaires via the Internet. The web-based monitor component of the OSS allowed national study centres to audit participation in real-time and to follow up those centres that returned incomplete questionnaires or did not return questionnaires.
- The *IEA Data Management Expert (DME):* The DME software enabled national study centre staff to capture the data through keyboard data entry and to perform a range of validity checks on the entered data. The DME databases included codebooks for each of the questionnaires, thus providing all the information necessary for producing data files for each instrument that adhered to the standard international format (see Chapter 8 for more details).

During the meetings with the NPMs and data management training sessions (in January 2017), the ISC described and explained the field operation procedures outlined in the survey operation procedure units, manuals and guidelines and provided guidance on how to use the software packages. Representatives of the national study centres also participated in hands-on training sessions to practise the correct handling of the software and some procedures. If anyone from the national study centres had queries or concerns regarding procedures, guidelines or software during implementation, ISC members were always available to support these people and to help them find appropriate solutions.[2]

6.2. Administering TALIS Starting Strong 2018

The ISC developed and agreed on individualised survey preparation schedules with each participating country. The complexity of instrument preparation in each country depended on whether it intended administering the survey to both populations of interest (see Table 1.1. in Chapter 1) and/or in more than one language (see Table 4.2. in Chapter 4). Complexity also depended on which questionnaire version(s) (e.g. combined centre, combined centre plus) the country intended using (see Table 3.7. in Chapter 3).

The ISC carefully considered these extra layers of complexity when setting up the individual survey preparation schedules. The schedules allowed countries to monitor their progress, plan for staff resources and ensure a smooth workflow.

Table 6.1 gives an overview of the international data collection windows and the time periods spent on instrument preparation for countries in both hemispheres.

Table 6.1. Instrument preparation periods

	Preparation time for instruments	Official main survey data collection window
Northern Hemisphere	1 Dec 2017–31 Mar 2018	1 Mar–31 May 2018
Southern Hemisphere	15 Feb–30 May 2018	1 Aug–31 Oct 2018

Countries following the Northern Hemisphere schedule started their national instrument preparation immediately after release of the international source questionnaires in December 2017 and used the first quarter of 2018 to finish this work. Countries following the Southern Hemisphere schedule started this process mid-February 2018 and finished the work at the end of May 2018.

As indicated above, countries that administered the survey questionnaires to both populations of interest or administered the questionnaires in more than one language or used additional questionnaire versions followed a somewhat complex and tight schedule. Norway and Israel administered the questionnaires in more than one language. Germany, Israel and Korea used additional questionnaire versions (combined centre questionnaire or combined centre questionnaire plus). Denmark did not use a combined questionnaire but adapted specific terms for staff working in home-based settings in the staff questionnaire.

Instrument preparation was relatively straightforward and timelines more comfortable for countries that administered the questionnaires to one population of interest (ISCED level 02) in one language. These countries included Chile, Iceland, Japan and Turkey.

6.3. Contacting centres and within-centre sampling procedures

The IEA's sampling team sent each NPM a selected centre sample based on the sampling frame the NPM had already submitted (see Chapter 5 for more details on centre sampling). To achieve the highest possible participation rates at the centre level, the sampling team sampled two replacement centres (assuming such centres were available) in addition to each originally sampled centre.

Once NPMs received their sample, the national study centres began contacting the designated centres. National study centres usually adhered to the technical standards and only contacted the first replacement centre if the originally sampled centre declined participation. If this replacement centre also refused participation, NPMs approached the second replacement centre.

One participating country (Germany) applied another process to achieve the required participation rates, namely contacting some replacement centres while the response from the original centre was still pending. If the original centre and the replacement centre both agreed to participate in the survey, the replacement centre was regarded as a national centre. However, because this procedure represented a deviation from technical standard 5.5, national centres were not included during computation of the participation rates that was conducted as part of the data adjudication process. Further mention of the adjudication procedure appears in Chapter 10.

National study centres that administered the survey in both populations of interest had to handle two samples in parallel, which meant they had to plan their resources carefully to meet the requirements of the multiple tasks.

Most of the participating countries asked each centre to nominate a centre co-ordinator to take responsibility for carrying out all survey-related tasks within the centre. In many cases,

NPMs reported that centre leaders took over the role of centre co-ordinator. One country reported that staff members took on this responsibility. One national study centre hired an external agency to complete the tasks. Close co-operation between centre co-ordinators and national study centres was crucial during all steps of staff listing, staff sampling and survey administration. To facilitate smooth communication, some countries established hotlines.

6.3.1. Identification number, listing forms and tracking forms

Information about staff members was gathered through the listing form and tracking form (see Annex F for more information on these two forms). National study centres used WinW3S to produce the forms. This software package provided hierarchical four-digit identification numbers that uniquely identified the sampled centres within each participating country. This number was also the identification code assigned to the person answering the leader questionnaire. Staff identification codes were derived from the centre codes by adding two additional digits at the end of the centre identifier, a process that created a hierarchical link between centres and staff members.

In accordance with the instructions in the centre co-ordinator manual, centre co-ordinators listed each eligible staff member by his or her name, year of birth, gender, staff role and leader role. One country reported that data protection/confidentiality laws and rules prevented it from providing the full names of the staff members. Two other countries used only numbers or codes instead of names due to confidentiality concerns.

The ISC pre-defined six staff roles at the international level: centre leader only (no pedagogical work), teacher, assistant, staff for individual children, staff for special tasks, and intern. National study centres introduced several additional country-specific staff roles to better reflect the situations in their country from the perspective of the questionnaire respondents. National study centres sometimes adapted these categories to encompass 12 and more staff roles. However, these were mapped and collapsed into the six international categories after data collection for international analysis and reporting purposes.

At times, centre co-ordinators found classifying staff members into the staff roles a demanding task, requiring close co-operation between the centre co-ordinators and their respective national study centres.[3] Although TALIS Starting Strong 2018 targeted ISCED level 02 staff members and U3 staff members, not every staff member was within scope. Out-of-scope staff included staff members who were volunteers who occasionally came into the centre for a special activity, auxiliary staff such as cleaners and cooking staff, medical and therapeutic staff such as speech therapists, occasional or emergency staff who filled in on a temporary basis, and staff on long-term leave.

Most countries reported that centre co-ordinators experienced difficulties in listing staff who were within scope of the survey. One country found it difficult to distinguish between staff working with children from birth to three years old from those working with children age three years and older. However, in all cases, the national study centres followed up via phone or e-mail and supported the centres in listing the staff correctly.

After the sampled centres returned their listing forms to their national study centre, the national study centre entered the information on those forms into WinW3S, enabling the software to draw a random sample of eight staff members per centre.[4] After completion of the within-centre sampling, WinW3S created tracking forms that listed all sampled staff members. The national study centres sent these tracking forms together with the questionnaires to the sampled centres so that centre co-ordinators knew which staff members should receive the questionnaires.

The tracking forms were also used to monitor the participation status of the sampled staff members and therefore included staff names (if used), ID, year of birth, gender and the questionnaire mode (online or paper). Each tracking form furthermore contained a column that allowed centre co-ordinators to document the questionnaire return status. Two countries said they did not use the tracking forms but used other means of obtaining the information. However, they gained approval to do this from the ISC beforehand.

The national study centres sent copies of the tracking forms, which included staff IDs but not staff names, to the ISC together with the survey data. Because the centre co-ordinators removed names on the tracking forms prior to submission, all names were kept confidential. Annex F contains a blank listing form and tracking form.

6.3.2. Assigning materials to staff and centre leaders

During TALIS Starting Strong 2018 in each participating country, the WinW3S software assigned the questionnaires to the centre leader and the staff members while the national study centre prepared one package of materials for each centre and sent it to the centre co-ordinator. This package contained all necessary questionnaires for the centre if that centre was administering the questionnaires in paper mode. Those centres conducting online administration received the cover letters containing the required access information. All packages also contained the tracking form and any other relevant briefing materials. To address confidentiality concerns, several countries chose to provide staff with pre-paid envelopes so they could return their questionnaires directly to their respective national study centre rather than hand it back to the centre co-ordinator.

The leader in each sampled centre was asked to complete one leader questionnaire. Additionally, the WinW3S software assigned the staff questionnaires within the same sampled centres. The centre co-ordinator handed out the questionnaires to the staff members according to the tracking form and the name of the staff member printed on the questionnaire's cover page (unless data protection regulations prevented the use of names).

Figure 6.1 outlines the different responsibilities of the national study centres and the centre co-ordinators with regard to the correct assignment of questionnaires to staff members.

6.4. Administering the questionnaires and conducting national quality observations

The ISC defined the international data collection windows as a guiding frame for countries. However, to offer flexibility for adjustments to country-specific needs (e.g. centres' closing times, holiday seasons), each participating country could select its own data collection period within the internationally prescribed time period. For Northern Hemisphere countries, the prescribed time for the main survey data collection lasted from 1 March until 31 May 2018; for the Southern Hemisphere countries, it was between 1 August and 31 October 2018. National study centres were requested to discuss any deviations outside of these periods with the ISC to receive the ISC's approval. Table 6.2 documents the data collection windows as they were originally planned, as well as changes to them (an empty cell in the columns relating to the actual data collection period means that the actual date was identical to the planned date). The table also highlights the dates that fell outside the internationally prescribed time periods. The data collection windows ranged from 10 days to about four and a half months.

Figure 6.1. Responsibilities of national study centres and the centre co-ordinator during survey administration

The data collection period for Denmark, Germany, Israel U3 and Korea deviated from the internationally prescribed time periods for different reasons. During the time scheduled for data collection, a public strike in Denmark influenced willingness to take part in the survey, so the ISC agreed Denmark should prolong its data collection. The data collection extension for Germany and Korea gave a considerable boost to the participation rates. Israel's split system, in which different authorities and settings provide education and care for children

under three years of age and for children at ISCED level 02, meant Israel required additional time to implement the questionnaires in the settings for the younger population of children.

Table 6.2. Originally planned and actual data collection windows for ISCED level 02 and U3

Participating country	Planned start of data collection window	Planned end of data collection window	Actual start of data collection window	Actual end of data collection window
Chile	01-08-18	31-10-18	23-07-18	
Denmark	19-03-18	31-05-18		08-07-18
Germany	15-03-18	15-07-18	15-03-18 08-10-18[a]	15-07-18 12-11-18
Iceland	15-03-18	19-06-18		
Israel ISCED level 02	22-04-18	17-05-18		
Israel U3	14-06-18	28-06-18		29-07-18
Japan	11-10-18	09-11-18		
Korea	01-06-18	15-07-18	09-07-18	23-11-18
Norway	06-03-18	06-06-18		
Turkey	21-05-18	31-05-18		

Note: [a] To increase participation rates, Germany's national study centre re-opened their data collection for a few weeks in October/ November 2018.

During the national data collection periods, centre leaders and staff were free to fill in the questionnaires whenever they preferred. The overall target was 100% within-centre participation. A centre was considered to be a participating centre if at least 50% of the sampled staff returned their completed staff questionnaires.

To maintain high survey standards, the ISC expected the national study centres to run a national quality control programme. The ISC provided a manual outlining the tasks to be carried out by the national quality observers (NQOs). However, NPMs could elect to arrange their own programme. NQOs could perform their national quality observations partly on the phone and partly by visiting participating centres. The ISC required each NQO to personally visit at least 10 centres (for more details, see Section 7.6. in Chapter 7 of this technical report).

Most of the participating countries chose national study centre staff to carry out the NQO programme. The remaining countries appointed external personnel, such as researchers involved in early childhood education and care or representatives of government agencies involved in education. After the TALIS Starting Strong 2018 main survey had been administered, NPMs reported the outcomes of the national quality control work in the survey activities questionnaire. Chapter 7 provides a discussion of these outcomes. The IEA Amsterdam was responsible for organising the international quality observation work; their role in this regard is also outlined in Chapter 7.

6.5. Monitoring the online questionnaires

Monitoring the administration of the survey was a demanding task for centre co-ordinators, especially if data protection laws prevented them from using the online monitor of the OSS software or if staff returned their paper questionnaires directly to their national study centre. In these instances, centre co-ordinators relied on national study centres to inform them of the need to follow up pending questionnaires.

National study centres monitored completion of the online questionnaires (see Chapter 8 for more details) and communicated the completion status to centre co-ordinators. Countries could manage this procedure according to their needs. Most national study centres preferred regular e-mail and/or telephone exchanges between the centre co-ordinators and themselves. Other national study centres either contacted respondents directly using e-mail or instant messaging, prepared reminder leaflets for staff and centres, asked union members to call centres, or created national TALIS Starting Strong 2018 websites where centre co-ordinators could log on individually to access all necessary information. Some national study centres granted centre co-ordinators access to the monitor so that they could organise the follow-up procedures themselves.

The monitoring work also included a participation-rate estimation tool that kept national data managers up to date on their current participation rate according to the already returned paper questionnaires or submitted online questionnaires.

After the main survey had been administered, each national study centre exported the questionnaire data availability status from the OSS monitor. National study centre personnel then imported this participation information, as well as the participation information from the DME with respect to administration of the paper questionnaire, into WinW3S, a practice that enabled the national study centres to verify the participation status of each sampled person. The ISC told each national centre that it was mandatory for them to verify participation before they submitted data to the TALIS Starting Strong 2018 International Consortium.

6.6. Receipt of materials and preparations for data entry

The major tasks for NPMs immediately after administration of the TALIS Starting Strong 2018 main survey included retrieving and collating the materials from centres and verifying their integrity. On receiving survey materials from the centres, NPMs:

- verified that all identification numbers on all questionnaires were accurate and legible
- checked that the participation status recorded on the tracking form matched the availability of the online questionnaires and the paper questionnaires
- followed up those centres that did not return all the survey materials or for which forms or paper questionnaires were missing, incomplete or otherwise inconsistent.

National study centres recorded all necessary information about centres, leaders and staff, including the return status of the questionnaires, in WinW3S. NPMs then organised the paper questionnaires and corresponding forms for data entry (see Chapter 8).

Notes

1. The international study centre (ISC) prepared software packages for sampling, online data collection, data entry and data processing to fit the needs of TALIS Starting Strong 2018, and it asked all participating countries to use only the IEA-provided software to perform the survey. For various reasons, the ISC agreed that Israel did not need to use the IEA eAssessment System.

2. The ISC used the survey activities questionnaire (SAQ), administered after the field trial and again after the main survey, to collect NPMs' feedback on their experiences during administration of TALIS Starting Strong 2018. The ISC considered the outcomes of the field trial activities questionnaire in terms of improvements to the main survey's procedures and the software in use. The outcomes of the main survey questionnaire are incorporated into this chapter.

3. In Iceland, the distinction between teachers and assistants was, in many cases, unclear to staff, centre leaders and centre co-ordinators. The same person was frequently grouped inconsistently across the listing form, the staff questionnaire and the leader questionnaire. The ISC consequently decided not to distinguish between these two staff categories when reporting the results for Iceland in the TALIS Starting Strong 2018 international report.

4. National study centres were given the opportunity to sample more than the recommended eight eligible staff members per centre, if desired.

Chapter 7. Quality control of data collection

To ensure high-quality data from the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018, the survey's international consortium developed and implemented a quality control programme designed to monitor participating countries' compliance with predefined procedures and standards. The quality control programme consisted of quality control measures at national and international levels as well as a questionnaire designed to elicit information from the national study centres. As summarised in this chapter, the findings of the quality control programme show that the participating countries generally complied with the procedural standards established for administration of the survey.

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

7.1. Overview

The Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 set out to gather internationally comparable data from leaders and staff members of early childhood and care centres. To ensure the international comparability of the data, the TALIS Starting Strong 2018 International Consortium stipulated standardised procedures for administering the survey instruments and collecting data. These procedures were issued to the participating countries via the TALIS Starting Strong 2018 technical standards (see Annex B) and the survey operations procedures units.

The consortium designed a comprehensive quality control programme designed to monitor the extent to which the participating countries complied with the procedural standards. The quality control programme consisted of the following components:

1. *The international quality control programme:* IEA designed and managed standardised quality control measures at the international level and recruited independent international quality observers (IQOs) whose role included visiting a subsample of the participating centres and interviewing the people responsible for co-ordinating survey activities in the centres.
2. *National quality control programmes (the responsibility of the national study centres):* The consortium required all participating countries to conduct quality control at the national level. The IEA aided this process by providing template documents that facilitated setting up the national quality control measures.
3. *Survey activities questionnaire:* The ISC complemented the quality control activities at the national and international levels by designing and administering a survey activities questionnaire developed to elicit information about the national project managers' (NPMs) experiences when preparing for and conducting the TALIS Starting Strong 2018 data collection.

This three-tier approach to quality control allowed the consortium to gather information about participating countries' compliance with standardised procedures from different perspectives. The different quality control components were also an essential source of information about the quality of the survey's data.

After the survey had been administered, the IEA evaluated the findings of the international quality control programme, paying special attention to any issues mentioned in those findings that could affect the quality of the data. This evaluation also included a review of NPMs' responses to the survey activities questionnaire. The information gained from the review subsequently informed data and sampling adjudication and the country-level rating of the data (see also Chapter 10).

This chapter provides an overview of the different components of the quality control programme, and focuses in particular on the design and approach of the international quality control activities and the survey activities questionnaire. The chapter also summarises the findings of the international quality control programme and evaluates the NPMs' answers to the survey activities questionnaire, which included questions on the national quality control activities in the participating countries.

7.2. The international quality control programme: the role of international quality observers

The IEA developed and implemented a standardised international quality control programme to monitor and document the survey administration and data collection activities in the participating countries.

7.2.1. Selection and training of international quality observers

An essential element of the programme was the recruitment and appointment of independent international quality observers (IQOs) who were tasked with monitoring the survey administration and data collection activities in the participating countries. Prior to the survey administration period, the IEA asked each NPM to nominate two individuals who could serve as an IQO in their country.

To ensure the independence of the quality control programme, the nominees could not be people working for the national study centre or people with personal ties to the NPM. In addition, NPMs were required to nominate candidates who were familiar with early childhood education and care (ECEC) environments and/or the day-to-day operations of childcare centres. Candidates also needed to be literate users of information and communications technologies and fluent in English and the language(s) in which the survey would be administered in their respective country.

After each NPM had nominated two candidates, the IEA selected one individual per country to serve as IQO[1] and then asked the appointed candidates to attend a mandatory one-and-a-half-day training seminar that took place at the IEA office in Amsterdam, The Netherlands, on 5 to 6 February 2018. During the training seminar, the IEA informed the IQOs about their role and responsibilities in relation to TALIS Starting Strong 2018 and instructed them on how to monitor the data collection in their respective countries. The IEA also provided the IQOs with the following materials to support their quality control work:

- the *International Quality Observer Manual*, outlining the IQOs' role and responsibilities
- the centre visit record (template), a standardised questionnaire guiding the IQOs' interviews with centre co-ordinators and their observations during the centre visits
- the translation verification report (template), guiding the IQOs' feedback on the translation verification process and the quality of the translated *TALIS Starting Strong 2018 Centre Co-ordinator Manual*(s)
- the NPM interview outline and question template, structuring the IQOs' visits to the NPMs and the process of selecting centres for the quality control programme
- the international version of the *TALIS Starting Strong 2018 Centre Co-ordinator Manual*
- the TALIS Starting Strong 2018 survey operations procedures units
- the TALIS Starting Strong 2018 technical standards
- the TALIS Starting Strong 2018 glossary, providing guidance for translators on the meaning of certain terminology used in the questionnaires
- the international source version of the survey instruments

- other supporting documents, including the centre visit tracking form, the centre visit travel form, travel expense forms, a checklist for collecting materials from the NPM, confidentiality agreements for assistants, and confidential login details the relevant individuals needed to enter data from the centre visit records into the IEA Online Survey System (OSS).

To manage the expected workload and to complete their responsibilities within the survey administration period, IQOs had the option to recruit assistants. The IQOs were responsible for training assistants and co-ordinating the international quality control activities within the respective country. Assistants were not expected to communicate directly with the IEA; instead, all communication with the IEA was channelled through the IQOs.

7.2.2. International quality observers' responsibilities

As part of the international quality control programme, IQOs had three main responsibilities. Their first responsibility involved visiting the national study centre to interview the country's NPM. During the meetings with NPMs, IQOs collected national survey materials (e.g. translated survey instruments and the translated centre co-ordinator manual) and selected 20 centres to be visited.

To determine which centres to visit, IQOs and NPMs drew a subsample from the national samples of participating centres. To ensure the feasibility of the centre visits, centres were excluded on the basis of practical constraints such as travel distance, time and cost. In countries that issued the survey to both populations of interest (i.e. ISCED level 02 and U3), the number of centre visits was split evenly between the two (i.e. 10 centre visits per population of interest).[2] The IEA provided IQOs with the NPM interview outline and question template to help them structure their visits to the national study centres.

The second responsibility required the IQOs to visit the 20 pre-selected centres to interview the centre co-ordinators and to observe the administration of TALIS Starting Strong 2018. The IEA provided IQOs with the centre visit record that they were to use for each centre visit. The centre visit record included questions asking the centre co-ordinators about preparations for administering the survey and the data collection processes in their respective centres. Additional questions asked the IQOs about their general observations and impressions of each centre co-ordinator's work. After completing all 20 centre visits, IQOs were required to enter the gathered information into the IEA Online Survey System.

Table 7.1 shows the number of visited centres per population of interest in the participating countries.[3] The IQOs conducted one centre co-ordinator interview in each centre. Section 7.3 of this chapter contains a detailed discussion of the IQOs' findings, as documented in the centre visit records.

Table 7.1. Number of centre visits per population of interest and participating country

Participating country	ISCED level 02	U3
Chile	20	0
Denmark	10	10
Germany	10	10
Iceland	19	0
Israel	10	10
Japan	20	0
Korea	20	0
Norway	10	10
Turkey	20	0

The IQOs' third responsibility was to provide feedback on the process used to verify the translations of the national survey instruments and of the centre co-ordinator manual(s). IQOs used the translation verification report to document and summarise their findings. They also compared the translated centre co-ordinator manual(s) against the international source version and documented all differences between the two. Section 7.4 of this chapter provides a summary of the IQOs' feedback in the translation verification reports.

Throughout their quality control work, IQOs were required to report their progress to the IEA and to advise them of any issues or observed irregularities. They were also expected to send two sets of deliverables to the IEA office in Amsterdam. The sets included the following materials and documents:

- a copy of the completed NPM interview outline and question template
- a copy of the completed translation verification report for each administered language
- copies of the 20 completed centre visit records
- 20 listing forms, as collected from the NPMs (one for every centre visited)
- 20 tracking forms, as collected from the NPMs (one for every centre visited)
- a copy of the translated national version(s) of the centre co-ordinator manual
- copies of all TALIS Starting Strong 2018 instruments administered in the respective countries
- confidentiality agreements signed by the IQOs' assistant(s) (if applicable).

7.3. Centre co-ordinator interviews: information from the centre visit records

The central part of the international quality control programme consisted of the centre visits as described above. During these visits, IQOs interviewed the centre co-ordinator responsible for administering the survey in the respective centre. IQOs used the centre visit record to document the centre co-ordinators' responses as well as any additional IQO observations. Prepared by the IEA, this questionnaire guided the IQOs' centre visits and interviews with the centre co-ordinators and ensured that the international quality measures followed standardised procedures. The centre visit record included open-ended and closed-ended questions covering four different aspects of centre co-ordinators' involvement in administering the survey. Together with the IQOs' general observations, the centre co-

ordinators' responses to these questions shed light on each country's compliance with the standardised procedures.

In keeping with the structure of the centre visit records, the subsequent paragraphs summarise the findings of the international quality control programme and thereby provide information on the extent to which the participating countries complied with the standardised procedures. Where appropriate in the following subsections, the centre co-ordinators' responses and the IQOs' observations are compared against the procedural standards applicable to the survey administration and data collection activities.

7.3.1. Centre co-ordinators' background and training

The first section of the centre visit record addressed the centre co-ordinators' professional background and experience as well as the training they had received for their role in TALIS Starting Strong 2018 and the information they were given about that role.

Centre co-ordinators' position in relation to the centre

The TALIS Starting Strong 2018 technical standards provided guidance on but did not strictly define how the centre co-ordinators should be affiliated with the centre(s) they were responsible for. However, the standards suggested that the centre co-ordinators could be active or retired centre leaders, staff members, or administrative or managerial personnel in the centre. Alternatively, members of the national study centres or individuals from external agencies could serve as centre co-ordinators.

Of the interviewed centre co-ordinators, 88.8% reported that their role was internal to the centre. Of those centre co-ordinators who were employees at the visited centres, 78.6% were centre leaders and the remaining 21.4% were other members of staff. Eighty-six per cent of the centre co-ordinators stated that they served as centre co-ordinator in only one centre. In Chile, all centre co-ordinators worked for an external agency and each co-ordinator was responsible for up to 26 centres.

Centre co-ordinators' previous experience

While previous experience of co-ordinating survey activities was not a requirement for individuals to act as centre co-ordinators, the technical standards recommended that centre co-ordinators should have experience with the work carried out in ECEC centres. The IQOs found that 25.7% of the centre co-ordinators had previously served as co-ordinators for other surveys or assessments at the national or international level, which meant TALIS Starting Strong 2018 represented the first time 74.3% of the respondents had filled the position of centre co-ordinator.

Centre co-ordinator manual

In order to support the centre co-ordinators in their work, the ISC provided the national study centres with a centre co-ordinator manual. The national study centres subsequently translated the manual into the national language(s) used for administering the survey. They also adapted the manual's content to fit the national context. The manual was designed to guide centre co-ordinators through their tasks and to provide information on and explanations of the survey administration procedures. As reported by the IQOs, the national study centres of seven of the nine participating countries used the manual to instruct centre co-ordinators. In the other two countries, centre co-ordinators received

instructions about their role in the form of training seminars or via a series of e-mails from their national study centre.

The IQOs reported that 90.6% of the interviewed centre co-ordinators who received the centre co-ordinator manual said that the manual worked well for them. Suggestions for improvement varied from country to country and included simplifying the explanations provided, shortening the manual, and adding more specific information about the survey administration and the data collection activities.

Centre co-ordinator training and comprehension of the survey

The consortium strongly recommended that national study centres organise some form of training session for the centre co-ordinators. The national study centres in Chile, Israel (U3), Korea, Norway, and Turkey arranged training sessions especially designed for centre co-ordinators. Of the centre co-ordinators in these countries, 62.2% attended a session offered by their national study centre. Some centre co-ordinators stated that they were too busy to attend the training session, did not know about the training session, or could not participate due to personal circumstances. In some countries, namely Denmark, Germany, Iceland, Israel (ISCED level 0.2) and Japan, the national study centres did not arrange training sessions but used alternative ways to prepare the centre co-ordinators for their role (see section 7.5.1 for more information).

Of those centre co-ordinators who did not attend a training session offered by their national study centre, 90.2% made other arrangements to prepare themselves for their role as centre co-ordinator. These arrangements included contacting the NPM to receive information about the survey, reading the centre co-ordinator manual, and learning about the survey administration and questionnaire return procedures prior to distributing the instruments.

A prerequisite for the successful administration of TALIS Starting Strong 2018 was having centre co-ordinators with a good understanding and thorough knowledge of the survey administration procedures. Table 7.2 summarises how centre co-ordinators evaluated their own understanding of different aspects of the survey administration. A large majority of respondents stated that they had no difficulty understanding the purpose of the survey and the survey procedures.

Table 7.2. Centre co-ordinators' comprehension of survey procedures

Aspect of survey administration	Did you have any difficulty understanding the survey procedures?	
	Yes (%)	No (%)
Purpose of the survey	7.3	92.7
Survey administration	5.6	94.4
Survey return procedures	5.6	94.4

7.3.2. Preparations prior to administration of the survey

Ensuring that the individual centres thoroughly prepared for the procedures involved in administering the survey was another crucial condition for the successful distribution of the survey instruments and the smooth collection of data. To ensure centre co-ordinators' preparedness, the ISC expected national study centres to maintain close contact with centre co-ordinators throughout the survey procedures and to respond to their questions and concerns. In most countries, centre co-ordinators were responsible for preparing the listing form(s) before the survey administration and for distributing the questionnaires and/or cover letters to the centre leaders and staff members listed on the tracking form(s).

Turkey was an exception in this regard because the national study centre provided the listing forms for all participating centres and distributed the survey instruments to the individual participants via e-mail.

Centre co-ordinators' communication with NPMs

As stipulated in the technical standards, comprehensive contact between the national study centres and the centre co-ordinators was necessary to ensure centre co-operation, proper within-centre sampling, and the smooth administration of the survey. The IQOs found that 88.8% of the centre co-ordinators experienced consistent communication with the NPMs, which allowed the centre co-ordinators to prepare administration of the survey in their centres in a timely and ordered manner. In addition, 86.6% of the centre co-ordinators stated that the NPMs and/or other personnel at the national study centres were responsive to their questions and concerns.

Preparing and using the listing and tracking forms

Preparing and accurately completing the listing forms and generating the tracking forms formed an essential component of the within-centre sampling process. Of the interviewed centre co-ordinators, 91.6% stated that they used the templates of the forms prepared by the ISC. Some centres did not use the templates but instead gained ISC approval for using alternative ways of documenting and tracking sampled participants. Of those centre co-ordinators who used the provided templates, 89.6% had no difficulties completing the forms.

The centre co-ordinators were also responsible for listing all eligible staff members on the listing forms. During the interviews, the IQOs asked the centre co-ordinators about the completeness of the forms used at the respective centres. In 91.1% of the visited centres, the centre co-ordinators stated that the listing form listed all staff members eligible to participate in the survey. In cases where incomplete forms were reported, the available information shows that the forms did not include those staff members who were soon to leave the centre, who were not selected to participate, or who did not want to participate in the survey. In 97.8% of the centres, as reported by the centre co-ordinators, the form listed only eligible staff members and no out-of-scope persons.

When the IQOs asked the centre co-ordinators about the proper use of the tracking form as an indicator of which people to distribute the survey instruments to, 86% of the co-ordinators said that they distributed the cover letters and/or questionnaires in accordance with the form. Most of the remaining centre co-ordinators were in Turkey, where centre co-ordinators were not responsible for distributing the survey questionnaires because the national study centre sent the surveys to the participants by e-mail.

Distributing the survey instruments

The timeline for distributing the survey instruments required national study centres to provide their centre co-ordinators with the questionnaires or cover letters one to two weeks before the survey administration period. When these materials arrived at the centres, centre co-ordinators were responsible for verifying the number of questionnaires or cover letters against the tracking form and for checking that the delivered questionnaires or cover letters were not defective and did not need to be replaced. Of the centre co-ordinators involved in overseeing paper administration of the survey, 92.1% reported that they did not have to arrange any replacement questionnaires, and 96.8% said they found no defective questionnaires and therefore did not need to replace any of the questionnaires during

administration of the survey. Of the centre co-ordinators responsible for the online questionnaires, 99.4% indicated that they did not have to arrange any replacement cover letters, and none of them detected defective cover letters requiring replacement during the survey administration.

The two different modes of administration meant that survey participants had to be sure they received the survey instruments in the expected format (online or paper). In regard to the paper administration, 96.9% of the centre co-ordinators said that the participants received the questionnaires in the expected format. One centre co-ordinator said that staff members who were expected to answer the paper questionnaire asked if they could complete the survey online instead.

In those centres where the survey was administered online, 80.6% of the interviewed centre co-ordinators confirmed that online participants received the questionnaire in the electronic format as expected. The remaining centre co-ordinators reported that they did not know if participants received the questionnaire in electronic format because they did not follow up with the participants after distributing the cover letters.

7.3.3. Data collection and survey administration

This section addresses the centre co-ordinators' activities during the data collection. It summarises the IQO findings with respect to the centre co-ordinators' interaction with individual survey participants and the confidentiality and data protection measures present in the visited centres.

Providing participants with assistance

Centre co-ordinators were not allowed to interfere with the participants' responses to the questions in the survey instruments. However, centre co-ordinators could assist any participants who were unsure about the survey administration or had any general questions about TALIS Starting Strong 2018. As shown in Table 7.3 survey participants asked centre co-ordinators questions about different aspects of the administration of the survey. For example, 22.3% of the centre co-ordinators said respondents asked them to clarify specific questions about the survey instruments and 1.1% of the co-ordinators said respondents asked them about linguistic errors they found in the questionnaires.

Table 7.3. Questions survey participants asked centre co-ordinators

Aspects of the survey administration	Did participants approach you to discuss any of the following?	
	Yes (%)	No (%)
Purpose of the survey	12.8	87.2
Survey return procedures	8.4	91.6
Clarification of any items	22.3	77.7
An error they spotted	1.1	98.9
Other questions about the survey	7.8	92.2
Any questions they could not answer	9.5	90.5

The IQOs also asked centre co-ordinators if they explained specific aspects of the survey administration procedures to staff members. Table 7.4, which summarises the IQO findings, shows that more than 90% of the centre co-ordinators explained particular aspects of the survey administration to the participants.

Table 7.4. Explaining the survey administration procedures

Aspects of the survey administration	Did you explain the following aspects to the participants?	
	Yes (%)	No (%)
Purpose of the survey	96.6	3.4
Estimated time to complete the survey	92.7	7.3
Confidentiality arrangements	92.2	7.8
Survey return procedures	91.6	8.4

Confidentiality and data protection

In anticipation of respondents' concerns about the use of their personal information and answers to the survey questions, the TALIS Starting Strong 2018 International Consortium advised the NPMs to ensure compliance with national confidentiality and data protection regulations. This section identifies what data protection measures were in place and whether centre co-ordinators dealt with the survey instruments in a confidential manner.

As documented in the centre visit records, 39.9% of centre co-ordinators reported that they had to adhere to particular data protection rules in their country. All interviewed centre co-ordinators in Denmark, Norway and Turkey stated that their countries had no particular data protection rules in place. In the remaining countries, the majority of the centre co-ordinators indicated that they had to adhere to such rules.

Of the interviewed centre co-ordinators, 83.2% stated that they distributed the survey instruments to the participants in a confidential manner. When IQOs asked the interviewed centre co-ordinators about access restrictions, 71.5% of the co-ordinators said that only they had access to the tracking form and listing form. Some centre co-ordinators said they kept digital copies of the lists on their computers and protected them with passwords. Physical copies of the lists were mostly kept in sealed envelopes and stored in locked closets or centre co-ordinators' desks. Of the co-ordinators in the centres that used paper questionnaires, 96.7% stated that only they had access to the completed questionnaires.

7.3.4. General impressions

The last two parts of the centre visit record contained questions designed to capture the centre co-ordinators' and IQOs' general impressions of the administration of TALIS Starting Strong 2018. The answers to these questions paint a more general picture of the data collection in the visited centres.

Staff attitudes towards the survey

The co-operation of staff members was paramount in ensuring the smooth administration of TALIS Starting Strong 2018. This section documents staff members' general cooperativeness and attitudes towards the survey. As reported by the IQOs, 95.5% of the centre co-ordinators described staff members as extremely or moderately co-operative. In addition, 80% of the centre co-ordinators stated that they endeavoured to increase the degree of co-operation when they encountered staff members who were less co-operative.

In 64.8% of the centres visited, the centre co-ordinators rated the overall attitude of staff members as positive. In 32.4% of the centres, the centre co-ordinators reported that staff members had a neutral attitude towards TALIS Starting Strong 2018. As reported by the IQOs, 73.7% of the centre co-ordinators stated that staff members generally felt comfortable answering the survey questions. However, in Korea, a particularly high

percentage (60%) of interviewed centre co-ordinators said staff members did not feel comfortable answering the questionnaire. In most of these cases, the centre co-ordinators reported that discomfort arose from practical constraints such as lack of time rather than emotional uneasiness.

In 19% of the visited centres across the participating countries, the centre co-ordinator stated that participants received encouragement to take part in the survey in the form of special instructions, motivational talks, or incentives. In most of these centres, the centre leader or other senior members of staff emphasised the importance of the survey in order to encourage staff members to answer the questionnaires.

Centre co-ordinators' impressions of the survey distribution process

As previously noted, the centre co-ordinators' key responsibility during administration of TALIS Starting Strong 2018 was distributing the questionnaires within the ECEC centres. According to the IQOs' reports, 79.3% of the centre co-ordinators stated that the survey distribution process went well and was without problems. Nineteen per cent of co-ordinators reported some minor issues with distributing the questionnaires but deemed the process satisfactory overall.

IQOs' reflections on the centre co-ordinators' performance

The centre co-ordinators' performance of their role was another aspect central to the success of TALIS Starting Strong 2018. Insight into the centre co-ordinators' performance was gained from the last few questions in the centre visit record. These questions asked the IQOs to document their impressions not only of the data collection but also of how well the centre co-ordinators' performed their role.

The IQOs reported that 92.7% of the centre co-ordinators were very prepared or somewhat prepared to administer the survey in the respective centres. The IQOs found that 86% of the centre co-ordinators were very willing to answer their questions about the survey administration activities. Less than 1% of the centre co-ordinators were unwilling to answer the IQOs' questions. The IQOs furthermore reported that the centre co-ordinators applied the survey procedures seriously and professionally in 95.5% of the visited centres. In most cases where centre co-ordinators were reported as less serious and/or professional, the IQOs indicated that the centre co-ordinators' overall workload prevented them from fully concentrating on administering TALIS Starting Strong.

7.4. Translation and manual review: information from the translation verification report

In addition to completing the centre visit records, the IQOs completed translation verification reports for each language in which the survey was administered. The completed reports contained the IQOs' feedback on the translation verification process, the quality of the translated survey instruments, and the content and translation of the centre co-ordinator manual.

As described in Chapter 4, all instruments administered during TALIS Starting Strong 2018 underwent translation verification. As part of the international quality control programme, the IEA asked IQOs to review the national survey instruments translated into the administered languages. The review included not only a numerical evaluation of each NPM's decisions about verifier feedback and interventions as documented in exports from the IEA eAssessment Translation System (hereafter Translation System), but also a

verification of the implementation of these decisions in the final survey instruments (i.e. the questionnaires used in the country). The rationale for asking the IQOs about the translation verification process was to gather feedback from independent native-language speakers and to reveal possible and otherwise unknown deviations from the standardised procedures.

In general, the number of verifier comments on major translation issues varied from language to language, with the range extending from 4 to 52. The NPMs accepted most of these comments, and the translations were adapted as suggested by the verifier. For all language versions of the survey instruments, the majority of the NPMs' reactions to the verifier comments were implemented in the final survey instruments in accordance with the documentation in the Translation System.

The second component of the translation verification reports addressed the overall utility of the verifiers' feedback and the quality of the final survey instruments. Several open-ended questions enabled the IQOs to summarise their qualitative evaluation of the translation verification process. According to the IQOs, the verifier comments were appropriate and useful for all language versions, while the translation verification process improved the quality and comprehensibility of the translated survey instruments.

The translation of the centre co-ordinator manual into the languages of administration was not included in the translation verification process described in Chapter 4. The national study centres were responsible for translating the manual and adapting it, where they deemed necessary, to fit the national context. Any deviations from the international source version of the manual required ISC approval. To determine whether the centre co-ordinator manual(s) used in each participating country deviated from the international source version in content or quality, the last part of the translation verification reports asked the IQOs to document any differences between the international manual and the translated national version(s).

Seven of the nine participating countries used the centre co-ordinator manual, which was translated into nine languages in total. The national study centres of the two countries that did not use the manual instructed centre co-ordinators via e-mails or at conferences held before data collection commenced. Most countries that used the manual template did not change its content at all or only minimally adapted the instructions to the national context (e.g. centre-based versus home-based settings).

Some countries that used the manual template adapted the centre co-ordinator instructions to the specific survey procedures at the national level. For example, the national study centre in Chile added a standardised protocol for contacting centres in order to account for the fact that centre co-ordinators were responsible for several centres. The national centre in Turkey changed the instructions about the within-centre sampling because it provided listing forms for all participating centres.

7.5. The NPMs' perspective: information from the survey activities questionnaire

The survey activities questionnaire gathered information on the national survey implementation from NPMs in all participating countries. Because Israel had two national study centres, one for ISCED level 02 and one for its U3 settings, there were also two NPMs, both of whom completed the survey activities questionnaire, thus yielding a total sample of 10 NPMs. The questionnaire contained nine content sections[4] featuring more than 100 questions in total. Questions focused on obtaining information from the national study centres about all survey-related activities and the extent to which all relevant parties had followed the procedures and guidelines. The questionnaire also gave NPMs an

opportunity to provide valuable feedback about all aspects of survey administration, including survey procedures and manuals, guidelines, support materials and software. The questionnaire was prepared by the ISC, who then sent it to all NPMs in May 2018 and asked them to complete it on line. All NPMs answered the questionnaire between June 2018 and January 2019. The following subsections present the results of this survey.

7.5.1. Contacting ECEC centres and preparing centre co-ordinators

First contact with the sampled ECEC centres was often made by the NPM together with other members of the national study centre (five countries). While respective ministries of education made the first contact in five countries, three of them did this in combination with the NPM and the national study centre staff. Three countries contracted staff from an external agency to co-ordinate contact with the centres.

Although the overall participation rates for TALIS Starting Strong 2018 were high for most of the participating countries, most NPMs (8 out of 10) said they experienced difficulty convincing centres to participate. The most common reason across the national study centres was lack of time to co-ordinate the study within the sampled centres combined with lack of time to respond to the survey. Strategies to overcome centres' reluctance to participate included multiple follow-up attempts via different communication channels (e-mail, phone calls), with those efforts directed towards explaining the importance of the study and the participation of staff in it and expressing ministry support for the survey. The consortium extended the data collection period in two countries (Germany and Korea) in an effort to improve their overall response rates. Participation rates show that, in most of the countries, these strategies were successful in overcoming difficulties and convincing centres to take part in the survey

The people nominated to fill the role of centre co-ordinator were generally easy to identify, and NPMs said they had no problems in appointing them. Eight NPMs reported that the centre leaders usually acted as centre co-ordinators; however, three of them stated that other staff members took on this role as well. In one country, mainly staff members fulfilled the centre co-ordinator tasks and one other country asked an external agency to appoint the centre co-ordinators.

One important challenge during TALIS Starting Strong 2018 concerned the definition of in-scope ECEC staff members. It was not always clear who the centre leader was, and which person should therefore complete the leader questionnaire. Three NPMs reported that they experienced difficulty explaining, to the centre co-ordinators, who were staff within the context (scope) of the study and three NPMs reported difficulty explaining how to accurately identify the centre leader.

All NPMs in the participating countries said they found adapting the centre co-ordinator manual to their national context and needs a relatively easy task. Adaptations varied and included excluding not applicable sections, shortening the manual or splitting the manual into smaller pieces so that centre co-ordinators would find it easier to handle. Two countries did not use the manual as such but shared the information via e-mails or at conferences prior to the data collection.

Because the co-ordinators played a key role within the survey, half of the national study centres provided them with formal training, while the other half used different communication channels, such as phone calls and written instructions. National study centres also used e-mails and phone calls to explain single tasks at the appropriate times throughout the survey process.

7.5.2. Within-centre sampling

All participating countries used the IEA Windows Within-School Sampling Software (WinW3S) provided by the ISC.

Only one country used pre-populated listing forms that contained information from that country's ministry of education. On completing the lists, the NPM sent them to the participating ECEC centres for verification. Half of the national study centres used means other than the paper/electronic forms created by WinW3S to list and track staff members. These means included online forms, e-mails and special software. However, all of these means conveyed the same type of information and applied the listing rules requested by and subject to the approval of the ISC.

Due to the complexity of the listing procedure within the context of ECEC, all but one country reported some difficulties during the listing process. These generally related to uncertainty over which centre staff belonged to the staff definition or the population of interest. If a participating centre detected and reported listing mistakes, the national centre addressed the mistakes via phone calls and ensured the errors were corrected before the survey administration process began.

While consideration was also given to data protection and confidentiality laws, only one country reported restrictions in this regard and these related to using staff members' full names on the listing and tracking forms and the national questionnaires. This problem was resolved by the country using sequential numbers on the forms.

7.5.3. Preparing questionnaires

Preparation of the questionnaires involved several steps, as outlined in Chapter 4. Mindful that translating and adapting survey instruments is a difficult exercise during any cross-national study, the ISC supported the process by bringing a structured approach to the instrument preparation process. The process included individualised translation verification schedules, additional support materials (e.g. national adaptation forms) and documentation guidelines.

Overall, all countries said that the information provided in the related survey operations procedures units was sufficient, while most said their NPMs found documentation of the national adaptations in the national adaptation forms clear and straightforward. However, one country said that the process involved in reaching agreement on national adaptations was more time-consuming than anticipated.

All countries found the information about the translation processes, provided in the related survey operations procedures units, sufficient. In addition, all countries regarded the provision of the existing field trial translations as a starting point for the main survey translations useful and considered the feedback from the international translation verifiers as very or somewhat useful.

After completion of the adaptation and translation verification processes for the paper instruments, the ISC prepared online questionnaires for all but the two countries that administered their questionnaires on paper only. Once the questionnaires were online, the ISC asked NPMs to review these questionnaires to ensure that no errors had occurred during the transfer from paper to online. The ISC then verified the layout of the paper and online instruments. All countries reported that the information provided for this process was sufficient. Two countries said they detected errors during the printing process but managed to correct these before handing the questionnaires over to the respondents.

7.5.4. Online administration

The majority of countries administered the questionnaires on line to staff and centre leaders. All but one country reported no technical problems during the online data collection. Although this country detected the issue directly after the start of the data collection period, the ISC was able to solve it immediately, which meant no data were lost. A few countries reported problems related to correctly logging in (e.g. respondents mistyping access information, an existing firewall blocking access through the link) or unstable internet connections.

To monitor the online data collection process and participation progress on a daily basis, the ISC provided a software tool called the IEA Online Survey System Monitor. This tool proved very helpful, and participants used it several times a week. The NPMs in all countries that collected data online reported that they used the tool several times a week or daily (four countries), approximately once a week (two countries) or less than once a week (one country, but this country collected the data primarily on paper).

7.5.5. Paper administration

While most data were collected on line, two countries, as already noted, administered exclusively paper instruments in the participating ECEC centres. Two other countries used a mixed-mode approach, that is, paper and online instruments. The four countries that administered paper questionnaires said they encountered no specific problems with this administration mode. When these countries were asked about the process they used to return the completed questionnaires, one of them said it asked each respondent to send back his or her completed questionnaire in a pre-paid envelope (i.e. one envelope per respondent). The other three countries said they asked the centre co-ordinators to collect the completed questionnaires and return them to their respective national study centre. In some instances, the national study centres in these countries had to remind centre co-ordinators to return the questionnaires.

7.5.6. Challenges for staff members filling out the questionnaires

Because TALIS Starting Strong 2018 was the first implementation of this study and because the questionnaires used for it were adapted from the TALIS 2018 context, concerns arose as to whether ECEC centre staff members would experience difficulties when responding to the questionnaires. Consequently, a question in the SAQ asked NPMs if they became aware of any such situations during survey administration. Four NPMs said some staff members needed help filling out the questionnaires because of challenges understanding some of the language in them. The questions staff asked related to concepts unfamiliar to them within the context of their work and to questions where they had to estimate percentages. The people who responded to these questions varied between the centre co-ordinators, colleagues, the centre leader and national study centre staff.

7.5.7. Monitoring participation

The ISC provided the national study centres with a software tool (the IEA Participation Rate Estimator) to help them track participation rates during data collection. NPMs sent the outcomes of the estimator to the ISC on a relatively regular basis, thus enabling the ISC to closely monitor participation rates in all participating countries and to contact national study centres if rates seemed too low or did not evolve fast enough. In some cases, the ISC recommended that national study centres apply specified strategies to improve participation rates (e.g. contacting specific centres to help them pass the 50% barrier).

7.5.8. Manual data entry and submission

All countries that administered the questionnaires on paper entered the responses to the paper questionnaires manually. One country used national study centre staff to do this task, two countries hired an external agency to do it, and the fourth country applied a combination of both. All NPMs of these countries said they used the provided IEA Data Management Expert (DME) software to enter the data manually and did not experience unacceptable levels of errors during the data entry process (see Chapter 8 for further details).

7.6. National quality control

In addition to the international quality control programme overseen by the IEA, national study centres were required to conduct quality control at the national level. The IEA helped the national study centres set up national quality control procedures by providing NPMs with a national quality observer manual that outlined the purpose and key components of quality monitoring during TALIS Starting Strong 2018. Countries could adapt the manual where necessary or choose their own individual design for national quality control. The IEA also provided national study centres with a centre visit record that they could use as part of their national quality control measures.

Eight of the 10 national study centres conducted national quality control programmes, with the number of visited ECEC centres ranging from 3 to 23. The two countries that did not implement the national quality control programme as suggested by the ISC maintained close contact with centre co-ordinators to ensure that they received all materials and followed the requested procedures.

Seven of the eight national study centres that conducted quality control at the national level used the national quality observer manual and centre visit records provided by the IEA. In most cases, the national study centres adapted the templates to their national context by adding or removing questions.

Notes

1. One exception was made for Israel, where two IQOs were selected. One IQO was responsible for visiting ISCED level 02 centres, while the other IQO was responsible for visiting U3 centres.

2. See Chapter 1 for an overview of the populations of interest in each of the participating countries.

3. Due to unforeseeable circumstances, Iceland managed only 19 centre visits.

4. The nine sections were centre contact; within-centre sampling; documenting and implementing national adaptations; translating instruments; preparing instruments and layout checks; administering questionnaires; manual data entry and submission; national quality observation; and miscellaneous.

Chapter 8. Creating and checking the international database

This chapter describes the approach and strategies used to create the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 international database (IDB). It describes the data entry and verification tasks carried out by the national study centres and the exchange of data and documentation between these centres and the international study centre at the IEA. It also describes the integration of data from the paper and online administration modes and the individual steps of the data editing and database creation procedures implemented there, including the detection and resolution of inconsistencies in the data. Finally, it gives an overview of the interim data produced and the steps that all involved centres took to confirm the accuracy, integrity and validity of the IDB.

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

8.1. Overview

Creating the TALIS Starting Strong 2018 international database (IDB) and ensuring its integrity required close co-ordination and co-operation among the international study centre (ISC), the national project managers (NPMs) and the OECD Secretariat.

The primary goals of this work were to ensure that

- all national adaptations to questionnaires were reflected appropriately in the codebooks and corresponding documentation
- all national information eventually conformed to the international data structure and coding scheme
- any errors, such as logical inconsistencies or implausible values given by respondents or those occurring during data entry, were minimised as much as possible.

The quality control measures applied throughout the process were identical, to the extent possible, for both populations of interest (the ISCED level 02 core and the U3 level).

The IEA Hamburg held a three-day data-management training session in Athens, Greece, in January 2017, covering software use, procedures for national adaptations, and rules and procedures for data entry. The seminar was specifically targeted at the national team member(s) responsible for data management and liaising with the IEA Hamburg. In addition, after administration of the TALIS Starting Strong 2018 field trial in 2017, the ISC informed NPMs during the third NPM meeting held in Reykjavik, Iceland, in November 2017 about the outcomes of the field trial and subsequent improvements to the relevant software.

This chapter describes any further steps taken to build the IDB and assure the quality and accuracy of the TALIS Starting Strong 2018 data.

8.2. Online data collection and verification

TALIS Starting Strong 2018 was administered in two modes. Participating countries could select between online or paper administration of their questionnaires or choose a combination of both. This alternative selection of one of the modes could be made for the two ISCED levels, for the staff of single centres or for particular staff within one centre. Denmark, Iceland, Korea, Norway and Turkey used the online mode exclusively. Japan and Israel administered all questionnaires on paper, and Germany and Chile applied a mixed mode design. National study centres had to ensure that individual respondents who did not want to participate in the online mode or did not have access to the Internet were provided with a paper questionnaire, thereby minimising non-response as a result of a forced administration mode.

Later, data from the two different collection modes were merged into a single set per ISCED level and country. Potential sources of error originating from the use of the two parallel modes had to be controlled for and reduced as much as possible to ensure uniform and comparable conditions across modes as well as countries to the greatest extent possible. The design established several general similarities to achieve this:

- questionnaires in both modes were self-administered and comparable in terms of layout and appearance

- the same sample design and procedures were used to identify respondents
- the same methods were used to contact respondents and to validate their participation
- data collection by both modes occurred over the same period of time.

Notable differences included approaches to skipping questions (manually on paper, automatic on line) and the possibility of validating responses in real-time in the online mode.

The electronic versions of the TALIS Starting Strong questionnaires could only be completed on line. No other options were permissible, such as sending/receiving PDF documents by e-mail or printing out the online questionnaires and mailing them to the national study centres. Because the online data collection for TALIS Starting Strong was designed to meet the requirements of a standardised educational survey and specific operations, successful administration of the electronic questionnaires relied on the technical standard requiring all participating countries to use the software provided by the IEA Hamburg.

To properly sequence preparation tasks and processes and to ensure comparability of data, the paper versions of the different questionnaire types (i.e. centre, leader and combined questionnaires) had to be finalised first in terms of translation and layout, even if the expectation was that all or almost all of the data would be collected on line. These final paper versions of the questionnaires were converted for the online mode, followed by final verification of their structure, text and layout.

In addition to these considerations, the design ensured that online respondents needed only an Internet connection and a standard Internet browser. No additional software, particular operating system or particular make or version of browsers were required.

The navigational concept for the online questionnaires had to be as similar as possible to that of the paper questionnaires. Respondents could use „next" and „previous" buttons to navigate to an adjacent page, as if they were flipping physical pages. In addition, a hyperlinked „table of contents" mirrored the experience of opening a specific page or question of a paper questionnaire. While most respondents followed the sequence of questions directly, these features allowed respondents to skip or omit questions just as if they were answering a self-administered paper questionnaire.

To further ensure the similarity of the two sets of questionnaires, responses in the online questionnaires were not made mandatory, evaluated or enforced in detail (e.g. through use of hard validations or a strict sequence). Instead, some questions used soft validation, such as respondents being asked to give several percentage numbers that would supposedly add up to 100%. For these questions, the sum was constantly updated according to the respondent's entries and was highlighted in red throughout the time it differed from 100%. However, even if a response remained red, respondents could proceed to the next question.

Certain differences in the representation of the two modes nonetheless remained. To reduce response burden and complexity, the online survey automatically skipped questions not applicable to the respondent, in contrast to the paper questionnaire, which instructed respondents to proceed to the next applicable question. Rather than presenting multiple questions per page, the online questionnaire proceeded question by question. While vertical scrolling was required for a few questions, particularly the longer questions with multiple „yes/no" or Likert-type items, horizontal scrolling was not. Because respondents could easily estimate through visual cues the length and burden of a paper questionnaire, the

online questionnaires attempted to offer this feature through progress counters and a „table of contents" that listed each question and its response status. Multiple-choice questions were implemented with standard HTML radio buttons. Once the respondents had reached the end of the online questionnaires, they were presented with a summary of questions that had been left unanswered. This summary was for informational purposes and did not affect the administration. Respondents could quit the online administration at any time.

The ISC provided national study centres with a tool to monitor online participation. On demand, a restricted version of this tool was also available for centre co-ordinators. While NPMs could see summary information (e.g. first login, last login, total of logins, progress of answers) and raw data on all centre leaders and their staff, centre co-ordinators were allowed to view summary information for their centre to initiate appropriate follow-up activities but could not access actual responses.

8.3. Data entry and verification of paper questionnaires at national study centres

The IEA Hamburg supplied the national centres with the IEA Data Management Expert (DME) software and the *TALIS Starting Strong Main Survey Operations Procedures: Unit 5*. This document described the process and rules for using the DME software during manual entry of data from the leader and staff questionnaires. It also provided information about the data structure.

Each national study centre was responsible for transcribing into computer data files the information from the leader and staff questionnaires. National study centres entered responses from the paper questionnaires into data files created from an internationally pre-defined codebook that contained information about the names, lengths, labels, valid ranges (for continuous measures or counts) or valid values (for nominal or ordinal questions) and missing codes for each variable in each of the different questionnaire types.

Before data entry commenced, national data managers (NDMs) were required to verify the nationally adapted codebook, which was provided by the IEA Hamburg. The codebook reflected all ISC-approved adaptations made to the national questionnaire versions (e.g. a nationally added response category). Each codebook then served as a template for creating the corresponding dataset.

In general, national study centres were instructed to discard any questionnaire that was unused or returned completely empty and to enter any questionnaire that contained at least one valid response. To ensure consistency across participating countries, the basic rule for data entry in DME required national staff to enter data „as is" without any interpretation, correction, truncation, imputation or cleaning. Any inconsistencies that remained after this data-entry stage were dealt with at the time of data cleaning (see below).

The rules for data entry meant that:

- Responses to categorical questions were generally coded as „1" if the first option (checkbox) was used, „2" if the second option was marked, and so on.
- Responses to „check-all-that-apply" questions were coded as either „1" (marked) or „9" (omitted or invalid).
- Responses to numerical or scale questions (e.g. centre enrolment) were entered „as is", that is, without any correction or truncation, even if the value was outside the originally expected range (e.g. if a staff member reported that he or she spent 80 hours a week on tasks related to his or her job at the centre). If countries needed

to enter values that exceeded the defined variable width, they submitted these few values using an Excel sheet, and the values were later included during data processing at the ISC.

- Likewise, responses to filter questions and filter-dependent questions were entered exactly as filled in by the respondent, even if the information provided was logically inconsistent.
- If responses were not given at all, not given in the expected format, ambiguous or in any other way conflicting (e.g. selection of two options in a multiple-choice question), the corresponding variable was coded as „omitted or invalid".
- During data capture, TALIS Starting Strong did not use a separate code to identify „not administered" questions, such as those that were misprinted. In these rare cases, the „omitted or invalid" code was used.

Data entered with DME were automatically validated. First, the entered respondent ID had to be validated with a five-digit code – the checksum, generated by the IEA Windows Within-School Sampling Software (WinW3S). A mistype in either the ID or the checksum resulted in an error message that prompted the data-entry person to check the entered values. The data verification module of DME also enabled identification of a range of problems such as inconsistencies in identification codes and out-of-range or otherwise invalid codes. These potential problems had to be resolved or confirmed before the data-entry person resumed data entry.

To check the reliability of the data entry within the participating countries, their national study centres were required to have at least 100 completed leader questionnaires and 5% of the total number of completed staff questionnaires (or at least a minimum of 100 staff questionnaires) entered twice by different staff members as early as possible during the data-capture period. This procedure allowed NDMs and the ISC to identify possible systematic or incidental misunderstandings or mishandlings of data-entry rules and to initiate appropriate remedial actions, for example, re-training national study centre staff. The acceptable level of disagreement between the originally entered and double-entered data was established at 1% or less; any value above this level required a complete re-entry of data. This restriction guaranteed that the margin of error observed for manually entered data remained well below the required threshold.

Before sending the data to the ISC for further processing, national study centres carried out mandatory verification steps on all entered data and undertook corrections as necessary. The corresponding routines were included in the DME software, and the data files were systematically checked for duplicate identification codes and data outside the expected valid range or values defined as valid. NDMs reviewed the corresponding reports, resolved any inconsistencies and, where possible, corrected problems by looking up the original survey questionnaires. NDMs also verified that all returned non-empty questionnaires were in fact entered and that the availability of data corresponded to the participation indicator variables and entries on the tracking forms.

While the questionnaire data were being entered, the national data manager at each national study centre used the information from the tracking forms to verify the completeness of the materials. Participation information (e.g. whether the staff concerned had left the centre permanently between the time of sampling and the time of administration) was entered in WinW3S.

In addition to the data files described above, national study centres provided the ISC with detailed data documentation, including hard copies or electronic scans of all original tracking forms and a report on data-capture activities collected as part of the online survey activities questionnaire. The ISC already had access to electronic copies of the national versions of all questionnaires and the final national adaptation forms as part of the layout verification process.

8.4. Data checking, editing and quality control at the IEA Hamburg

Once the data were submitted to ISC, data processing commenced. The objective of the process was to ensure that the data adhered to international formats, that information from centre leaders and their staff could be linked across different survey files and that the data accurately and consistently reflected the information collected within each participating country.

The ISC went to great lengths to ensure that the data received from participating countries were internationally comparable and of high quality. The foundation for quality assurance had been laid before the data were submitted to the ISC through the provision of manuals, training and software designed to standardise a range of operational and data-related tasks, and through the verification of the content and layout of the national adaptation forms, paper questionnaires, online questionnaires and codebooks. Two software programmes provided by the ISC were instrumental to the efficacy of the quality assurance process:

- *The WinW3S software* performed the within-centre sampling operations and provided strict adherence to the sampling rules defined for TALIS Starting Strong (see Chapter 5 for more details on centre sampling). The software created all necessary listing and tracking forms and stored centre- and staff-specific information, such as gender and participation status. The software also generated login credentials consisting of a unique ID and the corresponding checksum (the five-digit validation code). These login credentials were used for the online administration and at the same time for entering the paper questionnaires into the DME software in case the respondent answered paper questionnaires. For this purpose, WinW3s created questionnaire labels that included the generated login credentials. It also included a participation rate estimator that kept NDMs up to date on their current participation rate according to the already returned paper questionnaires or submitted online questionnaires. The ISC asked the NDMs to begin uploading the participation rate reports to them on a weekly basis from the moment that the data collection began.
- *The DME software* enabled entry of all questionnaire data in a standard, internationally defined format. Data entered with the DME software were automatically validated. This process included validation of login credentials and a range of other issues such as the uniqueness of the ID or out-of-range or otherwise invalid codes. When such issues were flagged by the software, the individuals entering the data were prompted to resolve or to confirm the inconsistencies before resuming data entry. In addition, special variables called „check variables" were used during data entry to avoid any individual entering values for the wrong question. If participating countries wanted to use different tools to enter their data, the DME provided them with the possibility of importing these data and verifying them with the same range of checks available to participating countries entering the data directly via the DME software. The software also included a range of checks

for data verification that NDMs were requested to perform during the process of data capture as well as after data entry.

A complex study such as TALIS Starting Strong required a correspondingly complex data-cleaning design. Accordingly, the ISC developed processing tools in Structured Query Language (SQL) and, where necessary, in Statistical Package for the Social Sciences (SPSS). The ISC took the following steps to ensure that programs ran in the correct sequence, that no special requirements were overlooked and that the cleaning process was implemented independently of the persons in charge.

- Before the data-cleaning programs were applied to real data, they were all thoroughly tested through use of simulated datasets containing all the expected problems or inconsistencies.
- To document versions and updates, all incoming data and documents were registered in a specific material-receipt database. The date of arrival was recorded, along with any specific issues meriting attention.
- All national adaptations and all detected deviations from the international data structure were recorded in a „national adaptation database" and verified against the national adaptation form (NAF), the national instruments, the codebooks and the contents of the data. Data analysts can find the reports from this process in the *TALIS 2018 and TALIS Starting Strong User Guide* (OECD, 2019).
- The cleaning was organised according to rules strictly and consistently applied to all national datasets, making deviations from the cleaning sequence impossible.
- All systematic or manual corrections made to data files were implemented and recorded in specific cleaning reports for the TALIS Starting Strong International Consortium and for NPM review and approval.
- On completion of the data cleaning for a participating country, all cleaning checks were repeated from the beginning to detect any problems that might have been inadvertently introduced during the cleaning process itself.

Figure 8.1 provides a schematic overview of this iterative process conducted in co-operation with the national study centres. The sequential data-cleaning steps displayed in the exhibit are described in more detail in the following sections.

8.4.1. Dealing with the roles of centre leaders

The diverse personnel structures across the centres that participated in TALIS Starting Strong 2018 required a complex handling of different scenarios, especially those in which centre leaders took on staff-related duties in addition to their role as centre leader.

To cope with centres that consisted of only one person responsible for a small number of children, the TALIS Starting Strong Questionnaire Expert Group (QEG) developed a questionnaire that combined suitable questions from the staff questionnaire and selected questions from the leader questionnaire. Several of the participating countries used this combined questionnaire and administered it to either or both of the two survey populations (i.e. ISCED level 02 and U3). For more details, see Chapter 3 of this report.

In addition, the QEG provided a longer version of this combined questionnaire to accommodate the specific case of the ISCED level 02 population in Israel, where centres are composed of a main early childhood education and care (ECEC) staff member who acts as the lead pedagogical staff person as well as the leader, but is supported by usually one

or two additional teacher assistants who help with some tasks. For more details, see Chapter 3.

TALIS Starting Strong 2018 also had to consider centres that consisted of a leader and additional staff but where the leader had staff duties in addition to his or her leadership role. In these cases, the leader completed only the leader questionnaire because the combined questionnaires were administered only to centres that consisted of just one person responsible for the children's care and education.

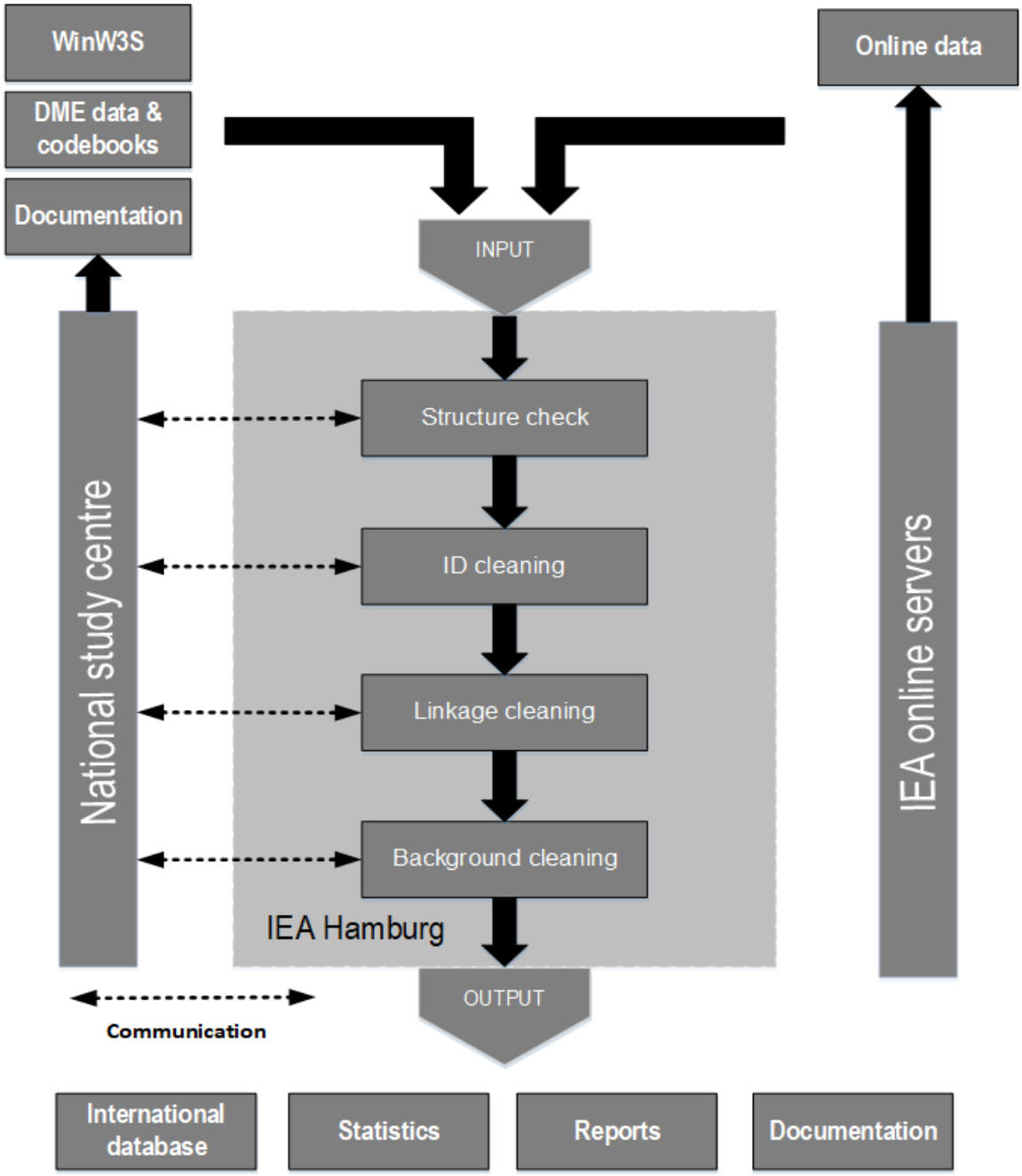
This complex structure had to be accommodated during data cleaning. This accommodation consisted of two data regulation steps. First, data from the two combined questionnaires were distributed amongst the staff and leader datasets, with additional records added to both datasets and the respondents' answers transferred to the relevant questions in the staff and the leader questionnaires. Second, an additional record was added to the staff dataset for those leaders who had staff duties but who received a leader questionnaire. Their responses were transferred to any questions that matched across the staff questionnaire and the leader questionnaire. (Information about the roles of staff in the centres was derived from the WinW3S software.)

The distribution of data from the combined questionnaires into the leader and staff dataset was one of the last cleaning steps within the data processing procedure. Prior to this distribution, the individuals carrying out the data cleaning resolved any inconsistencies in the questionnaire data separately for all four questionnaires, as described in the following sections. The questions addressed during the different types of data update are indicated in these sections by the abbreviations used for each staff question (SQ) and each leader question (LQ). The SQ or LQ question corresponding to questions in the combined questionnaires is additionally indicated in parentheses, where CQ refers to the combined questionnaire and CQ+ refers to the amended combined questionnaire developed for Israel.

8.4.2. Import, documentation and structure check

Data cleaning began with an analysis of the submitted data-file structures and a review of the data documented on the tracking forms. Most participating countries submitted all required documentation along with their data, which greatly facilitated the data checking. The ISC contacted the participating countries that returned incomplete data or documentation to obtain any missing material. Further data processing began as soon as the ISC received all required materials.

Figure 8.1. Iterative data-cleaning process



All available codebooks and data were imported from the source files and combined in the Structured Query Language (SQL) database. The database thus included the respondents' answers collected on line, the respondents' data entered into the DME, and all sampling and tracking information generated by and collected with WinW3S. During this step, the data originating from the paper questionnaires and the data originating from the online questionnaires were combined and checked for structural agreement, with sample and tracking information taken into consideration. In all cases, the data from both administration modes were structurally equivalent and used the same valid and missing codes. The early combination of these data during the import stage ensured that data

resulting from both administration modes were fed through the same data-processing steps and checks described in the remainder of this chapter.

Data from both ISCED levels were processed with the same type of checks. The content of the checks was nevertheless adapted by addressing the correct variable names, whenever necessary, to accommodate the different structure of the questionnaires used for the ISCED level 02 and U3 populations.

The structure check implemented at the ISC looked for differences between the international and the national file structures. As described above, some participating countries made structural adaptations to the questionnaires, the extent and nature of which differed greatly across these countries. Whereas some participating countries administered the questionnaires without any changes, except for translations and necessary cultural adaptations, others inserted or removed questions or options within existing international variables or added entirely new national variables.

Given the associated risk of deviating from the international data structure, NPMs wishing to make such changes had to follow certain strict rules to allow unequivocal integration of nationally adapted variables for international comparison. Where necessary, the ISC modified the data according to the international design to ensure that the resulting data were internationally comparable. For instance, additional national options in multiple-choice questions were recoded (mapped) in such a way that they adhered to the international code scheme. National variables were created so that the original values could be used later in reports presenting the national results of the survey.

NPMs and NDMs received detailed reports on structural deviations together with documentation on how the ISC resolved them. In a few cases, data were not available for certain variables because the corresponding question was not administered nationally (see the *TALIS 2018 and TALIS Starting Strong User Guide*, OECD, 2019).

8.4.3. Identification variable and linkage cleaning

To uniquely identify, track and document each respondent and each corresponding questionnaire in a survey, each record in a data file needs to have an identification number specific to it. The existence of records with duplicate identification (ID) numbers in a file implies an error. Because use of the DME software and the IEA Online Survey System (OSS) guaranteed the uniqueness of IDs in TALIS Starting Strong 2018, duplicate IDs could only occur in two situations: first, if the respondent answered the questionnaire online and completed a paper questionnaire as well, and, second, if a country administered the questionnaire in more than one language and the respondent completed the online questionnaire in two languages. If two records shared the same ID number and contained exactly the same data, one of the records was deleted and the other was left in the database. If the records contained different data (apart from the ID numbers), and it was impossible to identify which record contained the „authentic" data, NPMs were consulted to resolve the matter.

During TALIS Starting Strong, data collected at the centre level were recorded in the leader file. Ensuring that the records from these files could then be linked to the multiple staff-level records for that centre was crucial. The linkage was implemented through a hierarchical ID numbering system, and it was also cross-checked against the tracking forms.

Further ID cleaning focused on consistent tracking of information between the data used for listing, sampling and tracking in WinW3S and the actual responses in the questionnaire.

Whenever necessary, the variables pertaining to staff members' gender, year of birth, their staff role, exclusion status and participation status were verified and checked against the original paper version of the tracking forms.

Where possible, the ISC sought close co-operation with the national study centres in order to resolve ID or linkage inconsistencies. For this purpose, NPMs and NDMs received standardised reports containing each identified inconsistency. Once the ID, linkage, participation and exclusion information was finalised, data were transferred to the sampling unit of the ISC and used to calculate participation rates, exclusion rates and, finally, sampling weights.

8.4.4. Resolving inconsistencies in questionnaire data

After matching the national and international data structures as specified in the international codebooks and having resolved all ID and linkage issues, the ISC applied a series of standard cleaning rules to the data. The process, conducted through the SQL programs developed at the ISC, identified and, in many cases, automatically corrected inconsistencies in the data. The IEA Hamburg prepared detailed documentation of all cleaning checks, procedures and actions applied to the data, sent these to the national study centres and then explained them during the fourth NPM meeting in October 2018.

Filter questions, which appeared in certain positions in the TALIS Starting Strong 2018 questionnaires, were used to direct respondents to a particular question or section of the questionnaire. Filter questions and their dependent questions were treated automatically in most cases.

If the filter question contained a value and the respondent validly skipped the dependent questions, the dependent variables were coded as „logically not applicable". If a response to a filter question was equivalent to „no", meaning that the dependent questions were not applicable, and yet the respondent answered the dependent questions in an ambiguous pattern, the dependent variables were set to „logically not applicable" regardless of the value originally recorded in the dependent variable. The questions to which this setting applied were SQ-10/11 and 12 (CQ-10/11 and 12, CQ+-09/10–12); SQ-13/14 and 15 (CQ-21/22 and 23, CQ+-23/24 and 25); and SQ-27/28 (CQ-38/39, CQ+-48/49). Recodings for SQ-07/08 (CQ-07/08) were exceptions to this general rule (see the next section of this chapter, section 8.4.5).

Split-variable checks were applied to „yes/no" questions for which the responses needed to be coded into several variables. For example, question 11 in the staff questionnaire (SQ-11, CQ-11, CQ+-10) listed a number of education or training programmes that prepare people to work with children, and it asked respondents to mark whether they completed these programmes with „yes" or „no". Occasionally, respondents marked some yes or no boxes or just the yes boxes but left some of the no boxes unchecked, resulting in omitted values in the data file. In those instances where a combination of „yes" and „omitted or invalid" responses were found for a given question in the data, then the individuals involved in the data cleaning assumed that the unmarked boxes actually meant no. The questions to which this setting applied were LQ-04 (CQ+-12), LQ-08, LQ-25 (CQ-31, CQ+-36), LQ-28 (CQ+-39), LQ-31 (CQ+-40), SQ-12 (CQ-12, CQ+-11), SQ-13 (CQ-21, CQ+-23), SQ-14 (CQ-22, CQ+-24) and SQ-15 (CQ-23, CQ+-25).

Split-variable checks were also applied to „check-all-that-apply" questions for which the responses needed to be coded into several variables. For example, question 13 in the leader questionnaire (LQ-13, CQ-15, CQ+-15) asked respondents to consider statements that best

described the centre. Leaders could mark as many check boxes as appropriate. In line with the split-variable checks applied to „yes/no" questions, in those cases where some check boxes were left omitted while others were marked „checked", the check boxes coded as „omitted or invalid" were recoded as „not checked". If all of the check boxes within a question were marked as „not checked", the data cleaner recoded the whole question to „omitted or invalid". LQ-21(CQ-29, CQ+-32) was treated in the same manner.

The TALIS Starting Strong questionnaires included several questions that consisted of a list of numeric variables. These questions asked respondents to enter either a percentage or a number. In cases where some of the questions of the list were left omitted, while others were answered with a value, the omitted values were recoded to zero („0"). This was the case for LQ-05, LQ-17 (CQ+-19), LQ-18 (CQ+-20), LQ-22 (CQ+-33), SQ-06 (CQ-06, CQ+-06), SQ-35 (CQ+-56), and SQ-39 (CQ-48, CQ+-60).

The individual responses to percentage questions were summed. If these values fell outside the 90 to 110 range or if any of them were larger than 100, they were set to „omitted or invalid".

Finally, inconsistencies between the listing information in the tracking forms and the actual responses of the leaders and staff regarding their age and gender were resolved by giving precedence to the information provided by the leader and the staff member.

The number of inconsistent or implausible responses in the data files varied from one country to another, but no national data were completely free of inconsistent responses. Each problem was recorded in a database and identified by a unique problem number. The record also included a description of the problem and of the automatic action taken by the program or the manual action taken by ISC staff.

ISC staff referred issues that could not be corrected using systematic rules to the relevant NPM so that the original data-collection instruments and tracking forms could be checked to trace the source of the inconsistency. Whenever possible, staff at the ISC suggested a solution and asked the NPMs either to accept it or to propose an alternative. Data files were then updated to reflect the agreed-upon solutions. Both systematic corrections and those apparent on a case-by-case level were applied directly in SQL program syntax and carried out automatically for each cleaning run.

8.4.5. Final cleaning actions

If NPMs could not solve problems by inspecting the instruments and forms or could not suggest a satisfying solution or explanation, the TALIS Starting Strong Consortium defined the final cleaning rules. Any systematic content edits were agreed to by the ISC and the OECD Secretariat and documented for use by the NPM.

These final cleaning actions were implemented after the content from the combined questionnaires had been distributed to the staff and leader data set; they therefore apply to these two additional questionnaires. For more information, see section 8.4.1 of this chapter.

Final action for logically inconsistent filter/dependent responses

Except for the recodings described in section 8.4.4 of this chapter (i.e. the dependent questions recoded to the missing code „logically not applicable" if the respondent answered the filter question in a way that made a response to dependent questions logically impossible), the data management team at the IEA Hamburg conducted the following additional recodings as part of the final cleaning action.

First, they implemented a special treatment for LQ-21 (CQ-29, CQ+-32). Here, if the respondent marked the category „not applicable" but at the same time marked one of the other categories within the same dimension, the data cleaners set the category „not applicable" to „not checked". If, for example, the respondent marked the category „not applicable" for the first dimension (i.e. LQ-21A4, CQ+-32A4) and then additionally marked „me and/or other member of the staff", „ECEC centre governing board" or „local, municipality,… authority" (i.e. LQ-21A1–A3, CQ+-32A1–3), data cleaners set the category „not applicable" (i.e. LQ-21A4, CQ+-32A4) to „not checked".

Second, the IEA Hamburg and Statistics Canada agreed to a special treatment for SQ-07/08 (CQ-07/08) for weighting purposes and calculation of the staff multiplicity factor (WGTADJ4) (see Chapter 9 for more details).

- If SQ-07 (CQ-07) was „yes" (1) and SQ-08 (CQ-08) was „omitted or invalid" or zero (0), the data cleaners set SQ-07 (CQ-07) to „no" (2) and SQ-08 (CQ-08) to „logically not applicable".
- If SQ-07 (CQ-07) was „no" (2) and SQ-08 (CQ-08) was zero (0) or one (1), data cleaners set SQ-08 (CQ-08) to „logically not applicable".
- If SQ-07 (CQ-07) was „no" (2) but SQ-08 (CQ-08) was two (2) or more, the data cleaner set SQ-07 (CQ-07) to „yes" (1).

Final action for out-of-range/implausible numerical variables

- LQ-06 (CQ+-07): values higher than 120 were set to „omitted or invalid".
- LQ-17a (CQ+-19a): values higher than 15 were set to „omitted or invalid".
- LQ-17b–g (CQ+-19b–g): values higher than 500 were set to „omitted or invalid".
- LQ-18a–c (CQ+-20a–c): values higher than 500 were set to „omitted or invalid".
- LQ-19 (CQ-19, CQ+-21): if the value was zero (0), the value was set to „omitted or invalid".
- SQ-06a/b (CQ-06a/b, CQ+-06a/b): values higher than 64 were set to „omitted or invalid".
- SQ-06c–e (CQ-06c/d, CQ+-06c/d): values higher than 63 were set to „omitted or invalid".
- SQ-18/19 (CQ-26/27, CQ+-29/30): values higher than 120 were set to „omitted or invalid".
- SQ-35a–g (CQ+-56a–g): values higher than 500 were set to „omitted or invalid".
- SQ-37a–b (CQ-46a–b, CQ+-58a–b): if the sum of enrolment was zero (0) or larger than 100, the values were set to „omitted or invalid".
- SQ-39a–g (CQ-48a–g, CQ+-60a–g): values higher than 200 were set to „omitted or invalid". If the sum of enrolment was zero (0), the values were set to „omitted or invalid".

Final recoding for inconsistent leader and staff age and gender in listing and questionnaire information

The affected variables were LQ-01, LQ-02, SQ-01 (CQ-01, CQ+-01), SQ-02 (CQ-02, CQ+-02), GENDER_L and GENDER_S (gender on listing form), ITBIRTHY_L and ITBIRTHY_S (birth year on listing form), and the following rules were applied:

- Staff gender (SQ-01 (CQ-01, CQ+-01) vs. GENDER_S): (a) believe questionnaire information and replace the information taken from WinW3S Listing Forms on gender in case it was missing or inconsistent; (b) impute missing questionnaire value from listing if questionnaire variable was omitted.
- Leader gender (LQ-01 vs. GENDER_L): (a) believe questionnaire and replace the information taken from WinW3S Listing Forms on gender in case it was missing or inconsistent; (b) impute missing questionnaire value from listing if questionnaire variable was omitted.
- Staff age (SQ-02 (CQ-02, CQ+-02), ITBIRTHY_S): (a) believe questionnaire information and delete information taken from WinW3S Listing Forms in case it was inconsistent; (b) impute missing questionnaire value from listing form; (c) set SQ-02 (CQ-02, CQ+-02) to „omitted or invalid" if outside the range of 14 to 80; (d) set ITBIRTHY_S to „omitted or invalid" if outside the range of 1938 to 2004 (ITBIRTHY_S).
- Leader age (LQ-02, ITBIRTHY_L): (a) believe questionnaire information and delete information taken from WinW3S Listing Forms in case it was inconsistent; (b) impute missing questionnaire value from listing form.

8.4.6. Handling of missing data

During manual entry, via DME, of the TALIS Starting Strong data at the national study centres, two types of entry were possible – valid data values, and missing data values. Data-entry staff could assign either the valid values or a value for „omitted or invalid". Later, at the ISC, staff applied additional missing values to the data that would be used for further analyses and to differentiate response behaviour.

Four missing codes were used for the data in the international database (IDB):

- *Omitted or invalid (9):* Here, the respondent had the opportunity to respond to the question but did not do so or provided an invalid response. ISC staff also assigned the value in extremely rare cases where questions were misprinted or otherwise not legible.
- *Not administered (8):* If the returned questionnaire was empty, all variables referring to that instrument were coded as „not administered" (unit non-response). In addition, a country may have chosen not to administer a certain question in their national questionnaire as documented on the NAF. The variables corresponding to the question that was not administered were coded as „not administered". The same rule applied if all respondents for a questionnaire left out a particular variable.
- *Not reached (7):* A special missing code was assigned to questions that were deemed „not reached" to distinguish them from omitted responses. Omitted questions were those that a respondent most probably read but either consciously decided not to answer or accidentally skipped. In other words, the respondent started answering the questions but stopped answering them before the end of the

questionnaire, probably because of a lack of time, interest or willingness to co-operate. Not reached variables were exclusively located towards the end of the questionnaires. The following algorithm was used for the „not reached", code. First, the data management team at IEA Hamburg identified the last valid answer given in a questionnaire. The data management team then coded the first omitted response after this last answer as „omitted or invalid" but coded all following responses as „not reached". For example, the response pattern „1 9 4 2 9 9 9 9 9 9" (where „9" represents „omitted or invalid") would be recoded to „1 9 4 2 9 7 7 7 7 7" (with „7" representing „not reached"). When recoding „omitted or invalid" values to „not reached", the data management team at IEA Hamburg ignored all „not administered" values. Thus, for example, the pattern „3 1 5 2 9 9 9 8 9 9" would be recoded to „3 1 5 2 9 7 7 8 7 7".

- *Logically not applicable (6):* This code meant that the respondent answered a preceding filter question in a way that made the following dependent questions not applicable to him or her. This value was assigned during data processing only.

8.5. Interim data products

Building the TALIS Starting Strong international database was an iterative process during which the ISC provided the OECD Secretariat and NPMs with a new version of data files whenever a major step in data processing was completed. This process guaranteed that NPMs had a chance to review their data and run additional plausibility and statistical checks to validate the data. The data products that the ISC released to the OECD Secretariat and each NPM included the staff and leader data file as well as data summaries. All interim data were made available to the OECD Secretariat in full, whereas each participating country received only its own data.

The ISC sent the first version of the cleaned and weighted data from Denmark, Germany, Iceland, Israel, Norway and Turkey to the OECD Secretariat at the end of September 2018 and cleaned and weighted the data of all participating countries, including Chile, Japan and Korea, in January 2019. All known identification, linkage and content issues in these data had been resolved by that time. Estimation weights and variables facilitating variance estimation were also included. To protect respondents' identity, the data management team at IEA Hamburg scrambled respondents' IDs.

The data sent to the OECD Secretariat at the end of September 2018 were also used to produce the first set of draft tables for the international report, presented to NPMs at the fourth NPM meeting held in Seoul, Republic of Korea, in October 2018. The meeting directly followed the fourth TALIS 2018 NPM meeting that was conducted in the same place. The NPMs from Denmark, Germany, Iceland, Israel, Norway and Turkey received a version of their own cleaned and weighted data prior to this meeting, thereby giving them a chance to review their data and the tables produced by the OECD Secretariat for accuracy and validity. Chile, Japan and Korea received their cleaned data only after the meeting because of their different national study schedules. More information on timelines can be found in section 6.2 of Chapter 6.

NPMs were able to raise any issues concerning their data that had thus far gone unnoticed during the fourth NPM meeting. In addition, countries could comment on their data via e-mail until 15 November, while those countries with a different national study schedule could raise issues until the end of January 2019. This situation resulted in a second, updated

data version that concluded the main survey's field work and included scale scores. The ISC sent it to the OECD Secretariat and NPMs at the end of February 2019.

All interim data products were accompanied by detailed data processing and weighting documentation and summary statistics. The latter contained weighted and unweighted univariate statistics and frequencies for all questionnaire variables for each country. For categorical variables, which represent the majority of variables in TALIS Starting Strong, the percentages of respondents choosing each of the response options were displayed. For numeric or count variables, various descriptive measures were reported. These included the minimum, the maximum, the mean, the standard deviation, the median, the mode, percentiles and quartiles. For both types of variables, the percentages of missing information due to respondents omitting or not reaching a particular question were reported. These summaries were used for a more in-depth review of the data at the international and national levels in terms of plausibility, unexpected response patterns, suspicious profiles and so on.

8.6. Building the international database

For both the draft and the final international database (IDB), data cleaning at the ISC ensured that the information coded in each variable was, in fact, internationally comparable, that national adaptations were reflected appropriately in all concerned variables and that all records could be successfully linked across leader and staff levels.

The interim data products described above and the draft and final (public-use) IDBs had two key differences:

- All interim products included one record for each sampled unit (leader or staff) even if the corresponding questionnaire was not returned or returned empty. In contrast, the draft and final IDBs included only those records that satisfied the sampling standards. Data from those units that either did not participate or did not pass adjudication (e.g. because within-centre participation was insufficient) were removed.
- To protect the confidentiality of respondents' information, disclosure avoidance measures were applied. These measures applied to all international-level datasets and were implemented for all data versions of the IDB for use by all countries and public users.

The disclosure avoidance measures involved the following procedures.

- The staff (IDSTAFF) and centre identifiers (IDCENTRE) were scrambled and therefore did not match those used during data collection; however, the structural link between the centre and the staff level (the variable IDCENTRE in the staff file and the first four digits of any IDSTAFF) was maintained. For each country, unique matching tables were created and made available to authorised individuals within the country and at the international level.
- Variables used purely for the stratification of the staff sample, that is, birth year (ITBIRTHY_S and ITBIRTHY_L) and gender (GENDER_S and GENDER_L), were removed. The only variables collected via the questionnaires that were retained were gender (for staff only, SS1G01) and age (SS1G02 and SL1G02).
- Variables used purely for stratification of centres (IDSTRATE and IDSTRATI) were removed to avoid identification of geographical or organisational groups.

Because the stratum information is mostly of interest for national-level analysis, it was of course made available to the country concerned. Experience shows that researchers from other countries might also wish to conduct analysis by stratification. If so, they will need to request the stratification variables directly from the country.

- Information used in the calculation of final sample and replicate weights was removed (for the centre level, WGTFAC1 and WGTADJ1, and for the staff level, WGTFAC1, WGTADJ1, WGTFAC2, WGTADJ2 and WGTADJ4), because these could have allowed identification of the stratification cells.
- Replication zone and unit variables (BRRSZONE, BRRSREP, BRRTZONE and BRRTREP) that could indirectly identify centres were also dropped from the public-use micro-data.
- Answers to numeric questions that make centres and staff easily identifiable were categorized. This applies to questions about leader and staff age (LQ-02 and SQ-02), years of experience of the leader (LQ-05) and the centre enrolment (LQ-19). The original variables SL1G02, SS1G02, SL1G05B and SL1G19 therefore have been replaced by the categorized variables LAGEGR, SAGEGR, LEXPGR and NENRCHLD. The number of staff at the centre (LQ-17) was also recoded into categories based on the proportion of these roles within a centre's workforce. The original variables SL1G17A-G have been replaced by three categorised variables: LTEACHGR (combining leaders and teachers in the centre), LASSISTGR (reflecting assistants in the centre) and LSTOTH (reflecting other staff in the centre). The number of staff who permanently left the centre during the last 12 months (SL1G18B) has been replaced by the categorised variable LSTAFFLEFT based on the proportion of these former staff relative to the current workforce. The variable reflecting the sum of staff members has been derived from the sum of SL1G17A-G and then transformed into the categorised variable SUMSTAFGR. In addition, responses were collapsed for two questions on centre funding sources (non-governmental organisations and benefactors/donations, both from LQ-14). The original variables SL1G14C and SL1G14D have been replaced by the combined responses in variable SL1G14CD. For more information, please see Chapter 11.2.1 of the *TALIS Starting Strong 2018 Technical Report* (OECD, 2019[1]).
- Additional variables that make leaders and thereby centres easily identifiable were dropped from the public-use micro-data. These include leader gender (SL1G01), details of leader work experience (SL1G05A and SL1G04C-E), hours worked each week (SL1G06), the number of centres where the leader works (SL1G07), the centre location (SL1G11 and SL1G13A-G), centre for-profit status (SL1G16), the number of staff who began work at the centre (SL1G18A), the number of staff who temporarily did not work at the centre (SL1G18C) and the assigned mode of the leader questionnaire (i.e. online or paper, MODE_L).
- Several staff variables were dropped from the public-use micro-data that could be sensitive if individual leaders were identified. Staff responses about job satisfaction (SQ-44) were dropped, with the exception of items not specific to staff work at a particular centre (SS1G44A, SS1G44I, SS1G44J and SS1G44K). The most likely reason for staff to leave their role as an ECEC staff member (SS1G46) was also dropped.

- Finally, data on leaders' educational attainment (SL1G03) have been replaced by the categorical variable LISCED. Furthermore, the variable LISCED has been dropped additionally for Denmark to reduce the risk of identifying individual leaders in this country.

The building of the IDB complied, at all times, with the OECD's rules for processing and transferring personal data and with rules stipulated by national or regional legislation for the protection of personal data, such as the European Union's General Data Protection Regulation (GDPR).

To protect its respondents' privacy, Iceland decided to withdraw all data from the IDB. This information can now be obtained only through direct contact with Iceland.[1]

After each NPM and the OECD Secretariat had agreed on data-release policy and confidentiality agreements, the IEA Hamburg released a draft IDB on behalf of the OECD. This draft database contained data from all participating countries as of March 2019, prior to publication of the international report in October 2019. This release enabled participating countries to replicate results presented in the draft chapters of the international report. This data version was also used in an international database training session that staff from the IEA Hamburg held in Lüneburg, Germany, in March 2019 for accredited members of the participating countries' teams.

The final, public-use international database scheduled for release in October 2019 is supplemented by full documentation in the *TALIS 2018 and TALIS Starting Strong User Guide* (OECD, 2019[2]). The database provides a unique resource for policy-makers and analysts. It contains data from ECEC centre leaders and staff from 13 different samples in nine participating countries and follows the TALIS 2018 design model. It therefore provides a strong link to TALIS 2018, which contains data from schools and teachers from 83 different samples in 48 participating countries/economies across five continents. The overlap between the indicators used in TALIS 2018 and TALIS Starting Strong 2018 adds analytical value for researchers comparing indicators across populations.

Although data are included as part of the IDB, the sample adjudication process determined that the ISCED level 02 and U3 leader and staff data for Denmark does not reliably represent the population from which the sample was drawn. The sampling adjudication variable IN3S18 was therefore set to zero. Database users are encouraged to use only cases where IN3S18 equals 1 to ensure analyses are representative of the sampled populations.

Note

1. Please contact the TALIS Starting Strong team at the OECD to be put in contact with Iceland.

References

OECD (2019), *TALIS 2018 and TALIS Starting Strong 2018 User Guide*, OECD, Paris, http://www.oecd.org/education/talis/TALIS_2018-TALIS_Starting_Strong_2018_User_Guide.pdf. [2]

OECD (2019), *TALIS Starting Strong 2018 Technical Report*, OECD, Paris, http://www.oecd.org/education/talis/TALIS-Starting-Strong-2018-Technical-Report.pdf. [1]

Chapter 9. Estimation weights, participation rates, and sampling error

This chapter covers three important aspects of the quality of the Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018 outcomes. The first is the weighting of the data to reflect the sampling design and to produce unbiased estimates or to account for unequal probabilities of selection. Descriptions cover how each component of the final estimation weight was defined and how these components were assembled into the final estimation weight. The second aspect, participation rates, is detailed as a significant factor affecting data quality. The third and final aspect described is balanced repeated replication (BRR), the method chosen for this survey to achieve unbiased estimates of sampling error.

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

9.1. Overview

This chapter covers three important aspects of the quality of the TALIS Starting Strong 2018 outcomes: the weighting of the data to produce unbiased estimates or to account for unequal probabilities of selection, survey participation rates, and estimation of sampling error.

Although the international sampling plan was prepared as a self-weighting sampling plan of staff in early childhood education and care (ECEC) centres (whereby each staff member ultimately had the same final estimation weight), the actual conditions in the field, non-response at centre and staff level, and the co-ordination of multiple samples made it impossible to fulfil this ideal plan. In the end, in most participating countries, the national sampling plan was established as a stratified multi-stage probability sampling plan with unequal selection probabilities.

The following sections of this chapter detail how each component of the final estimation weight was defined and how these components were assembled into the final estimation weight. Annex G provides the resulting participation rates and estimated population sizes for each participating country and population of interest, that is, ISCED level 02 and centres serving children under three years of age (U3).

Due to unequal weights and the complex structure of the survey sample, the sampling error had to be estimated by accounting for the survey's design and weights. Failure to take this approach could have translated into severely biased estimates of sampling error.

Correctly estimating sampling error is often a daunting task, but simple and approximately unbiased methods are available. TALIS Starting Strong 2018 opted for balanced repeated replication (BRR) not only because of this method's statistical properties (consistency, asymptotic unbiasedness) and its portability (one formula fits all types of parameter estimates) but also because it is comparatively easy to compute.

The last section of this chapter explains how the variables ("replicates") needed for conducting this method were created and how the BRR estimates of sampling error were computed. These estimates of the sampling error are another key element of the statistical quality of survey outcomes.

A more detailed description of the survey design and its implementation can be found in Chapter 5 of this report and in the *TALIS Starting Strong 2018 Survey Operations Procedures: Unit 1*.

9.2. Computing estimation weights

9.2.1. Elements of the estimation weight (final weight)

The statistics produced for TALIS Starting Strong 2018 were derived from data obtained through samples of centres, centre leaders and centre staff. To achieve meaningful statistics for a participating country, those statistics needed to reflect the whole population from which they were drawn and not merely the sample used to collect them. The process of going from the sample data to information about the survey population is called estimation. Note, however, that sample counts do not suffice as estimates of population totals (e.g. the total number of ISCED level 02 staff in a participating country).

The *estimation weight* or *final weight* is the device that allows the production of country-level estimates from the observed sample data. The estimation weight indicates the number

of population units that a sampled unit (e.g. a survey respondent) represents. The final weight is the combination of the many factors reflecting the selection probabilities during the various sampling stages and the response obtained at each stage.

Other factors may also come into play as dictated by special conditions directed towards maintaining unbiasedness of the estimates (e.g. adjustment for staff working in more than one centre). Because TALIS Starting Strong 2018 consisted of two populations of interest (ISCED level 02 and services for children under the age of three years (U3)), the estimation weights had to be computed independently for each population of interest. This requirement held true even if the samples were co-ordinated across the populations of interest.

Basically, final weights are the product of a design or base weight and of one or many adjustment factors; the former is the inverse of the selection probability, the latter compensates for random non-response within strata that could, if not accounted for, induce bias in the estimates. These design weights and adjustment factors are specific to each stage of a sample design and to each explicit stratum. Because the countries participating in TALIS Starting Strong 2018 adapted the general sample design to their own conditions, the estimation weights had to conform to these national adaptations.

The following are the conventional notations used in this chapter. The letters *h*, *i*, and *j* are used as subscripts, the lower-case letters *d*, *m*, *n*, *q*, *r* and *t* refer to the sample, and the upper-case letters *H*, *L*, *M* and *N* refer to the population (see also Table 9.1):

- Each participating country had 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 centres was drawn from the N_h centres comprising explicit stratum h. The index $i = 1, \ldots, n_h$ therefore points to the i^{th} sampled centre in explicit stratum h.
- Each centre $i = 1, \ldots, n_h$ within the explicit stratum h had a measure of size (MOS) noted as M_{hi}; the sum of the individual measures of size is therefore noted as M_h. Ideally, this MOS comprised the expected number of eligible staff members in each centre, but if this information was unavailable, the MOS could be another size measure such as number of children. If no MOS was available, M_{hi} was set to 1 for all centres in a stratum. The number of staff (without the centre leader) per centre at the time of staff sampling is given by L_{hi}.
- In each responding centre, a sample of m_{hi} staff was drawn from the listing of L_{hi} staff. If the selected centre was large enough, $m_{hi} = 8$ by design. The index $j = 1, \ldots, m_{hi}$ points to the staff, and m_{hi} may differ from 8 if local conditions dictated that the sample size should differ. The centre leader was always selected, regardless of whether he or she had staff duties as well.

Table 9.1. Conventional notations used in this chapter

Unit	Indices	Units sampled	Units participating	Eligible units non-participating	Ineligible units	Units in population
Explicit stratum	h					H
Centres	i	n	r		d	N
Measure of size						M
Staff (without centre leader)	j	m	t	q		L
Centre leaders with staff duties			t^l	q^l		

9.2.2. Centre base weight (centre design weight)

The first sampling stage in TALIS Starting Strong 2018 consisted of drawing the sample of centres. In most participating countries, this sample followed a systematic random sampling scheme with probability proportional to size (PPS). In participating countries where most centres were small or no measure of size (MOS) was available, systematic simple random samples were drawn instead, which meant each centre in a stratum had the same probability of being selected. Consequently, a centre base weight was needed to represent this first sampling stage. If all centres were selected (a census) in an explicit stratum of a participating country, then the centre base weight was set to 1 for that stratum.

Using the notation given above, for each centre $i = 1, \ldots, n_h$ and each explicit stratum $h = 1, \ldots, H$, the centre base weight was given by

$$WGTFAC1_{hi} = \frac{M_h}{n_h \times M_{hi}}.$$

In Iceland, because all centres were selected (i.e. $n = N$), $WGTFAC1_i = 1$ for all $i = 1, \ldots, N$.

In Israel, a three-stage sampling design was used for home-based settings that were part of the services for children under three years of age. All home-based settings in Israel are assigned to people called “co-ordinators”. Israel could not provide a full list of home-based settings, but it could provide a full list of co-ordinators. Co-ordinators were therefore sampled as the first stage of sampling. Each co-ordinator was then asked which home-based settings he or she worked with. The settings for which that co-ordinator was responsible were then listed and two of those home-based settings from that list were selected. The centre base weight was calculated as the product of the co-ordinator base weight as the first stage and the home-based setting base weight as the second stage.

9.1.1. Centre non-response adjustment factor

Despite all efforts to secure the participation of all selected centres in the participating countries, some centres were unable or unwilling to participate. The centre non-response adjustment factor was computed separately for the staff and the leader populations. The sampling team treated a centre as non-responding with regard to the staff population if less than 50% of the selected staff members participated. They treated a centre as non-responding with regard to the leader population if the centre leader did not return a completed questionnaire. Hence, a centre could be treated as participating for the staff population and non-participating for the leader population, or vice versa.

The non-participating centres needed to be represented by those that did participate. In accordance with the assumption that the respondents and the non-respondents were similar within the stratum, a non-response adjustment factor was required within each explicit stratum. Thus, for each explicit stratum $h = 1, \ldots, H$, if r_h centres participated in TALIS Starting Strong 2018 out of the n_h selected centres, and if d_h centres were found to be closed or out of scope, the non-response adjustment factor was

$$WGTADJ1_h = \begin{cases} \dfrac{n_h - d_h}{r_h}, & \text{for participating centres} \\ 0, & \text{for non-participating centres} \end{cases}.$$

9.1.2. Final centre weight

The centre estimation weight is the product of the centre base weight and the centre non-response adjustment factor. It was used to estimate the centre-related parameters, with the information drawn from the leader questionnaires.

The final centre weight (centre estimation weight) for each participating centre $i = 1, \ldots, r_h$ and each explicit stratum $h = 1, \ldots, H$ was

$$CNTRWGT_{hi} = WGTFAC1_{hi} \times WGTADJ1_h = \frac{M_h}{n_h \times M_{hi}} \times \frac{n_h - d_h}{r_h}.$$

9.1.3. Staff base weight (staff design weight)

In general, staff and centre leaders constituted different target populations and were selected separately within selected centres. However, in some centres, centre leaders also had staff duties. In an effort to keep the survey response burden at a tolerable level, these individuals were asked to complete the leader questionnaire but not the staff questionnaire, even though they remained in scope for the staff survey.

Some parts of the two questionnaires contained the same questions, so the data for these variables collected from the centre leaders with staff duties was copied over to the staff data file. All centre leaders with staff duties had a staff base weight of 1 because they had been selected with certainty in every centre.

In each participating centre, a systematic random staff sample was selected with equal probability from all staff who were not the centre leader. The nominal sample size within each centre was set at $m_{hi} = 8$, but the number of in-scope staff in each selected centre could mean that the staff sample size needed to be modified. Thus, for each selected staff $j = 1, \ldots, m_{hi}$ of centre $i = 1, \ldots, n_h$ in explicit stratum $h = 1, \ldots, H$, the staff base weight was

$$WGTFAC2_{hij} = \begin{cases} \dfrac{L_{hi}}{m_{hi}}, & \text{for staff who are not the leader} \\ 1, & \text{for leaders with staff duties} \end{cases}.$$

9.1.4. Staff non-response adjustment factor

Unfortunately, because not all selected staff were able or willing to participate in TALIS Starting Strong 2018, the non-participating staff needed to be represented by the participating ones. Under the assumption of "missing at random", representation was achieved by way of the staff non-response adjustment factor.

Each participating centre $i = 1, \ldots, r_h$ of each explicit stratum $h = 1, \ldots, H$ had three kinds of staff: those who responded to the survey questionnaire (noted t_{hi}), those who did not respond but were still at the selected centre (noted q_{hi}), and those who left the centre permanently after the sample had been selected. The staff non-response adjustment factor for the staff without centre leader duties was

$$WGTADJ2_{hij} = \begin{cases} \dfrac{t_{hi} + q_{hi}}{t_{hi}}, & \text{for responding staff} \\ 0, & \text{for non-responding staff} \end{cases}.$$

9.1.5. Non-response adjustment factor for centre leaders with staff duties

As mentioned earlier, the leader questionnaire contained data that were equivalent to the data collected in the staff questionnaire. Because these data were added to the staff dataset, centre leaders with staff duties are represented in the staff datasets. However, non-responding centre leaders were adjusted differently on the assumption that leaders with staff duties could not be adjusted by increasing the staff weights from the same centre because their answers might have been systematically different due to their role within the centre. Therefore, other centre leaders with staff duties from the same stratum represented non-participating centre leaders with staff duties.

The staff adjustment factor for centre leaders with staff duties was

$$WGTADJ2_h = \begin{cases} \dfrac{t_h^l + q_h^l}{t_h^l}, & \text{for responding leaders with staff duties} \\ 0, & \text{for non-responding leaders with staff duties} \end{cases},$$

where t_h^l is the number of participating leaders with staff duties in a specific stratum and q_h^l is the number of non-responding leaders with staff duties in a specific stratum.

9.1.6. Staff multiplicity adjustment factor

Some staff were working in more than one centre. Because the MOS of each centre was taken independently, these staff were counted more than once. Also, because the staff samples were independent from one centre to the next, selecting the same staff member more than once was possible. An adjustment was needed to account for the number of centres ($nb\_centres_{hij}$) in which a given staff member worked. This information was collected through the TALIS Starting Strong 2018 questionnaires. For most staff, the adjustment factor was 1 (i.e. for those staff working in only one centre); for the others, it was the reciprocal of the number of centres in which they worked.

For each staff $j = 1, \ldots, t_{hj}$, in each participating centre $i = 1, \ldots, r_h$, in explicit stratum $h = 1, \ldots, H$, the staff adjustment factor for multiplicity was

$$WGTADJ4_{hij} = \begin{cases} \dfrac{1}{nb\_centres_{hij}}, & \text{for responding staff} \\ 1, & \text{for non-responding staff} \end{cases}.$$

9.1.7. Final staff weight

The final staff weight (estimation weight) was the product of the staff base weight, the two adjustment factors associated with each participating staff member, and the final centre

weight. All estimates pertaining to the staff populations therefore needed to use the final staff weight.

For each participating staff $j = 1, \ldots, t_{hij}$, in each participating centre $i = 1, \ldots, r_h$, in explicit stratum $h = 1, \ldots, H$, the final staff weight was

$$STAFFWGT_{hij} = \{CNTRWGT_{hi}\} \times \{WGTFAC2_{hij} \times WGTADJ2_{hij} \times WGTADJ4_{hij}\}$$

$$= \left\{\frac{M_h}{n_h \times M_{hi}} \times \frac{n_h \times d_h}{r_h}\right\} \times \left\{\frac{L_{hi}}{m_{hi}} \times \frac{t_{hi} + q_{hi}}{t_{hi}} \times \frac{1}{nb\_centres_{hij}}\right\}.$$

In the simplest case, the sampling design prepared for TALIS Starting Strong 2018 would have yielded equal weights for all staff.

If we assume that (1) the centre sample size was distributed among the explicit strata proportionally to the number of staff in each stratum, (2) the MOS used to select centres was the same as the number of listed staff, (3) samples of eight staff could be selected from every selected centre, (4) the listing forms contained nobody but in-scope staff, (5) there was no leader with staff duties, (6) each selected centre and staff member participated, and (7) each staff member was working in only one centre, then the final staff weight would effectively have been the same for all the staff in the sample:

$$STAFFWGT_{hij} = \{CNTRWGT_{hi}\} \times \{WGTFAC2_{hi} \times WGTADJ2_{hi} \times WGTADJ4_{hij}\}$$

$$= \left\{\frac{M_h}{n_h \times M_{hi}}\right\} \times \left\{\frac{L_{hi}}{8} \times 1 \times 1\right\} = \frac{M}{180 \times M_h}\frac{M_h}{8} = \frac{M}{1440}$$

for $h = 1, \ldots, H$; $i = 1, \ldots, n_h$; $j = 1, \ldots, m_{hi}$, and where M is the total number of staff in the population of interest.

The general calculation of the final staff weight for participating leaders with staff duties is the same as the staff weight given immediately above:

$$STAFFWGT_{hij} = \{CNTRWGT_{hi}\} \times \{WGTFAC2_{hij} \times WGTADJ2_{h} \times WGTADJ4_{hij}\}.$$

And because the staff non-response adjustment factor differed, this factor translates into:

$$= \left\{\frac{M_h}{n_h \times M_{hi}} \times \frac{n_h \times d_h}{r_h}\right\} \times \left\{\frac{L_{hi}}{m_{hi}} \times \frac{t_h^l + q_h^l}{t_h^l} \times \frac{1}{nb\_centres_{hij}}\right\}.$$

9.2. Participation rates

The quality requirements for TALIS Starting Strong 2018 translated into participation rates (response rates) for centres and for staff. Reaching the required participation levels does not preclude some degree of error in the results but should reduce the reliance on the missing at random assumptions made for the non-response adjustments.

9.2.1. Participation rate for centres

The minimum centre participation rate set for TALIS Starting Strong 2018 was 75% after replacement. Although replacement centres could be called upon as substitutes for non-responding centres, the survey's national project managers were encouraged to do all they could to obtain the participation of the centres in the original sample. A high number of replacement centres could have led to the sample losing its probabilistic features and

becoming increasingly "purposive", a situation with the potential to undermine the reliability, validity and interpretability of a participating country's results.

The unweighted centre participation rate was computed as

$$UNWCPART = \frac{\sum_{h=1}^{H}\sum_{i=1}^{r_h} 1}{\sum_{h=1}^{H}\sum_{i=1}^{n_h-d_h} 1} = \frac{\sum_{h=1}^{H} r_h}{\sum_{h=1}^{H}(n_h - d_h)},$$

where r_h, n_h and d_h are as defined above. This formula represents the crude proportion of centres that achieved a response from their centre leaders.

The weighted centre participation rate was computed as follows:

$$WTDCPART = \frac{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{t_{hi}} WGTFAC1_{hi}}{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{t_{hi}} CNTRWGT_{hi}}.$$

This formula represents the proportion of the centre leader population accounted for by the participating centres.

Both rates were computed once over the complete set of participating centres (after replacement) and once over the subset of participating centres in the original selection (before replacement).

9.2.2. Participation rate for staff

The centre participation rate was calculated separately for the staff population as well, using the same formulas. Responding centres that reached at least 50% of responding staff were considered to be "participating" centres. Centres that failed to meet this threshold were considered to be "non-participating". As such, a centre could count as participating for the leader population but not for the staff population, or vice versa.

TALIS Starting Strong 2018 also expected at least 75% of the selected staff in participating centres (original sample or replacement centre) to take part in the assessment. Because the staff participation rate was calculated across all participating centres, whether the centres were in the original sample or used as a replacement, the participation rate for staff was a requirement only at the national level, not at the centre level.

The unweighted staff participation rate was defined as

$$UNWSPART = \frac{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{t_{hi}} 1}{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{m_{hi}} 1} = \frac{\sum_{h=1}^{H}\sum_{i=1}^{r_h} t_{hi}}{\sum_{h=1}^{H}\sum_{i=1}^{r_h} m_{hi}}.$$

This formula gives the crude ratio of the number of responding staff in participating centres with respect to the expected sample size from the participating centres.

The weighted staff participation rate was given as

$$WTDSPART = \frac{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{t_{hi}} WGTFAC1_{hi} \times WGTFAC2_{hij} \times WGTADJ4_{hij}}{\sum_{h=1}^{H}\sum_{i=1}^{r_h}\sum_{j=1}^{t_{hi}} STAFFWGT_{hij}}.$$

9.2.3. Overall staff participation rates

The overall unweighted and weighted participation rates were the product of the respective centre and staff participation rates.

9.2.4. Reporting participation rates

Both weighted and unweighted centre participation rates, with and without replacement centres, were produced. The weighted and unweighted participation rates for staff were also computed. The analytical results for each participating country were annotated according to whether or not the response rate requirements were adequately met (see Chapter 10 for further details). Annex G provides the unweighted and weighted centre participation rates before and after (1) replacement of the non-participating centres, (2) the staff participation rate in participating centres, and (3) the overall participation rate for each participating country.

9.3. Sampling error with balanced repeated replication (BRR)

Estimation of sampling error for surveys with complex sampling designs such as the one for TALIS Starting Strong 2018 requires special attention. Both the survey design and the unequal weights are needed to obtain (approximately) unbiased estimates of sampling error. Failure to do this can lead to severe underestimation of the sampling error. While exact formulae exist in theory for stratified PPS sample designs, the required computations become practically impossible as soon as the number of primary units selected per stratum exceeds two.

Approximate solutions for this problem have been proposed over the years. An important class of solutions is that of *resampling* or *replication*. *Interpenetrating sub-samples* (Mahalanobis), *balanced half-samples* or *balanced repeated replication* (McCarthy, Fay), the *jackknife* (Quenouille, Tukey, Durbin, Frankel), and the *bootstrap* (Efron) are the best-known examples of replication methods (for a review of these methods, see, for example, Lohr, (1999[1]); Rust and Rao (1996[2]); Wolter, (2007[3])).

In a similar vein to what was done for PISA (OECD, 2017[4]) and TALIS (OECD, 2014[5]), TALIS Starting Strong 2018 used the balanced repeated replication (BRR) to estimate the sampling error of the estimates. BRR is a replication method suited to sample designs where exactly two primary sampling units (PSUs) are selected in each stratum.

The BRR principle maintains that each of the two PSUs can provide an unbiased estimate of the total (or other parameter of interest) of its stratum. If the sampling design comprises H strata, then 2^H possible unbiased estimates of the parameter of interest are obtained by combining either PSU from each of the H strata. The sampling error of the estimate of the parameter of interest can be directly computed by comparing each of the 2^H estimates with their mean, as is usually done in simple basic statistics. Even with moderate values of H, the number of unbiased estimates may be quite large (e.g. $2^5 = 32$, $2^{10} = 1\,024$, $2^{20} = 1\,048\,576$). BRR provides a means of extracting from the complete set of 2^H possible replicates a much smaller subset that gives the same measure of sampling error as the full set.

9.3.1. Creating replicates for BRR

BRR was developed for sample designs that use only two PSUs per stratum. Clearly, none of the countries participating in TALIS Starting Strong 2018 implemented this type of design. Fortunately, the implemented sample design could be approximated through use of a superimposed "BRR-ready" sample plan. The participating centres (of the original sample or the replacements), listed in the order in which they appeared on the sampling frame and sorted by stratification and size, were paired within explicit strata, and each pair was dubbed "pseudo stratum" or "zone".

If the number of participating centres in an explicit stratum was odd, then a triplet was formed with the last three centres. If the sample was a census, as was the case in Iceland, the centres had the function of explicit strata, and the staff within centres were paired. In the very rare cases where a centre had only one staff member, it was not possible to form either a pair or a triplet, and this staff member stayed alone in a zone. The pairs (or triplets) were then numbered sequentially from 1 to *G*, thus spanning the whole sample, and within each pseudo stratum or zone, each centre was assigned a random pseudo PSU number 1 or 2 as depicted in Table 9.2. In those instances where there was an odd number of centres in an explicit stratum, the first centre was randomly designated as either 1 or 2, the second centre was designated as the opposite of the first, and the third centre was again randomly designated as 1 or 2.

Table 9.2. Example of "BRR-ready" sample design and random assignment of pseudo PSUs

Explicit stratum	Centre ID	Zone = pseudo stratum	Pseudo PSU
1	1001	1	1
1	1002	1	2
1	1003	2	2
1	1004	2	1
2	1005	3	1
2	1006	3	2
2	1007	3	1
3	…	…	…
…	…	…	…
H	…	G-1	2
H	…	G-1	1
H	…	G	1
H	…	G	2

One of the two pseudo PSUs will be dropped, and the weight of the remaining PSU will be doubled and used to compute an estimate of the parameter of interest. Rather than randomising which PSU will be dropped, we can use a special matrix (of order *4t*) of +1's and -1's – the so-called Hadamard matrix – to indicate which PSU is to be kept (+1) and which is to be dropped (-1) from each pseudo stratum in BRR, thereby associating the +1's with the PSUs numbered 1 and the -1's with the PSUs numbered 2.

For example, the Hadamard matrix of order 8 can be written as:

$$Hadamard_8 = \begin{pmatrix} +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & +1 & -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

In this matrix, each column is a BRR replicate and each line is a pseudo stratum or zone; the matrix entry indicates which pseudo PSU should be kept from each pseudo stratum to create the BRR replicate. The Hadamard matrix of order 8 translates into the pattern depicted in Table 9.3.

Table 9.3. Example of which pseudo PSU to select per zone

Zone	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5	Replicate 6	Replicate 7	Replicate 8
1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2
2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2
3	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2
4	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2
5	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2
6	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2
7	Pseudo PSU 1	Pseudo PSU 1	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 1	Pseudo PSU 2
8	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2	Pseudo PSU 2

Like TALIS and PISA, TALIS Starting Strong 2018 implemented a variation of the BRR developed by Fay (Fay, 1989[6]). Rather than completely dropping a PSU and doubling the weight of the other one, this variation multiplies the final weight ($STAFFWGT_{hij}$) of every staff member in the PSU indicated by the Hadamard matrix by a replicate factor of 1.5 to get the replicate weight. The replicate weight is similarly obtained by multiplying the final weight ($STAFFWGT_{hij}$) of the staff member in the remaining PSU by the replicate factor of 0.5. This strategy removes the risk of completely deleting some domain.

In those cases in TALIS Starting Strong 2018 where there was an odd number of PSUs in an explicit stratum, three centres constituted the last zone. For each replicate, as indicated by the Hadamard matrix, the weight of the selected unit was multiplied by 1.7071 if it was the single unit, and the weights of the remaining pair were multiplied by 0.6464. If the matrix indicated that the pair should be selected, then the weights of the paired units were multiplied by 1.3536, and the weight of the single unit was multiplied by 0.2929. This strategy, developed by Judkins (OECD, 2002[7]), ensures that the sum of the factors is 3.

The resulting factors based on the Hadamard matrix of order 8 look like the pattern in Table 9.4for the first seven centres of the example above.

Table 9.4. Example of BRR factors

Centre ID	Zone	Pseudo PSU	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
1001	1	1	1.5	1.5	1.5	0.5	1.5	0.5	0.5	0.5
1002	1	2	0.5	0.5	0.5	1.5	0.5	1.5	1.5	1.5
1003	2	2	1.5	0.5	0.5	0.5	1.5	0.5	1.5	1.5
1004	2	1	0.5	1.5	1.5	1.5	0.5	1.5	0.5	0.5
1005	3	1	0.6464	0.6464	1.3536	1.3536	1.3536	0.6464	1.3536	0.6464
1006	3	2	1.7071	1.7071	0.2929	0.2929	0.2929	1.7071	0.2929	1.7071
1007	3	1	0.6464	0.6464	1.3536	1.3536	1.3536	0.6464	1.3536	0.6464

The nominal sample size for TALIS Starting Strong 2018 was $n = 180$ centres. The smallest multiple of 4 that is at least as large as the number of needed zones for this number of centres is 92. Therefore, $G = 92$ zones or pseudo strata were created for each participating country; a series of $G = 92$ BRR replicate weights was also computed and stored. The creation of BRR weights was then applied to all participating countries, regardless of the sample size and the sample selection method.

9.3.2. Estimating the sampling error

Let θ be the population parameter of interest. Let $\hat{\theta}^*$ be the full sample estimate for θ obtained by using the final weight, and let $\hat{\theta}_g$, $g = 1, \ldots, 92$, be the $G = 92$ BRR replicate estimates of the same parameter of interest obtained by using the BRR weights described earlier. Then, with k set to equal 0.5, Fay's BRR estimate of the sampling variance and sampling error of $\hat{\theta}^*$ are respectively given by

$$\hat{V}_{FAY}(\hat{\theta}^*) = \frac{1}{G(1-k)^2}\sum_{g=1}^{92}(\hat{\theta}_g - \hat{\theta}^*)^2 = \frac{1}{23}\sum_{g=1}^{92}(\hat{\theta}_g - \hat{\theta}^*)^2.$$

$$se_{FAY}(\hat{\theta}^*) = \sqrt{\hat{V}_{FAY}(\hat{\theta}^*)}$$

9.3.3. Using sampling error when comparing estimates

When estimates (either variables or groups within a participating country) are compared across two participating countries, or when a participating country value is compared to the international average, the comparison must be scaled through use of the appropriate estimate of sampling error.

The standard error for the difference of two estimates from one participating country, say $\hat{\theta}_1$ and $\hat{\theta}_2$ (e.g. $\hat{\theta}_1$= hours spent on tasks related to the job at the centre and $\hat{\theta}_2 =$ hours spent with children at the centre), is given by

$$se(\hat{\theta}_1 - \hat{\theta}_2) = \sqrt{\hat{V}_{FAY}(\hat{\theta}_1) + \hat{V}_{FAY}(\hat{\theta}_2) - 2\widehat{Cov}(\hat{\theta}_1, \hat{\theta}_1)}\,.$$

The standard error for the difference of the estimates for two participating countries, say $\hat{\theta}_A$ and $\hat{\theta}_D$, is given by

$$se(\hat{\theta}_A - \hat{\theta}_D) = \sqrt{\hat{V}_{FAY}(\hat{\theta}_A) + \hat{V}_{FAY}(\hat{\theta}_D)}.$$

A regression on a dummy variable, as is illustrated by the following example, or the IEA IDB Analyzer software (analysis type "percentages and means") can be used to compare subpopulations within a participating country.

Suppose that the difference between male and female staff for some characteristic (e.g. professional development needs, content knowledge) is of interest. We can set a dummy variable *Gender* = 0 if male, *Gender* = 1 if female. A regression model can then be written as $Score = a_0 + a_1 Gender$. Clearly, if *Gender* = 0, then $Score_{\text{male}} = a_0$. Likewise, by setting *Gender* = 1, we obtain $Score_{\text{female}} = a_0 + a_1$. Design-based estimation of the regression parameters a_0 and a_1 can be done using appropriate software that uses the replicate weights to estimate the standard errors of regression parameters. If the significance test on a_1 cannot reject the null hypothesis H_0: $a_1 = 0$, then we must conclude that the scores for male and female staff are not significantly different.

If $\hat{\theta}$ is one of the statistics described above and $se(\hat{\theta})$ is the standard error of $\hat{\theta}$, then confidence intervals about zero can be easily obtained by computing the following boundaries:

$$lower_\alpha = \hat{\theta} - t_{\frac{\alpha}{2};df} se(\hat{\theta}) \text{ and } upper_\alpha = \hat{\theta} + t_{\frac{\alpha}{2};df} se(\hat{\theta}),$$

where $1-\alpha$ is the pre-set confidence level (e.g. $1-\alpha = 0.95$) and $t_{\frac{\alpha}{2};df}$ is $1-\frac{\alpha}{2}$ percentile of the Student distribution with *df* degrees of freedom. In most applications, *df* will be large enough to allow the use of the standard normal deviate $z_{1-\frac{\alpha}{2}}$ (e.g. $z_{1-\frac{\alpha}{2}} = 1.96$ for $\alpha = 0.05$). However, in order to confirm the number of degrees of freedom, there is still a need to verify how many zones actually contribute to the statistic $\hat{\theta}$ and how many BRR replicates contribute to the computation of $se(\hat{\theta})$. More on this matter is covered in the *TALIS 2018 and TALIS Starting Strong 2018 User Guide* (OECD, 2019[8]).

9.3.4. Design effect (deff) and effective sample size

The complex sampling designs in surveys such as TALIS Starting Strong 2018 are known to be "less efficient" than simple random samples of the same size. Usual explanations include the fact that respondents are selected in groups of individuals sharing many characteristics, such as centre environment, professional training, equipment and so on. The loss in efficiency is often summarised in a statistic called "design effect" or *deff* (Kish, 1965[9]). The design effect for a statistic and a sampling plan is the ratio of the variance of the estimate under the sampling plan to the variance of the same estimate under simple random sampling of the same size. In the case of TALIS Starting Strong 2018, the true design effect was approximated by

$$deff(\hat{\theta}, BRR) = \frac{\hat{V}_{BRR}(\hat{\theta})}{\hat{V}_{SRS}(\hat{\theta})}.$$

Alternatively, the design effect can be regarded as the ratio of sample sizes. We can then speak of "effective sample size" to describe the sample size of the complex survey adjusted for the design effect:

$$n_{effective} = \frac{n}{deff}.$$

The following tables (Table 9.5, Table 9.6, Table 9.7 and Table 9.8) give the estimated design effect for selected variables from the staff and the leader questionnaires, by participating country and population of interest. Deff values near 1 mean that the design is as precise as a simple random sample of the same size. Deff values larger than 1 mean that the sampling design is less efficient than a simple random sample of the same size. Deff values smaller than 1 indicate a gain in precision. Alternatively, the deff values indicate by what factor the sample size is affected. The larger the deff, the smaller the effective sample size; reciprocally, the smaller the deff, the larger the effective sample size. High design effects result in higher sampling errors for samples of the same size.

Because the value of the deff depends on the design itself (efficiency of the stratification, clustering, sample size) and on the true sampling variance of the variable in the population, the deff will vary from one variable to the next.

Table 9.5. Estimated design effects for selected variables, ISCED level 02, staff data

Participating country	Professional development needs, content knowledge	Time spent with children	Process quality, prosocial behaviour	Pedagogical practices in the target group	Job satisfaction
Chile	1.62	1.64	1.39	1.29	1.56
Denmark	1.04	1.09	2.46	1.34	1.42
Germany	1.50	1.69	1.77	1.20	1.59
Iceland	0.97	1.76	0.99	1.17	1.26
Israel	1.65	1.63	1.77	1.43	1.55
Japan	1.81	1.76	2.22	2.53	2.47
Korea	1.81	1.65	2.03	1.52	1.36
Norway	1.45	1.01	1.50	1.99	1.43
Turkey	1.96	3.28	0.99	3.12	2.84

Notes:

Professional development needs, content knowledge

For each of the areas listed below, please indicate the extent to which you currently need professional development:

- working with parents or guardians/families
- learning theories (e.g. socio-cultural, behavioural, cognitive, constructivist)
- facilitating learning in literacy and oral language
- facilitating learning in mathematics/numeracy
- facilitating learning in science and technology.

(SQ-16g–k, scale S1PDPE)

Time spent with children

Approximately how many hours did you spend together with children at this ECEC centre? (SQ-19)

Process quality, prosocial behaviour

To what extent do the following apply to ECEC staff at this ECEC centre?

- If invited, ECEC staff join in with the children's play.
- When ECEC staff play with children, the children are allowed to take the lead.
- ECEC staff show enjoyment when joining the children's play.
- ECEC staff arrange settings to allow extensive play amongst groups of children.
- ECEC staff encourage sharing amongst children.
- ECEC staff encourage children to help each other.
- ECEC staff encourage children playing in small groups to include other children.
- ECEC staff encourage children if they comfort each other.
- ECEC staff hug the children.

(SQ-29a–h & SQ-30a, scale S1PRDEV)

Pedagogical practices in the target group

Thinking about your work with the target group, how often do you do the following?

- I set daily goals for the children.
- I explain how a new activity relates to children's lives.
- I give different activities to suit different children's interests.
- I give different activities to suit different children's level of development.
- I adapt my activities to differences in children's cultural background.

(SQ-41a–e, scale S1PRACT)

Job satisfaction

How strongly do you agree or disagree with the following statement?

- All in all, I am satisfied with my job. (SQ-44k)

Table 9.6. Estimated design effects for selected variables, ISCED level 02, leader data

Participating country	Work experience	Time spent on leadership tasks	Diversity beliefs	Job satisfaction
Chile	1.14	0.77	0.97	0.78
Denmark	0.62	1.13	1.32	1.47
Germany	0.92	0.87	1.11	0.87
Iceland	0.89	1.00	0.98	0.59
Israel	1.02	1.07	a	1.18
Japan	2.90	1.54	1.51	1.60
Korea	1.38	1.94	2.01	1.25
Norway	1.64	1.37	1.31	1.58
Turkey	1.03	1.59	1.15	2.43

Notes:

Work experience

How many years of work experience do you have, regardless of whether you worked full-time or part-time?

- Years working as an ECEC centre leader in total. (LQ-05b)

Time spent on leadership tasks

On average throughout the last 12 months, what percentage of time in your role as an ECEC centre leader do you spend on the following tasks in this ECEC centre?

- Pedagogical leadership, including meeting with staff to promote quality learning environment, organising staff development programmes about new pedagogical approaches, child development, etc. (LQ-22b)

Diversity beliefs

In your view, approximately how many ECEC staff in this ECEC centre would agree with the following statements?

- It is important to be responsive to differences in children's cultural backgrounds.
- It is important for children to learn that people from other cultures can have different values.
- Respecting other cultures is something that children should learn as early as possible.
- It is important to recognise the different interests of children regardless of gender.

(LQ-29a–d, scale S1LDIVB)

[a] This question was not administered in Israel.

Job satisfaction

How strongly do you agree or disagree with the following statements?

- All in all, I am satisfied with my job. (LQ-36i)

Table 9.7. Estimated design effects for selected variables, services for children under the age of three years, staff data

Participating country	Professional development needs, content knowledge	Time spent with children	Process quality, prosocial behaviour	Pedagogical practices in the target group	Job satisfaction
Denmark	1.80	1.31	1.23	1.46	1.85
Germany	1.10	1.69	1.59	1.20	1.59
Israel	2.13	1.81	1.79	1.41	3.90
Norway	1.83	1.97	1.68	1.30	2.22

Notes:

Professional development needs, content knowledge

For each of the areas listed below, please indicate the extent to which you currently need professional development:

- working with parents or guardians/families
- learning theories (e.g. socio-cultural, behavioural, cognitive, constructivist)
- facilitating learning in literacy and oral language
- facilitating learning in mathematics/numeracy
- facilitating learning in science and technology.

(SQ-16g–k, scale S1PDPE)

Time spent with children

Approximately how many hours did you spend together with children at this ECEC centre? (SQ-19)

Process quality, prosocial behaviour

To what extent do the following apply to ECEC staff at this ECEC centre?

- If invited, ECEC staff join in with the children´s play.
- When ECEC staff play with children, the children are allowed to take the lead.
- ECEC staff show enjoyment when joining the children´s play.
- ECEC staff encourage sharing amongst children.
- ECEC staff encourage children to help each other.
- ECEC staff encourage children if they comfort each other.
- ECEC staff hug the children.

(SQ-29a–f & SQ-30a, scale S1PRDEV)

Pedagogical practices in the target group

Thinking about your work with the target group, how often do you do the following?

- I set daily goals for the children.
- I explain how a new activity relates to children's lives.
- I give different activities to suit different children's interests.
- I give different activities to suit different children's level of development.
- I adapt my activities to differences in children's cultural background.

(SQ-41a–e, scale S1PRACT)

Job satisfaction

How strongly do you agree or disagree with the following statement?

- All in all, I am satisfied with my job. (SQ-44j)

Table 9.8. Estimated design effects for selected variables, services for children under the age of three years, leader data

Participating country	Work experience	Time spent on leadership tasks	Diversity beliefs	Job satisfaction
Denmark	1.84	1.44	1.28	1.85
Germany	0.97	0.95	1.09	0.80
Israel	2.49	3.16	3.35	3.03
Norway	1.36	2.31	0.95	1.51

Notes:

Work experience

How many years of work experience do you have, regardless of whether you worked full-time or part-time?

- Years working as an ECEC centre leader in total. (LQ-05b)

Time spent on leadership tasks

On average throughout the last 12 months, what percentage of time in your role as an ECEC centre leader do you spend on the following tasks in this ECEC centre?

- Pedagogical leadership, including meeting with staff to promote quality learning environment, organising staff development programmes about new pedagogical approaches, child development, etc. (LQ-23b)

Diversity beliefs

In your view, approximately how many ECEC staff in this ECEC centre would agree with the following statements?

- It is important to be responsive to differences in children's cultural backgrounds.
- It is important for children to learn that people from other cultures can have different values.
- Respecting other cultures is something that children should learn as early as possible.
- It is important to recognise the different interests of children regardless of gender.

(LQ-29a–d, scale S1LDIVB)

Job satisfaction

How strongly do you agree or disagree with the following statements?

- All in all, I am satisfied with my job. (LQ-36i)

In Chapter 5 (Table 5.1), the sample size for staff was derived through use of a hypothesised deff of 2. The tables above show how the actual sample designs implemented in the various participating countries outperformed the design hypotheses for most presented variables, thus resulting, in many cases, in effective samples that were much larger than the expected nominal 400 staff. However, as these tables also show, the various samples were not always as efficient as simple random samples, despite stratification.

References

Adams, R. and M. Wu (eds.) (2002), *Programme for International Student Assessment (PISA): PISA 2000 Technical Report*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264199521-en. [7]

Fay, R. (1989), "Theoretical application of weighting for variance calculation", *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 212–217. [6]

Kish, L. (1965), *Survey Sampling*, Wiley, New York. [9]

Lohr, S. (1999), *Sampling: Design and Analysis*, Duxbury Press. [1]

OECD (2019), *TALIS 2018 and TALIS Starting Strong 2018 User Guide*, OECD, Paris, http://www.oecd.org/education/talis/TALIS_2018-TALIS_Starting_Strong_2018_User_Guide.pdf. [8]

OECD (2017), *PISA 2015 Technical Report*, OECD, Paris, https://www.oecd.org/pisa/data/2015-technical-report/. [4]

OECD (2014), *TALIS 2013 Technical Report*, OECD, Paris, http://www.oecd.org/education/school/TALIS-technical-report-2013.pdf. [5]

Rust, K. and J. Rao (1996), "Variance estimation for complex estimators in sample surveys", *Statistics in Medical Research*, Vol. 5, pp. 381–398. [2]

Wolter, K. (2007), *Introduction to Variance Estimation*, Springer Verlag, New York. [3]

Chapter 10. Data adjudication

This chapter covers the adjudication of TALIS Starting Strong 2018. The chapter's first two sections focus on the reasons for adjudication and on what is adjudicated. The chapter then presents the criteria used to assess the data quality and briefly describes the final adjudication process. The chapter ends with a description of the recommended usage rating for each of the TALIS Starting Strong 2018 participating countries and populations.

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

10.1. The general adjudication process during TALIS Starting Strong 2018

This chapter covers the adjudication of TALIS Starting Strong 2018.

The TALIS Starting Strong 2018 International Consortium was responsible for providing participating countries and the OECD with databases containing data that (1) inform policy relevant to participating countries; (2) allow international comparisons; (3) allow the development of indicators; (4) contain valid and reliable data obtained using rigorous methods; (5) provide estimates that can be interpreted across participating countries; and (6) were obtained in a timely and resource-conscious manner.

It is thus the task of the consortium to demonstrate and document that the survey's processes were designed and implemented to meet these requirements, and to alert the participating countries and the OECD to shortcomings or limitations affecting use of any data elements.

The chapter begins by defining and describing the adjudication process followed during TALIS Starting Strong 2018. The elements that were presented for adjudication are then listed and commented on in turn, while the criteria used to assess the quality of each element are given in the third section. The fourth section provides the practical meaning of each rating. It also presents the outcomes of the adjudication, that is, the recommended rating for each participating country and each population of interest in TALIS Starting Strong 2018. Annex G of this report provides the full unweighted and weighted participation rates.

10.1. What was adjudicated during TALIS Starting Strong 2018?

The basic principle guiding the adjudication was to determine, for each participating country and for each of the TALIS Starting Strong 2018 populations of interest, whether the data released to the participating countries and to the OECD were "fit for use", as intended.

To establish fitness for use, a number of quality assurance processes were designed and activated throughout the survey process. Some processes relied on expert advice and opinion, some relied on qualitative information and learned judgement, and some relied on quantitative information. One such example is the international quality control programme conducted by the consortium in each participating country and described in Chapter 7 of this technical report.

In general, the overall quality of the survey implementation and the data yielded were high. However, several issues arose throughout the survey, for example during national adaptations or with respect to the extension of the collection window. The consortium clarified and discussed each issue with the concerned participating country and sometimes sought the advice of the OECD Secretariat during this process. When discussion produced a solution that was acceptable to the participating country and complied with the TALIS Starting Strong technical standards, the consortium considered the issue settled and were satisfied that it would have no impact on data quality. Some problems were discovered only during data processing, weighting or scaling. The impact of each of these issues on data quality, that is, their potential to limit the utility of the data, was assessed and documented.

The adjudication session held at the IEA Hamburg in September 2018 in the presence of representatives of the OECD Secretariat (a second session was held online in January 2019 to accommodate countries participating according to the Southern Hemisphere survey

schedule; see Chapter 6) submitted each individual dataset to the same scrutiny. Each dataset was a combination of participating country, population of interest (ISCED level 02 or services for children under three years of age) and population (centre leaders and centres, or staff, respectively). The member of the consortium responsible for a particular key step in the survey presented, discussed and assessed any unresolved issues arising out of the step that might reduce the final fitness for use of a particular dataset.

Outstanding and unresolved situations that could diminish the overall utility of a dataset could arise from any step in the survey process. For example:

- *Questionnaire adaptation to national context:* completeness of questionnaires in terms of including all introduction texts and questions in the international source version as well as the corresponding notes, instructions, response categories and coding schemes; questions removed or modified without agreement
- *Translation and verification:* national translations of items that would make comparisons difficult; translation verifier's report of low-quality translation; high number of rejections of verifier's suggested improvements to the national translations
- *Quality of the sampling frame (before sampling and confirmation at weighting):* measure of size that seems disproportionate to what is known of the participating country; missing data elements for some establishments or centres.
- *Handling of out-of-scope units and refusal units:* replaced unit found to be out of scope
- *Handling of replacement and "volunteer" units:* replacement centres participated but the original centre is also on the database
- *Staff rosters and within-centre sampling:* apparent incompleteness; apparent biased selection of staff sample
- *Main survey administration:* compliance with technical standards, survey operation procedure units, rules and guidelines; failure to administer the data collection within the agreed-upon data collection window; failure to administer the questionnaires according to the tracking form.
- *Data collection (paper and online modes):* missing records (physical or electronic)
- *Data cleaning and editing:* post-collection call-backs to national project managers (NPMs)
- *Quality observers' reports:* differences found between documentation of national decisions on translation verification feedback and the actual implementation in the final national instruments; reports of national translations viewed as unnatural and non-fluent; reported deviations in sampling procedures for listing all eligible participants in a centre; failure to administer questionnaires according to tracking form; any other observed deviation from standardised administration procedures
- *Weighting:* issues with miscoded centres, "volunteer" centres or staff; rough comparisons with sampling frame information; call-backs to NPMs; estimates of population size not matching information on sampling frame; centre listings too short; centre ID with no matching unit on the sampling frame

- *Participation rates of centre leaders and staff members:* weighted and unweighted rates strikingly different; participation rates computed within 1% of a rating borderline.

If no outstanding and unresolved issue was raised, the adjudicators established the recommended rating in accordance with the technical standards and based on participation rates (see Table 10.1 and also Chapter 9).

10.2. The criteria for assessment

As a general criterion, the consortium members at the adjudication considered that any problem that had been satisfactorily resolved was no longer a problem and therefore did not need to be discussed. The consortium member mainly responsible for the issue at hand presented the issue if unusual circumstances made that problem stand out.

In those instances where a problem had been only partially resolved, the consortium gained an impression of its impact on data quality by working through the following questions:

- What is the problem?
- What solutions have been tried and failed?
- What solutions have been partially worked through and to what extent?
- Is the perceived impact such that words of caution should be issued to users?
- Is the problem likely to compromise or limit international (or national) comparisons?
- Is more information from participating countries needed to assess the issue in full?

Once each survey process had been assessed, the consortium members formulated a recommended rating that accounted for the participation rates and for any unresolved issue.

10.3. Recommended usage ratings for participants

For easy reference, the following two adjudication tables (Table 10.1 and Table 10.2) found in the TALIS Starting Strong 2018 technical standards are replicated with minor changes here.

Table 10.1. Adjudication rules for centre or centre leader data in TALIS Starting Strong 2018

Centre participation (returned leader questionnaires)		Risk of centre non-response bias	Rating	Reference
Before replacement	After replacement			
⩾ 75%	⩾ 75%		Good	
⩾ 50% but < 75%	⩾ 75%		Fair	A
	⩾ 50% but < 75%	Low	Fair	C
		High	Poor	D
< 50%			Insufficient	

Note: Chapter 5 of this report discusses the centre-replacement procedure.
Source: OECD TALIS Starting Strong 2018 technical standards (see Annex B).

Table 10.2. Adjudication rules for staff data in TALIS Starting Strong 2018

Centre participation (minimum of 50% staff participation)		Staff participation after centre replacement	Risk of staff non-response bias	Rating	Reference
Before replacement	After replacement				
⩾75%	⩾75%	⩾75%		Good	
		⩾ 50% but < 75%		Fair	A
⩾ 50% but < 75%	⩾ 75%	⩾ 75%		Fair	B
		⩾ 50% but < 75%	Low	Fair	C
			High	Poor	D
⩾ 50% but < 75%	⩾ 50% but < 75%			Poor	E
< 50%	⩾ 75%			Poor	F
< 50%	< 75%			Insufficient	

Note: Chapter 5 of this report discusses the centre-replacement procedure.
Source: TALIS Starting Strong 2018 technical standards (see Annex B).

The following bulleted list provides a simple guide aimed at helping data users appreciate the limitations on data use or quality:

- *Good:* the participating country's data can be used for all reporting and analytical purposes and should be included in international comparisons.
- *Fair (line A):* national and subnational estimates can be produced; some staff characteristics may suffer from larger standard errors, hence the warning "Fair" but with no additional warnings to users appearing necessary.
- *Fair (line B):* national and subnational estimates can be produced; some subnational estimates may be of lower precision (larger standard errors) if sample size is locally low, hence the warning "Fair" but with no additional warnings to users appearing necessary.
- *Fair (line C):* national and subnational estimates can be produced; some subnational estimates may be of lower precision (larger standard errors) if sample size is locally low, hence the warning "Fair" but with the possible inclusion of a note on data quality pointing out the outcome of the non-response bias analysis (NRBA); since centre participation is somewhat lower than under (B), comparing subnational estimates should be done with care, as some of those results were based on just a few centres; comparing small subnational estimates with similar groups from other participating countries is unlikely to uncover any statistically meaningful differences because the standard errors are probably too large.
- *Poor (line D):* in addition to the warnings issued for the previous category, a note should warn users of indications of non-response biases in some estimates; comparisons of subnational estimates should be limited to the groups with the larger sample sizes; at this point, the sample represents between 37% and 56% of the workforce because it includes a relatively small sample of centres; comparisons with similar groups in other participating countries are not encouraged.
- *Poor (line E):* subnational estimates are not recommended; there should be a note pointing out the difficulty of obtaining a representative sample of centres.

- *Poor (line F):* limitations similar to those of line E, but there should be a note pointing out the difficulty of obtaining at least 50% participation of the selected sample of centres; risks a non-representative sample of centres.
- *Insufficient:* weights should not be calculated for any official tabulations; hence, data should not be incorporated[1] in international tables, models, averages etc.

The final ratings thus depended on participation rates before and after replacements and on the apparent severity of the non-response biases.

The next four tables (Table 10.3,Table 10.4, Table 10.5, and Table 10.6) present the recommended rating for each participating country, by ISCED level and population.[2]

Table 10.3. Services for children under the age of three years: centre leader participation rates and recommended ratings

Participating country	Leader participation before replacement (%)	Leader participation after replacement (%)	Recommended rating
Denmark[a]	35.4	47.5	Insufficient
Germany[b]	50.7	57.2	Poor
Israel[c]	93.3	97.4	Good
Norway[d]	66.8	92.6	Fair

Notes:
A centre was deemed a participating centre if the centre leader returned his or her questionnaire. Chapter 5 of this report discusses the centre-replacement procedure.
[a] *Denmark:* low response rates during the survey may have resulted in bias in the estimates reported, thus limiting comparability of the data. The data collection period has been extended.
[b] *Germany:* there has been under-coverage in the sampling frame. To increase participation rates, Germany's national study centre re-opened their data collection for a few weeks in October/November 2018. The non-response bias analysis could not show that there was no evidence of a high risk of non-response bias.
[c] *Israel:* the data collection period has been extended.
[d] *Norway:* home-based settings were excluded; the exclusion rate exceeded 5%.

Table 10.4. Services for children under the age of three years: staff participation rates and recommended ratings

Participating country	Centre participation before replacement (%)	Centre participation after replacement (%)	Staff participation in participating centres (%)	Overall staff participation (%)	Recommended rating
Denmark[a]	31.7	43.7	86.1	37.6	Insufficient
Germany[b]	50.3	57.0	89.7	51.1	Poor
Israel[c]	90.7	95.3	97.4	92.8	Good
Norway	67.1	91.1	86.5	78.8	Fair

Notes:
A centre was deemed a participating centre if at least 50% of the selected staff members returned their questionnaires. Chapter 5 of this report discusses the centre-replacement procedure.
[a] *Denmark*: low response rates during the survey may have resulted in bias in the estimates reported, thus limiting comparability of the data. The data collection period has been extended.
[b] *Germany*: there has been under-coverage in the sampling frame. To increase participation rates, Germany's national study centre re-opened their data collection for a few weeks in October/November 2018. The non-response bias analysis could not show that there was no evidence of a high risk of non-response bias. In two centres, staff listings were found to be incorrect; these centres were therefore categorised as "non-participant".
[c] *Israel:* the data collection period has been extended.

Table 10.5. ISCED level 02: centre leader participation rates and recommended ratings

Participating country	Leader participation before replacement (%)	Leader participation after replacement (%)	Recommended rating
Chile	94.8	98.6	Good
Denmark[a]	40.9	55.3	Insufficient
Germany[b]	52.8	69.0	Poor
Iceland	75.4	75.4	Good
Israel[c]	97.4	98.3	Good
Japan	87.2	98.6	Good
Korea[d]	61.5	76.1	Fair
Norway	64.6	83.8	Fair
Turkey[e]	99.8	99.8	Good

Notes:
A centre was deemed a participating centre if the centre leader returned his or her questionnaire. Chapter 5 of this report discusses the centre-replacement procedure.
[a] *Denmark:* low response rates during the survey may have resulted in bias in the estimates reported, thus limiting comparability of the data. The data collection period has been extended.
[b] *Germany:* Home-based settings were excluded; the exclusion rate exceeded 5%. To increase participation rates, Germany's national study centre re-opened their data collection for a few weeks in October/November 2018. The non-response bias analysis could not show that there was no evidence of a high risk of non-response bias.
[c] *Israel:* ultra-orthodox centres were excluded after the survey because of the low participation rates in this sector; the exclusion rate therefore exceeded 5%.
[d] *Korea:* the data collection period was extended.
[e] *Turkey:* centres belonging to the Ministry of Family and Social Issues were excluded; the exclusion rate therefore exceeded 5%.

Table 10.6. ISCED level 02: staff participation rates and recommended ratings

Participating country	Centre participation before replacement (%)	Centre participation after replacement (%)	Staff participation in participating centres (%)	Overall staff participation (%)	Recommended rating
Chile	94.0	98.3	98.8	97.1	Good
Denmark[a]	32.8	45.7	84.3	38.5	Insufficient
Germany[b]	50.3	57.0	89.7	51.1	Poor
Iceland[c]	87.0	87.0	84.8	73.8	Good
Israel[d]	95.7	96.7	97.4	94.1	Good
Japan	87.2	98.6	99.6	98.2	Good
Korea[e]	59.7	75.4	96.8	72.9	Fair
Norway	60.7	78.7	83.8	65.9	Fair
Turkey[f]	99.7	99.7	99.6	99.2	Good

Notes:
A centre was deemed a participating centre if at least 50% of the selected staff returned their questionnaires. Chapter 5 of this report discusses the centre-replacement procedure.
[a] *Denmark:* low response rates in the survey may have resulted in bias in the estimates reported, thus limiting comparability of the data. The data collection period has been extended.
[b] *Germany*: To increase participation rates, Germany's national study centre re-opened their data collection for a few weeks in October/November 2018. The non-response bias analysis could not show that there was no evidence of a high risk of non-response bias.
[c] *Iceland:* comparisons between teachers and assistants are not possible because participants were not familiar with the distinction between these roles.
[d] *Israel:* ultra-orthodox centres were excluded after the survey due to low participation rates in this sector; the exclusion rate therefore exceeded 5%.
[e] *Korea:* the data collection period was extended.
[f] *Turkey:* centres belonging to the Ministry of Family and Social Issues were excluded; the exclusion rate therefore exceeded 5%.

Notes

1. At their meeting held in Paris in November 2018, the TALIS Starting Strong 2018 Technical Advisory Group recommended that data from participating countries that had not reached 50% participation should be nonetheless weighted and displayed in tables.

2. Table 10.3 to Table 10.6 display the participation-rate estimates that were the most favourable for the adjudication rating. The most favourable estimates could have been weighted or unweighted depending on the characteristics of the country, the staff and centre leader population and the population of interest (ISCED level 02 or services for children under three years of age).

Chapter 11. Validation of scales and construction of scale scores

To enable reporting on a latent trait (sometimes referred to as a construct) or other abstract trait, some questions in the TALIS Starting Strong 2018 questionnaires were combined into an index or scale. This chapter explains how the indices were created and describes the methodology used to validate scales and construct scale scores. It details latent trait evaluation and the procedure involved in computing scale scores. It also illustrates the implications of the evaluation results for using scale scores in further analyses. The chapter furthermore describes the possibilities and limitations of using scale scores for cross-country comparisons and presents each scale in more detail together with its statistical properties.

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

11.1. Overview

The TALIS Starting Strong 2018 questionnaires include numerous items pertaining to, for example, the characteristics of early childhood care and education (ECEC) settings and the attitudes, perceptions and practices of ECEC centre leaders and staff. Use of suitable statistical procedures makes it possible to combine responses to these items into indices or scales. In line with TALIS 2018 and its previous cycles, two types of combinations were considered for TALIS Starting Strong 2018:

- simple indices (e.g. ratios), constructed through simple arithmetical transformations or by recoding one or more items
- scale scores, derived using latent modelling within the framework of confirmatory factor analysis (CFA).

CFA treats items as indicators of the unobservable personal characteristics of survey respondents, such as self-efficacy or beliefs, and uses combinations of items to develop a model of the latent construct. Once the stability of a latent construct model is tested, scale scores are estimated that serve as numeric values for the latent constructs.

While simple indices enhance the analysis of observable (manifest) characteristics, such as student–teacher ratio, scales enable analysis of non-observable (latent) characteristics such as attitudes or other personal traits. This chapter begins by outlining the procedures used to compute simple indices. It then describes the procedures involved in scale evaluation and scale score estimation. A detailed description of the scales, including the items used to compose each scale and the results of scale evaluation, follows. The chapter ends by exploring the implications of these results for further analyses, especially in relation to cross-country comparisons.

11.2. Computation of simple indices

This section describes the construction of simple indices that are part of the publicly available international database. The indices were constructed through arithmetical transformation or recoding one or more of the items in the survey's questionnaires. Chapter 12 provides more details on this process.

11.2.1. Ratios and recoded variables

Index of ECEC centre diversity practices

The index of ECEC centre diversity practices was calculated at the staff level according to the information derived from staff members' responses to a question about the frequency of their daily interactions with the children in their centre. These interactions included those that involved using diverse learning materials with the children within the target populations. The index (DIVPRAC) was derived by adding together the information in all four question categories (SS1G34A–D), but it was calculated only if all the staff members answered all four items (i.e. A–D) of the question (and assuming they were administered).

Sum of all staff members working with the target group

The sum of all staff members working with the target group was calculated at the staff-level by drawing on the information derived from staff members' responses to the question that asked them to state how many staff were working with the target group of children on

their (staff members) last working day before the day that they responded to the questionnaire. The sum of staff working with the target groups (SUMSTAFFTG) was derived by adding up the number of all staff working with children in the target group (SS1G35A–G). However, the value was only calculated if staff answered all of the first five items (i.e. A–F) of the question (and assuming they were administered). The last item (G) "Other <ECEC staff>" was not considered in the same way and did not influence the decision on the need to calculate the SUMSTAFFTG. The person filling in the questionnaire was asked not to include herself/himself in the number of staff working with the target group. Consequently, "+1" is added to the sum of staff members working with the target group.

Sum of boys and girls in the target group

The sum of all boys and girls in the target groups was calculated at the staff-level based on the information derived from the staff members' responses to the question about the number of boys and girls in the target group on their (staff members) last working day before the day they answered the questionnaire. The sum of boys and girls in the target group (SUMCHILD) was derived by adding up the number of boys and girls in that group (SS1G37A–B).

Headcount of children per adult in the target group

The headcount of children per adult was calculated at the staff-level according to the information derived from staff members' responses to questions about the total number of all staff members working with the target group of children (headcounts) and the number of all boys and girls in the target group (headcounts). The variable (HEADCHPAD) was derived by dividing the sum of all boys and girls in the target group (SUMCHILD) by the sum of all staff members working with the target group (SUMSTAFFTG).

Sum of all staff members – categorised variable

The sum of all staff members at the centre level was calculated through reference to the information that ECEC centre leaders provided when asked to state the number of staff currently working with all age groups of children in their respective centres. The index thus reflects the overall sum of staff working in each centre rather than being restricted to the number of staff working with the survey's target population of children.

The sum of staff (SUMSTAFF) was derived by adding up the number of all staff working in the centre (SL1G17A–G). This value was only calculated if the centre leader answered all of the first five items (i.e. A–F) of this question (and assuming they were administered). The last item (G) "Other <ECEC staff>" was not considered in the same way and did not influence the decision on the need to calculate the SUMSTAFF or not.

The variable SUMSTAFF was then categorised. The new index (SUMSTAFGR) was calculated by recoding the values from SUMSTAFF into three categories. The first category included all centres with fewer than 15 staff members, the second category included all centres with between 15 and 29 staff members and the third category all centres with 30 or more staff members. The variable SUMSTAFF is not available in the public use database.

Proportion of staff members with different roles – categorised variables

The proportion of staff members with different roles in the centre was calculated through reference to the information that ECEC centre leaders provided when asked to state the number of staff currently working with all age groups of children in their respective centres. The resulting three indices reflect the overall proportion of staff with different roles in each centre rather than being restricted to staff working with the survey's target population of children.

The first index (LTEACHGR) was derived by adding up the number of leaders and the number of teachers working in the centre (SL1G17A and SL1G17B) and then dividing this number by SUMSTAFF. The result was then categorised so that LTEACHGR included three categories. The first category included all centres where the proportion of leaders and teachers in the total staff is 50% and below, the second category included all centres where the proportion of leaders and teachers is between 50% and 75% and below and the third category included all centres where the proportion of leaders and teachers is greater than 75%.

The second index (LASSISTGR) was derived by dividing the number of assistants working in the centre (SL1G17C) by SUMSTAFF. The result was then categorised so that LASSISTGR included three categories. The first category included all centres where the proportion of assistants in the total staff 25% and below, the second category included all centres where the proportion of assistants is between 25% and 50% and below and the third category included all centres where the proportion of assistants is greater than 50%.

The third index (LSTOTH) was derived by adding up the number of staff for individual children, the number of staff for special tasks, the number of interns and the number of "other" staff working in the centre (SL1G17D–G) and then dividing this number by SUMSTAFF. The result was then categorised so that LSTOTH included two categories. The first category included all centres where the proportion of these four types of staff members in the total staff is 25% and below, the second category included all centres where the proportion of these four types of staff members is greater than 25%.

Proportion of staff who left the centre – categorised variable

The proportion of staff who left the centre was calculated through reference to the information that ECEC centre leaders provided when asked to state the number of staff who permanently left the ECEC centre during the 12 months prior to the survey (SL1G18B). This number was then divided by SUMSTAFF and categorised to form a new index (LSTAFFLEFT) with two categories. The first category included all centres where 10% or fewer staff left the centre and the second category included all centres where more than 10% of staff left the centre.

Leader age groups – categorised variable

The leader age group variable was calculated by drawing on the responses of centre leaders to the question about their age (SL1G02). The new index (LAGEGR) was calculated by recoding the values from the original question into six categories. The first category included all leaders under the age of 50 and the second category included all leaders age 50 and older.

Staff age groups – categorised variable

The staff age group variable was calculated from the responses of staff to the question about their age (SS1G02). The new index (SAGEGR) was calculated by recoding the values from the original question into six categories. The first category included all staff under the age of 20, the second category included all staff between 20 and 29, the third category all staff between 30 and 39, the fourth category all staff between 40 and 49, the fifth category all staff between 50 and 59 and the sixth category all staff 60 years of age and older.

Leader education – categorised variable

The leader education variable was calculated from the responses of centre leaders to the question about the highest level of formal education they completed (SL1G03). The new index (LISCED) was calculated by recoding the values from the original question into two categories. The first category included all leaders whose highest level of formal education was ISCED level 5 or below and the second category included all leaders whose highest level of formal education was ISCED level 6 or above.

Leader experience – categorised variable

The leader experience variable was calculated from the responses of centre leaders to the question about their year(s) working as an ECEC centre leader in total (SL1G05B). The new index (LEXPGR) was calculated by recoding the values from the original question into three categories. The first category included all leaders with less than five years of experience, the second category included all leaders with between five and nine years of experience and the third category included all leaders with ten or more years of experience.

Number of enrolled children – categorised variable

The number of enrolled children variable was calculated through reference to the responses of centre leaders to the question about their centre's current enrolment level (SL1G19). The new index (NENRCHLD) was calculated by recoding the values from the original question into two categories. The first category included all centres with fewer than 100 children; the second category included centres with 100 and more children.

Centre funding from non-governmental organisations and/or benefactors – combined variable

Whether a centre received funding from either non-governmental organisations (including religious institutions and employers) or benefactors, donations, bequests, sponsorships, parent/guardian fundraising (including subsidies through non-profit ECEC providers) was calculated through reference to the responses of centre leaders about sources of funding for the centre (SL1G14C and SL1G14D). The combined variable SL1G14CD is coded as "yes" only if leaders indicated their centres received at least one of these types of funding. Otherwise, SL1G14CD is coded as "no" if leaders indicated their centres received none of these two types, and it is coded as "don't know" if leaders indicated they did not receive funding from one of the types but did not know about the other one.

11.3. Scaling procedures

The primary aim of the TALIS Starting Strong 2018 was to collect robust and rich information about the characteristics of centre staff and leaders as well as about their

centres. Many of the specific personal traits, for instance, staff members' attitudes and self-efficacy, could not be observed directly, but only through expressed opinions or intended and observed behaviour. Such traits are considered latent, that is, not directly observable. In large-scale studies such as the TALIS Starting Strong, sets of items drawn from a study's instruments are used to estimate these latent traits. The instruments used in the TALIS Starting Strong were the staff questionnaires and the leader questionnaires, the items of which were designed to reflect specific facets of the envisaged latent traits. The procedure used to combine responses to these questions into a single scale score representing the latent construct of interest is called *scaling*.[1]

11.3.1. Scale development and methods

The TALIS Starting Strong 2018 scale development work began with a theoretical identification of items that seemed indicative of the specified latent constructs. These identifications were based on lessons learned from the TALIS series of surveys, research theories drawn from relevant fields, and expert knowledge on item and scale construction. These steps served as an initial validity check of the scales (Messick, 1995[1]). The data intended to represent constructs and to be used for scale score computation were subjected to extensive quality checks that included the use of item-level statistics. Checks focused on the distribution of missing data, number of responses per category and the shape of response distributions, and reliability analysis.

Field trial data were used to evaluate the latent constructs and then modify them for the main survey. The TALIS Starting Strong 2018 International Consortium, the survey's Questionnaire Expert Group (QEG) and the OECD used the results from these field trial analyses to make decisions about item and scale modifications. In some instances, the decision was not to create scale scores but instead to report the values at the level of individual items only. In addition, some items were removed from the instruments after the field trial to reduce the length of the questionnaires (for details on this matter, see Chapter 3 of this report). All decisions considered the state of instrument development at the time of the field trial and the low number of cases in the field trial data.

For the main study data, the analysis of items and scales involved another thorough evaluation of the extent to which each scale measured its construct consistently (construct reliability) (Lyubomirsky and Lepper, 1999[2]) and the extent to which the scale measured the same construct across the participating countries (construct invariance) (Cheung and Rensvold, 2002[3]; Vandenberg and Lance, 2000[4]). The processes involved in evaluating the scales and constructing scale scores during the main survey encompassed the following steps:

1. Descriptive and internal consistency analyses
    a. item analysis of missingness
    b. item analysis of distribution
    c. item analysis of item-total correlation
    d. initial reliability checks.
2. Model analysis: confirmatory factor analysis (CFA)
    a. model analysis of the predefined construct, involving a joint analysis of data from all participating countries (CFA on a pooled sample[2])

   b. model analysis at the country level (separate CFA for both the ISCED level 02 and the U3 (if applicable) populations in each country).

3. Measurement invariance testing
   a. comparability of the constructs within populations across countries (i.e. measurement invariance testing conducted across countries within each population)
   b. comparability of the constructs within countries across populations (i.e. measurement invariance testing conducted across populations within each country).

4. Final scale modelling
   a. modelling of the constructs to account for the invariance levels evident for each cross-country within population measurement invariance result and each cross-population within country measurement invariance result[3]
   b. reliability analysis of the final models.

5. Scale score computation
   a. scale score estimation
   b. scale score standardisation
   c. composite scale scores.

The following (sub)sections of this chapter describe the steps mentioned above in more detail. The subsection titled *Descriptive and internal consistency* analyses of the initial descriptive analysis and the reliability analysis from steps 1d and 4b. The next subsections, *Model analysis* and *Measurement invariance testing: across countries and across populations*, describe procedures from the model analysis (steps 2a and 2b) and the measurement invariance testing (steps 3a and 3b), respectively. The subsection *Final scale modelling* describes how the initial analyses led to the final models and parameter estimates (4a). The section *Scale scores computation* is divided into three subsections. The first describes the scale score estimation from step 5a, the second the scale score standardisation from step 5b, and the third the scale score computation for special types of scales, namely composite scales (step 5c).

11.3.2. Scale evaluation

Descriptive and internal consistency analyses

As initial checks, items were analysed in relation to missingness, distribution and corrected item-total correlation. In addition, Cronbach's alpha, as an initial reliability check, was calculated and evaluated. Where applicable, items were reverse coded, which meant that the higher the value on an item the higher the level of the latent construct.

Scale reliability was tested by examining the internal consistency of the scale. The weighted omega statistic was used to measure reliability of the scales because it does not assume equal factor loadings (tau-equivalent measurement model) in the measurement model (Viladrich, Angulo-Brunet and Doval, 2017[5]; Zhang and Yuan, 2016[6]). Therefore, compared with Cronbach's alpha, weighted omega is relatively unbiased when items exhibit unequal factor loadings for a single scale, which is true of all scales in the TALIS Starting Strong 2018.

The weighted omega value is equal to the square of factor score determinacy (FSD) obtained from the final scale model, which is an estimated correlation between the latent variable and the items (Beauducel, Harms and Hilger 2016[2]). However, Mplus FSD is used only to calculate models with no binary items. Consequently, the reliability statistic reported for scales measured by binary items is Cronbach's alpha obtained from the initial reliability checks (step 1d).

To estimate the reliability of multidimensional scales, the omega coefficients from each scale's subscales were used to calculate stratified coefficient alpha as follows:

$$\gamma_{STRAT,\alpha} = 1 - \frac{\sum_i \sigma_i^2 (1 - \gamma_i)}{\sigma_c^2}$$

where, $\gamma_{STRAT,\alpha}$ is the reliability of the composite/multidimensional scale; σ_i^2 is the variance of the *i*th subscale, or the variance of the standardised factor scores of this subscale; γ_i is the reliability of the *i*th subscale or the omega for this subscale; and σ_c^2 is the variance of the composite/multidimensional scale or the variance of the sum of the standardised factor scores of the subscales (He, 2010[7]).

Model analysis

Confirmatory factor analysis (CFA) was conducted to evaluate how well the empirical data reflected the predefined latent construct. CFA allows inference on the scale from the scale items, as it establishes associations between the two. The association between each item and the scale is described in a regression line (OECD, 2014[8]). Use of model fit indices as evaluation criteria (see further details in Table 11.1) make it possible to assess the specified (or theoretical) model of each scale with respect to its alignment to the empirical data (Hu and Bentler, 1999[9]).

Various procedures can be used to estimate scale scores, including computation of a sum or mean score over all items that measure the same construct; computations based on classical test theory (CTT) or structural equation modelling (SEM) using CFA; and computation of the person parameters based on item response theory (IRT). Although the results from these methods are typically highly correlated, the methods are not completely congruent, and each method has its respective advantages and disadvantages.

As with the current and past TALIS cycles, scale score computation based on CFA was used in TALIS Starting Strong (OECD, 2010[10]; OECD, 2014[8]). This method continues to be one that has a solid scientific basis and offers great flexibility, given the invariance results for the scales. Also, when employed with certain modelling software (Mplus[4]), this form of computation is well equipped to deal with missing values.

Analysis for the TALIS Starting Strong 2018 was based on the general SEM framework, where CFA is a specific type of model classified within this framework (Schreiber et al., 2006[11]). All constructs with ordinal response categories were scaled via continuous CFA (estimated using robust full-information maximum likelihood estimation on the matrix of Pearson's correlations).[5] Constructs with binary items were scaled via categorical CFA modelling (estimated using robust WLSMV[6] estimator on the pair-wise matrix of tetrachoric correlations). Design weights and replicate weights were used for all analyses, and weights were rescaled so that each country participating in the TALIS Starting Strong 2018 contributed equally to the estimates.[7]

A number of items in the survey were identical to their counterparts in the TALIS cycles and were therefore used to construct latent scale scores. Item selection for the TALIS

Starting Strong 2018 emphasised maximising the overlap of items between TALIS 2018 and the TALIS Starting Strong 2018 to allow for the possibility of item-level comparisons across studies. However, the scaling procedure does not allow for direct comparisons of scale scores across studies.

To evaluate the overall performance of the scale models, the first step was to conduct CFA on a pooled sample composed of data from all countries for ISCED level 02 (the TALIS Starting Strong's core target population surveyed by all participating countries). The initial examination of the pooled CFA models of each scale was based on model fit indices.[8] The models that passed the cut-off criteria (see Table 11.1) proceeded to further steps, while improvements were made to the models of the scales that originally failed to pass the cut-off values. Model improvements included, for example, changes to the model structure (e.g. item exclusion, inclusion of residual covariances between specific items). These modifications were included in the models for all following steps.

Most pooled model improvements were suggested by the modification indices produced by the CFA implementation software. These indices reflect the approximate change to χ^2 if certain fixed/constrained parameters are freely estimated (Brown, 2006, p. 119[12]). To maintain a balance between improving the model and keeping the model as parsimonious as possible, the only model re-specification implemented and tested was the one resulting in the biggest (or, for some cases, second biggest) change to χ^2. The same reason applied to the implementation (in rare cases) of cross-loadings in the multidimensional scales, but these were generally avoided if other alternatives could significantly improve the model.

In certain cases, additional improvements were made to the models during testing for measurement invariance. Changes were made programmatically to the model, which meant that changes should be ones that did not (1) bias the results and (2) change the content of the scale construct. Such changes included, for example, fixing the negative residual variance for items in some scales to be positive and close to zero.

Most of the χ^2-based model modifications included correlations between single items (residual covariance). All scale modifications were accompanied by plausibility checks conducted by the scaling team, as well as by the QEG experts, thus acknowledging that changes to the model structure had relevant implications for scale content. In other words, items were only correlated if the content of the items reflected a plausible and substantive correlation.

In cases where the model did not show acceptable fit, even after improvements, exploratory factor analysis (EFA) was used to analyse scale dimensionality. If the analysis revealed multidimensionality, one of the following modifications was applied:

1. reducing the number of items in order to build a unidimensional construct as indicated by the pattern matrix provided by the oblimin rotation of the principal axis factoring[9] solution
2. splitting the scale into two constructs as indicated by the EFA results
3. modelling the scale as a multidimensional, hierarchical construct, as indicated by the EFA results, thus obtaining one composite scale score
4. keeping the scale as unidimensional during further analysis if there were strong content-related or other reasons for doing so; the latter was relevant in those instances where theoretical arguments supportive of multidimensionality were absent.

The multidimensional scales that were originally treated as unidimensional were re-specified and re-evaluated. Constructs that only marginally failed to pass the cut-off criteria were kept for further analysis but eventually dismissed if additional scale modifications did not succeed (see the Excluded scales sections of this chapter for staff and leaders). The use of EFA for model improvement provided meaningful enhancement of one latent construct, Sources of work stress. The scale was modelled as a multidimensional, hierarchical construct.[10]

Scales that could not be improved through modifications were dropped from the analysis. A deleted scale typically had a fragmented internal structure in which items did not fit together, meaning that the scale was either composed of several factors or poorly defined. The assumption in relation to the latter case was that the items failed to measure the predefined construct and that the content validity of the scale was therefore low. Decisions made at this stage were discussed with the TALIS Consortium and the QEG.

During the second step, CFA models were tested via use of country-level data from each of the analysed populations (in all countries ISCED level 02 and U3, as applicable[11]). Improvements to the pooled models were applied to these models. In total, 11 single-country single-population level models were analysed for each scale.[12]

The evaluation procedure for these models was very similar to that for the pooled model. If a single-country, single-population model failed to meet the fit index cut-off criteria (Table 11.1), additional improvements for that specific population were implemented. If one of these scale models could not be improved, then that single-country, single-population was removed from further parameter estimation and included in the scale score construction involving use of the fixed parameters from the final scale model. Table 11.1 provides information on the cut-off criteria for the CFA model fit evaluation (Brown, 2006[12]; Chen, 2007[13]; DeVellis, 2003[14], pp. 94-96; Hoyle, 2014[15]). It is important to stress that these statistical criteria were used for the decision-making process that was based on an iterative process involving content-related considerations between the IEA scaling team and the QEG members.

Table 11.1. Cut-offs for CFA model evaluation for the TALIS Starting Strong 2018

Statistic/index	Description	Cut-offs
Missing total (%)	The percentage of missing values (out of those administered).	More than 80% of valid answers at the item level for each participating country
Cronbach's alpha	Scale reliability index: internal consistency index. Higher values indicate greater internal consistency.	0.600–0.699 (acceptable) ≥0.700 (good)
Omega	Scale reliability index: factor score determinacy (FSD) squared. Higher values suggest greater scale reliability.	0.600–0.699 (acceptable) ≥0.700 (good)
Stratified coefficient alpha	Scale reliability index: internal consistency index for composite/multidimensional scales. Higher values indicate greater internal consistency.	0.600–0.699 (acceptable) ≥0.700 (good)
Corrected item-total correlation	Correlation between the responses to an item and the observed total scores on all other items in the scale (the sum of all the other items). The correlation ranges between -1 and +1. Higher value suggests that the item has better discrimination power.	≥0.300 (acceptable)

CFI (comparative fit index)	Index of the model–data goodness of fit. This compares the targeted factor structure model and the baseline model (all relationships fixed to zero). Higher value indicates a better model fit.	≥ 0.900 (acceptable)
TLI (Tucker-Lewis index)	Index of model–data goodness of fit. This gives the distance between the targeted factor structure model and the baseline model as a proportion of the distance between the baseline and the target model. Higher value indicates a better model fit.	≥ 0.900 (acceptable)
RMSEA (root mean square error approximation)	Index of the model–data misfit. The value indicates the degree of model misspecification. It approaches 0 as the fit of the model improves.	≤ 0.080 (acceptable)
SRMR (standardised root mean square residual)	Index of the model–data misfit. The value indicates the degree of model misspecification in terms of the model average of squared residuals between the observed covariances and the model-implied covariances as well as differences in the observed and model-implied item intercepts. The value is provided in a standardised metric ranging from 0 to 1, with smaller values indicating better model fit.	≤0.060 (acceptable)
WRMR (weighted root mean square residual)	Residual-based fit index (experimental fit statistic). This index is suitable for models with varying variances of sample statistics and when sample statistics are on different scales. It is also used with categorical outcomes and was used in TALIS as the model–data misfit for categorical CFA. Smaller values indicate a better model fit.	≤0.900 (acceptable)
Standardised factor loadings	These indicate the strength of the relationship between each item and the latent scale.	0.450–0.600 (moderate) ≥ 0.600 (strong)

Notes:

1. Cronbach's alpha, omega, and stratified coefficient alpha are all different estimators of the same reliability/internal consistency; the criteria are therefore the same.
2. The cut-off criterion for the SRMR was less strict in TALIS 2013 (SRMR≤.1). To enhance alignment with the established cut-off criteria for model evaluation proposed in the literature (Hu and Bentler, 1999[4]; OECD, 2014[4]; Schermelleh-Engel, 2003[8]; Steiger, 1990[8]; Yu, 2002[8]), TALIS Starting Strong 2018 applied a stricter cut-off of SRMR. The cut-off was justified because the scale construction in TALIS Starting Strong 2018 benefitted from prior TALIS cycles as well as from the field trial.
3. When the intercepts are fixed to either a very high or low value, the SRMR can be misleading. In these cases, the variances may be very low, resulting in an extremely large SRMR value (Asparouhov, 2018[13]). Thus, decisions relating to the performance and the measurement invariance of the scale are based primarily on CFI, TLI and RMSEA. SRMR was used for model evaluation if other fit indices revealed inconsistent results and in models with varying intercepts (configural and metric models).

Treating missing items in scales

The TALIS Starting Strong 2018 used two versions of questionnaires for ECEC centre staff and leaders, with those versions depending on the population being targeted (ISCED level 02 or U3). A combined centre questionnaire was also used in centres with just one person, and a longer version (combined centre questionnaire plus) was used for Israel ISCED level 02, where centres consisted of one main teacher and assisting staff (i.e. no "leader"). These questionnaire versions contained (combined) suitable questions from the staff questionnaire and selected questions from the leader questionnaire. Some items deemed not suitable for the U3 populations were administered only to the ISCED level 02 populations. Also, one or more items were not administered in some countries that chose not to use them. Ability to analyse scales with items not administered in certain populations required specific adaptations to the procedure.

Any comparison of countries or education levels that uses multiple group confirmatory factor analysis (MGCFA) requires all groups to have variance-covariance matrixes of the same size. If this condition is not fulfilled, *phantom indicators* can be created in order to include all groups (countries and populations) in the scaling procedure. A phantom

indicator is a data-driven, randomly generated indicator that biases neither the results nor the model fit nor the parameters in any way, but simply allows groups with missing data to be included in the scale models.

Residual centring was used to create the phantom indicators in the TALIS Starting Strong, a process that produced random standard normal indicators orthogonal to all the other items within the group. In these cases, the models' degrees of freedom were also adjusted to correctly estimate the unbiased fit indices in the CFA models (Geldhof et al., (2013[14]), pp. 37–38[21]). Table 11.2 presents the groups for which phantom indicators were created.

Table 11.2. TALIS Starting Strong 2018 groups requiring phantom indicators

	Population	Scale	Item(s)
Staff			
	Israel (both)	S1COLL	SS1G23C
	U3 (all countries)	S1PRDEV	SS1G29D & SS1G29G
	U3 (all countries)	S1SOCIAL	SS1G41J
	Turkey (ISCED level 02)	S1PDIV	SS1G16Q
Leaders			
	Germany (U3)	S1LPANE	SL1G26D

Full datasets (with phantom indicators) were used throughout the work (described below) involved in evaluating the TALIS Starting Strong scales and constructing the scale scores. The only exception to this practice was the reliability analysis conducted at the country level, where the model was estimated for each country separately and there was no need to adapt the usual procedure. When interpreting the scale construct, readers need to consider that some constructs were based on fewer than the usual number of items for specific populations, as indicated in Table 11.2.

Measurement invariance testing: across countries and across populations

The TALIS Starting Strong 2018 data can be used to gain a cross-national perspective focused on comparing results across the participating countries. Of crucial importance during any comparison of scales across groups (such as education systems, participating countries, the ISCED level 02 and U3 populations) is making sure that the scales are equivalent in meaning in each of the groups being compared. The extent to which a scale offers comparability across groups varies, a situation that has direct implications for which statistical calculations (e.g. regression, correlation, mean comparisons) should be used to analyse the scale score.

The statistical procedure used to analyse the comparability of latent scales in different groups is called *measurement invariance testing* (Cheung, 1999[15]; Cheung and Rensvold, 2002[3]; Davidov, 2008[16]; OECD, 2014[8]; Steenkamp, J.-B. and E. M. Baumgartner, 1998[17]; Van de Vijver et al., 2019[18]). The TALIS Starting Strong 2018 scales were evaluated with respect to their equivalence (comparability) across participating countries and populations, within the CFA framework. For more details on the measurement invariance testing within the CFA framework, see OECD (2014[8]).

It is important to remember that the construction of latent scales is based on associations between several items and the underlying latent construct and the mean structure of the items. The latent construct, including its indicators (items), is reflected within a specified joint model, the *measurement model*. The measurement model specified within the CFA framework contains different parameters that are estimated along with the model: item

factor loadings, intercepts (or thresholds in the case of categorical CFA) and residual variances,[13] as well as latent means and variances (Davidov, E. et al., 2014[19]).

During the procedure involved in testing measurement invariance, different models are specified and compared to one another. The models differ from one another in terms of certain parameters being either unconstrained (i.e. freely estimated) between groups, which assumes lack of comparability across groups, or constrained (i.e. to the same value) between groups, which assumes comparability across groups. The degree to which the stricter model (i.e. a model that assumes equal parameters across groups), as compared to the less strict model (i.e. a model that assumes some degree of flexibility of parameters between groups), suits the data is evaluated via model fit indices and the direction (i.e. better or worse fit) and degree of change between the fit indices of each model. [14]

In the least restrictive model, all parameters are freely estimated separately for each group[15] (e.g. participating country), which means that the parameters are unconstrained and vary across these groups. The model implies that there is no comparability between groups because all parameters are group-specific and therefore no statistical comparisons are permitted.

As the number of models being estimated increases, they become gradually more restrictive (i.e. contain a greater number of parameters that are restricted to be equal across groups), resulting in increasingly equal measurement models with greater levels of comparability between and across the groups. As the models become more restrictive, the justification for statistical comparability and analysis between and across the groups (e.g. correlation or mean comparisons) increases. The levels of comparability are called *measurement invariance levels* and are specifically defined by the parameters that are restricted in the model representing each of the levels. The three most common levels of measurement invariance testing are (1) configural, (2) metric and (3) scalar (Cheung, 1999[15]; Cheung and Rensvold, 2002[3]; Davidov, 2008[16]; Davidov, E. et al., 2014[19]; OECD, 2014[8]; Steenkamp, J.-B. and E. M. Baumgartner, 1998[17]).

The lowest level of measurement invariance, the configural level, assumes that the underlying latent construct is specified by a particular configuration of items in all analysed groups in the same way. The *configural level of measurement invariance* applies when the construct is measured by the same items. This type of invariance implies that the structure of the construct indicated by the configuration of items is equivalent across participating countries. If a scale reaches only the configural level of measurement invariance, then any statistical method applied to compare the scale scores across groups will violate the basic assumption of the comparability of the measured construct. The comparability occurs at a conceptual level only, which means score comparability is not achieved. Therefore, results (e.g. correlations) from different groups can be discussed only through reference to each specific group.

The meaning of the scale is defined by the content of the questions respondents were asked and that were used to create the scale. If the strength of the associations (i.e. the magnitude of the regression parameters) is the same across groups, then the latent construct is assumed to have the same meaning. This is the second level of measurement invariance, the metric level. The *metric level of measurement invariance* applies when (1) the structure of the construct is the same across groups, and (2) the strength of the associations between the construct and the items (factor loadings) is equivalent across groups. Metric invariance makes it possible to claim that one unit of change in the construct will lead to the same amount of average change in the items that constitute the construct across different groups (e.g. participating countries). If a scale establishes the metric level of invariance, it can be

assumed that comparisons of correlational analyses (such as correlation or regression analysis) are free of cross-group bias. Of note with respect to TALIS Starting Strong, this level of measurement invariance testing was omitted for scales with binary indicators because use of the Mplus software with these models presented identification issues.

During TALIS Starting Strong 2018, the level of comparability was deemed sufficient if all parameters of a model (except the residuals of the items[16]) were fixed to be the same across groups (e.g. participating countries). This approach is called the scalar level of measurement invariance. The *scalar level of measurement invariance* applies when (1) the structure of the construct is the same across groups, (2) the strength of the associations between the construct and the items (factor loadings) are equivalent, and (3) the intercepts/thresholds for all items across groups are equivalent. If the intercepts of the items for all groups are equivalent, then the expected value of the items becomes the same across groups when the value of the construct is zero, which means that the value/degree of the construct for a certain value of the observed item can be claimed to be equivalent across different groups. In this case, cross-group comparisons of scale means are justified, and the results can be assumed to be free of cross-group bias (e.g. cross-cultural bias).

TALIS Starting Strong sought to use up-to-date and valid techniques with solid analytical backgrounds for the scaling procedure to ensure the resulting model of each scale was an accurate representation of the characteristics of ECEC centre staff and leaders in the participating countries. This aim resulted in an examination of the measurement invariance across both participating countries and populations.

The measurement invariance testing referred to as "cross-country" examined invariance within a single population (either ISCED level 02 or U3). In other words, for the participating countries that participated with the ISCED level 02 population, measurement invariance testing was conducted in a way that considered each participating country to be a separate group during the analysis. This same procedure was followed separately for the U3 population. The measurement invariance testing referred to as "cross-population" examined invariance within a single country across both populations. Therefore, measurement invariance testing was conducted for each participating country that participated in the study at both ISCED level 02 and U3, which meant that the invariance testing treated ISCED level 02 and U3 as separate groups during the analysis. The measurement invariance testing for both the cross-country and the cross-population was then used to develop each scale's final model containing the parameter constraints suggested by the results of the invariance testing for each of the two populations conducted for that particular scale.

The modelling method used to conduct the measurement invariance during TALIS Starting Strong 2018 was MGCFA (the same method used for the three TALIS cycles). During the cross-country measurement invariance testing, the analysis evaluated whether the model was invariant/equal across participating countries within a certain population.[17] These models were compared at the configural, metric, and scalar levels, and the purpose of the analyses was to investigate if the statistical analysis of the scale scores could be compared across participating countries within each population.

During the cross-population measurement invariance testing, the analysis evaluated whether the model was invariant/equal across populations within a country. To be specific, for each country that participated with both populations, two CFA models[18] were created and compared at configural, metric, and scalar levels. The purpose of this analysis was to investigate whether statistics obtained from the analysis of the scale scores could be

compared across populations within a single country. The scale score estimation was based on the evaluation of each scale's results as described in the following sections.

The changes to the model fit indices CFI, TLI, RMSEA and SRMR/WRMR[19] were used to evaluate the measurement invariance level of each scale. The criteria used to conduct the evaluation were as follows:

- The *configural* level of invariance (no restriction on factor loadings or intercepts) was established if the model passed the following criteria: CFI≥0.90 or TLI≥0.90 and RMSEA≤0.08 or SRMR≤0.06/WRMR≤0.90.
- The *metric* level of invariance (with factor loadings set to be the same for different groups) was established if the difference in the fit indices between the metric and configural models passed the following criteria: ΔCFI<0.010 or ΔTLI<0.015 and ΔRMSEA>-0.015 or ΔSRMR>-0.03 (Chen, 2007[20]).
- The *scalar* level of invariance (factor loadings and intercepts set to be the same for different groups) was established, if the difference in the fit indices between the scalar and metric models passed the following criteria: ΔCFI<0.010 or ΔTLI<0.015 and ΔRMSEA>-0.015 or ΔSRMR>-0.01/ΔWRMR>-0.40 (Chen, 2007[20]).

In some instances, the configural model was "just identified", meaning that the model fit could not be evaluated due to the lack of degrees of freedom. The fit indices therefore indicated perfect model fit because the model was derived directly from the data (Brown, 2006, p. 66[12]). These cases occur in unidimensional scales with three items, resulting in a model that perfectly describes the empirical structure of the data. Comparison of the fit indices of a perfectly fitting configural invariant model with a metric invariant model can lead to detection of differences that are often greater than the recommended thresholds. Thus, in the TALIS Starting Strong 2018 scale evaluation, the models with three indicators were considered to be metrically invariant if the metric model fitted the data well (with the same criteria used for the configural model used for the evaluation). The reason for these special evaluation criteria is that the addition of constraints to a just identified model typically leads to worse fit indices beyond the difference criterion between the metric and the configural models outlined above, with this worsening having the potential to influence the resulting level of invariance.

Final scale modelling

The specification of parameters in the final model depended on the level of invariance established during the previous steps of the analysis both cross-country and cross-population. For example, if scalar invariance was established for one scale across countries participating at ISCED level 02, then the final model allowed equal factor loadings and intercepts across those countries. If, for the same scale, metric invariance had been established across the countries participating at the U3 level, then the final model also allowed equal factor loadings (while allowing the intercepts to vary) across those countries *in the same model*. Finally, the invariance testing results of the same scale's cross-population testing within each participating country were also allowed the necessary constraints within the same model.

In summary, the measurement invariance testing results for both populations across all participating countries and all participating countries' across populations for a single scale were modelled together in the scale's final model. Therefore, the final scale models accounted for all invariance results both cross-country and cross-population. The final scale

models underwent MGCFA, during which each individual country and population was modelled. The final scale models underwent the same evaluation with regard to their fit assessments, with final improvements being made to the models where necessary.

The final scale models for those participating countries that were excluded from the scale evaluation (see Table 11.6) used fixed parameters. The parameters were fixed according to the cross-country measurement invariance results for the ISCED level 02 population, TALIS Starting Strong 2018 study's target population. Therefore, if a scale reached scalar invariance cross-country at ISCED level 02, the factor loadings and intercepts were constrained to be equal to the unstandardised parameters in the final scale model. If metric invariance was reached, then only the factor loadings were constrained, and if configural invariance was reached, then no constraints were imposed on the factor loadings and intercepts.[15]

Once the final scale models had been specified for all participating countries and populations, factor score determinacies from the model were used to calculate the omega reliability coefficient as part of the evaluation of the scale. All parameters estimated in these models are reported below in the section titled *Results from scale evaluation*. Unstandardised factor loadings and intercepts are presented for the scales that reached metric or scalar levels of invariance for the designated population so as to show the model's equality restraints. In addition, standardised factor loadings are shown to aid interpretation of the model's level of reliability.[20]

11.3.3. Scale scores computation

Scale score estimation

Completion of the scale evaluation was followed by the creation of a scale score for each of the constructs. Using scale scores instead of analysing single items or sum scores offers many methodological and practical advantages. Although scale scores, when compared to individual variables, are not completely free of measurement error, this error is often minimised, thus increasing the reliability of these scores (Brown, 2006[7]; Hansen et al., 2006[16]). In practice, using a single score enhances the readability, interpretation and implementation of analysis as compared to analysis based on a set of variables. Another advantage of scale scores compared to simple sum scores is that the former accounts for differences in the relative strength of the relationships between the latent construct and the items (Cheung and Rensvold, 1998[28]). In addition, scale score computations account for missing data while still producing a score for each observation, an outcome that is more difficult to obtain through use of a simple sum score.

The computation of the scale scores in the TALIS Starting Strong 2018 was based on the CFA models previously established within the model evaluation. Thus, specification of the CFA models included the model modifications and considered the results of the measurement invariance testing. This approach meant that the model parameters in each scale were kept constant or allowed to vary according to established levels of measurement invariance (i.e. the approach considered the measurement invariance testing results for both cross-country and cross-population).

CFA models also allow development of scores known as *factor scores*. Within the MGCFA framework, the parameters can be estimated separately, in the multiple group models, for each single-country, single-population level. This approach makes it possible to constrain or freely estimate the factor loadings and item intercepts/thresholds, depending on the measurement invariance results. The factor scores are specified as continuous normally

distributed and the software program Mplus was used for computation of the scale scores representing the latent constructs.[4]

According to the SEM framework, an item y is predicted from the latent factor η, which is multiplied with the matrix of factor loadings $\boldsymbol{\Lambda}$. The vector of item intercepts $\boldsymbol{\tau}$ and the vector of residuals ε are both added to the product. This is written as:

$$y = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \eta + \varepsilon$$

To estimate factor models from ordinal items, the MLR estimation procedure for continuous latent constructs was used because it is robust to non-normality. Mplus uses the maximum of the posterior distribution of the factor, which is known as the maximum a posteriori method (Muthén, L. K. and B. O. Muthén, 1998–2017[21]). This method is similar to the latent regression approach (Skrondal, A. and P. Laake, 2001[22]). If all y items are continuous, the factor score estimate η for individual i is based on a regression method with correlated factors (Muthén, 1977[23]), where the factor score is computed from the mean vector of y items, denoted as $\boldsymbol{\mu}$, the factor score coefficient matrix $\boldsymbol{C}$, the vector of observations $\boldsymbol{v}_i$, the vector of intercepts $\boldsymbol{\tau}$, and the matrix of factor loadings $\boldsymbol{\Lambda}$ multiplied by the mean vector $\boldsymbol{\mu}$:

$$\hat{\eta}_i = \boldsymbol{\mu}_y + \boldsymbol{C}(\boldsymbol{v}_i - \boldsymbol{\tau}_y - \boldsymbol{\Lambda}_y \boldsymbol{\mu}_y)$$

The factor score coefficient matrix, in turn, is based on the item covariance matrix $\boldsymbol{\Sigma}$, the matrix of factor loadings $\boldsymbol{\Lambda}$, and the matrix of residual variances and covariances $\boldsymbol{\Theta}$:

$$C = \boldsymbol{\Sigma}_y \boldsymbol{\Lambda}_y^T (\boldsymbol{\Lambda}_y \boldsymbol{\Sigma}_y \boldsymbol{\Lambda}_y^T + \boldsymbol{\Theta}_y)^{-1}$$

These formulas imply that higher factor loadings on an item are associated with a stronger influence of this item on the factor score estimate. Likewise, the larger the residual variance of an item, the smaller its influence on the factor score estimate. The factor loadings, item intercepts, the mean vector and the variance of the latent variable affect the estimated scores.

The WLSMV estimation procedure was used to estimate factor models with scaled binary items. This method produces weighted least square parameter estimates by using a diagonal weight matrix, robust standard errors, and a mean- and variance-adjusted χ^2 test statistic (Brown, 2006, p. 388[12]). The method also takes a slightly different approach to estimating factor scores. First, the probability of observed binary response "1" is defined as

$$f_j(y_{ij}|\boldsymbol{\eta}_i) = \Phi\left[(\zeta - \boldsymbol{\lambda}_j' \boldsymbol{\eta}_i)\theta_{jj}^{-\frac{1}{2}}\right]$$

and the probability of observed categorical response "0" is therefore $1 - f_j(y_{ij}|\eta_i)$, where ζ is the item threshold based of an item j, $\boldsymbol{\lambda}_j'$ is the jth row of the matrix of factor loadings $\boldsymbol{\Lambda}$, and θ_j is the jth diagonal of the matrix of residual variances and covariances $\boldsymbol{\Theta}$, while $\boldsymbol{\eta}_i$ is a vector of true factor scores.

The factor score estimates $\hat{\eta}_i$ are then found as the mode of the posterior distribution of $\boldsymbol{\eta}_i$ by minimising, through use of quasi-Newton techniques, the following function F with respect to $\boldsymbol{\eta}_i$:

$$F = \frac{1}{2}(\boldsymbol{\eta}_i - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\mu}_i) - \sum_{j=1}^{p} \ln f_j(y_{ij}|\boldsymbol{\eta}_i)$$

where $\boldsymbol{\mu}$ is the mean vector of y items. Contrary to the factor score estimation for models with categorical items, this approach assumes uncorrelated residual variances even if residual covariances are allowed.

For the scales in TALIS Starting Strong 2018 based on both continuous and categorical data (mixed data), WLSMV estimation was used. The procedures for computing factor scores follow those used for scales with strictly binary items described above. However, an additional function was added for computing probabilities for the ordinal variables as follows:

$$f_j(y_{ij}|\eta_i) = e^{-\frac{(y_{ij}-v_j-\boldsymbol{\lambda}_j'\boldsymbol{\eta}_i)^2}{\theta_{jj}}}$$

where v_j is a vector of item intercepts.

For both continuous and categorical data, Mplus provides a model-based approach to estimating parameters in a model with missing data. Model-based approaches account for the missing data and estimate the missing parameters in one step (Lüdtke, O. et al, 2007[24]), and to do this Mplus uses the expectation maximisation algorithm (for a detailed description see Dempster, A. P., N. M. Laird and D. B. Rubin, 1977[18]. This procedure assumes that the data are missing-at-random, which means that the probability of a missing observation depends not on the true score of a person regarding the latent construct but may depend on non-missing observed data and be correlated with other covariates of the scale (Schafer, J. L. and J. W. Graham, 2002, p. 147[25]).

For each TALIS Starting Strong scale, the expectation maximisation algorithm was used to compute a scale score for respondents who responded to at least one of the items belonging to the respective scale (this algorithm made it possible to deal with missing data and provide the appropriate estimator for the continuous or categorical nature of the respective scales). The residual variances of the items were allowed to be freely estimated in all models.

Scale score standardisation

To enhance interpretation of the scale scores, the scores were standardised so that the value 10 corresponds to the mid-point of the scale. This approach differs from standardisations in which a specific value is set to be equal to the mean of the scale. The formula is presented below, but preparatory to it, Figure 11.1 presents an explanation of how to interpret the scale score values. The figure displays the questionnaire items that were used to create the scale *Staff need for professional development in pedagogical content knowledge* (S1PDPE). The metric of scale scores was transformed to indicate the relative midpoint of the original scale items' categories. More specifically, S1PDPE was measured by the responses to items H through L (with variable names SS1G16H through SS1G16L) from question 16 in the staff questionnaire (question 12 in the combined questionnaire). These items contained the responses "no need at present", "low level of need", "moderate level of need" and "high level of need", coded "1" to "4", respectively.

Numerically, the midpoint for each item is 2.5, as shown in the figure. To calculate the *item midpoint value* (IMV) for each individual, a simple average is calculated for all item responses. Conceptually, if an individual's IMV is less than 2.5, this suggests that, on average, items can be considered as *no or lesser need*. An IMV greater than 2.5 suggests the items are, on average, of *greater need*. A value of 2.5 suggests indifference.

Data from the TALIS Starting Strong ISCED level 02 populations (with the exception of those in the participating countries not meeting the technical standards; see Table 11.6) were used during standardisation of the estimated scale scores. The metric used to standardise the scale scores had a standard deviation of 2.0 and a mean of 10. The mathematical transformation ensured that all or almost all values were positive, thus allowing for a convenient interpretation.

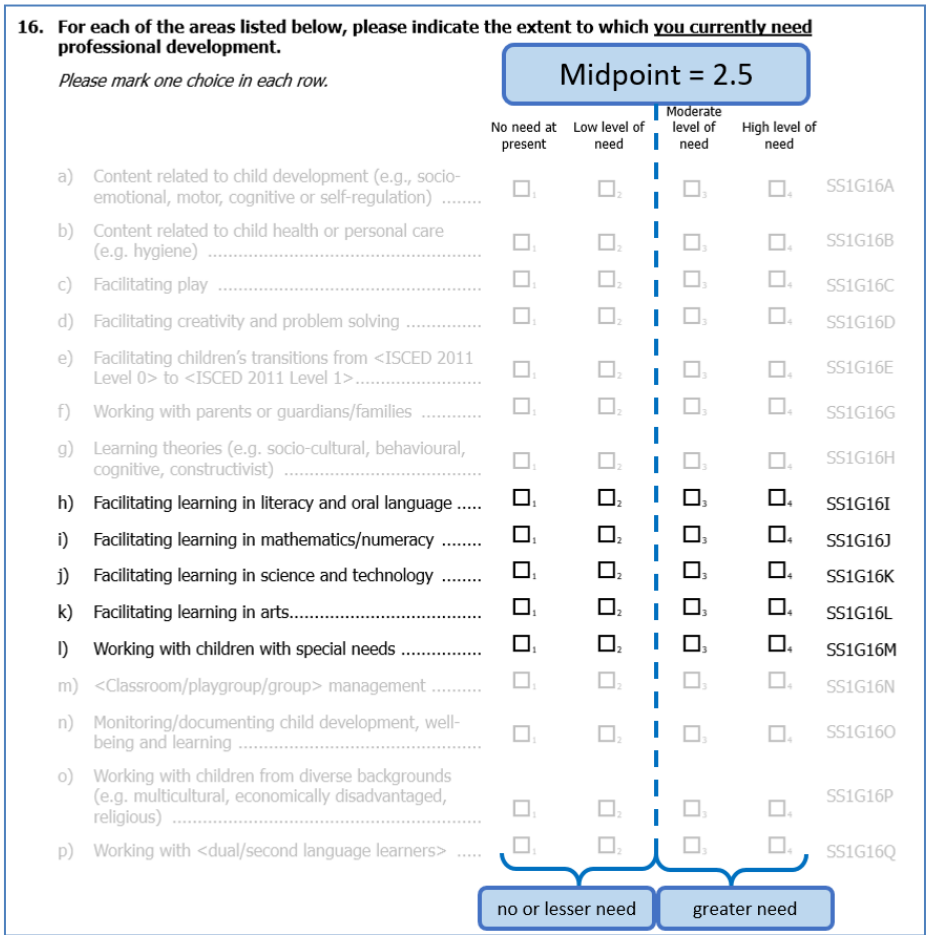
Once the scores were adjusted with a standard deviation of 2.0 and a mean of 10, a second adjustment was made as follows. The average scale score for the set of those individuals from ISCED level 02 whose IMV was equal to the midpoint of the scale items (in this example, 2.5) was computed and then subtracted from 10. This difference ($\bar{F}_M^*$) was then added to the scale score for each individual, resulting in the final standardised scores.

Mathematically, the standardisation is represented as follows:

$$X_i = 10 + 2\frac{F_i - \bar{F}^*}{\sigma_{F^*}} + \bar{F}_M^*$$

where X_i is the standardised scale score of individual i, F_i is the raw estimated scale score of that individual, $\bar{F}^*$ is the mean scale score of the ISCED level 02 population, σ_{F^*} is the standard deviation of the scale score of the ISCED level 02 population, and $\bar{F}_M^*$ is as described above.

Figure 11.1. Illustration of the midpoint of a scale



The addition of $\bar{F}_M^*$ shifts the standardised scale scores so that the scale score average becomes 10 for those individuals from the target population with an IMV equal to the midpoint of the response scale. This shift in the scale score mean to 10 plus the constant $\bar{F}_M^*$ allows for easy interpretation of scale scores: scores above 10 suggest positive associations with the scale (e.g. *agreement*, *no need or lesser need*), scores below 10 suggest negative associations with the scale (e.g. *disagreement*, *greater need*), while scores of exactly 10 suggest indifference.

Table 11.3 provides an example of this procedure for the scale S1PDPE. The column on the left shows all the possible IMVs for individuals obtained from aggregating the responses to the four items of the scale. The column on the right shows the mean factor scores of individuals from the target population with the corresponding IMV. As shown, individuals with an IMV of 2.5 have an average scale score of 10. In general, these means rise as the IMV rises, and fall as the IMV falls.[21]

Table 11.3. Items' mean-scaled score equivalent table for the scale S1PDPE

Average item score on the set of items	Mean factor score
1.00	6.70
1.20	7.05
1.25	6.79
1.33	-
1.40	7.24
1.50	7.60
1.60	7.78
1.67	-
1.75	7.72
1.80	8.23
2.00	8.59
2.20	8.94
2.25	9.11
2.33	9.93
2.40	9.45
2.50	**10.00**
2.60	9.98
2.67	9.33
2.75	10.72
2.80	10.41
3.00	10.79
3.20	11.28
3.25	11.70
3.33	11.33
3.40	11.99
3.50	12.34
3.60	12.67
3.67	12.85
3.75	12.26
3.80	13.09
4.00	13.50

Note: Some means are blank because no cases in the target population had an item average of that value.

Composite scale scores

Among the scales created for the TALIS Starting Strong 2018 dataset are two multidimensional scales, the scores of which are based on three subscale components.[22] The subscales were constructed within a multidimensional model to account for the multidimensional latent construct's structure, and the factor scores were then computed by taking the whole structure into account. To calculate the factor scores for the overall multidimensional latent constructs, composite scores were computed by taking a simple average of the corresponding standardised scores of the subscales.[23] The computation of the composite scores for individual can be summarised as follows:

$$Y_i = \frac{\sum_{j=1}^{N} X_{ij}}{N}$$

where Y_i are the composite scores for a certain multidimensional scale for individual i, X_{ij} are the raw factor scores of the subscale j for individual i, and N is the number of subscales

of this multidimensional scale. These scores were then standardised in the same manner as described in the subsection of this chapter titled *Scale score standardisation*.

Interpretations of results from any analysis between the subscale scores and the composite scores need to consider the following limitations: (1) the subscale scores should not be used simultaneously in a correlation or regression analysis due to collinearity of these subscales; and (2) the weight of the subscale on the general latent construct was not taken into account (i.e. assumed to be equal for all subscales).

11.3.4. Recommendation for analysis and interpretation

Cross-country comparability

Results from the scale reliability analyses and the cross-country or cross-population measurement invariance testing described above have major implications for (1) the construction of scale scores, and (2) the use of the scale scores in further analysis.

As a guide for the comparability of the scale scores, the labels for the scale scores in the international datasets give the cross-country levels of invariance. For example, the scale with the variable name S1PDPE has the label, "Prof. development – needs pedagogy / Scalar (U3) – Metric (0.2)", which indicates that the scale reached scalar invariance within the U3 population and metric invariance within the ISCED level 02 population.

Table 11.4 shows how many scales reached a particular level of invariance. For a more detailed look, Table 11.5 presents the specific invariance levels for each scale, listed by its variable name and the label name found in the international database.

Table 11.4. Scale counts of the invariance levels for both TALIS Starting Strong 2018 populations

	Invariance levels		
Population	Configural	Metric	Scalar
Staff			
ISCED level 02	5	13	0
U3	7	9	2
Leaders			
ISCED level 02	4	3	0
U3	2	3	2

Table 11.5. Invariance level reached for each scale by population

	Scale label	Variable name	ISCED level 02	U3
Leader scales	Professional development: needs leadership	S1LDEV	Metric	Configural
	Diversity beliefs	S1LDIVB	Configural	Scalar
	Distributed leadership	S1LEADP	Configural	Metric
	Leader support for pedagogical learning	S1LEADS	Metric	Metric
	Neighbourhood	S1LNIGHB	Configural	Metric
	Networking	S1LPANE	Metric	Scalar
	Opportunities for parental participation	S1LPART	Configural	Configural
Staff scales	Activities to enhance child development	S1ACTIV	Metric	Configural
	Engagement in collaborative practices	S1COLL	Metric	Metric
	Satisfaction with working conditions	S1JOBSA	Metric	Configural
	Facilitating parent/guardian engagement	S1PAREN	Metric	Metric
	Professional development: needs diversity	S1PDIV	Configural	Scalar
	Staff participation in collaborative prof. development	S1PDNET	Configural	Configural
	Professional development: needs pedagogy	S1PDPE	Metric	Scalar
	Adaptive pedagogical practices	S1PRACT	Metric	Metric
	Facilitating socio-emotional development, overall	S1PRDEV	Metric	Metric
	Facilitating emotional development	S1PREM	Metric	Configural
	Facilitating child initiative	S1PRINT	Configural	Metric
	Facilitating language	S1PRLAN	Configural	Metric
	Facilitating learning and development, overall	S1PROLD	Configural	Configural
	Facilitating literacy development	S1PROLI	Metric	Configural
	Facilitating numeracy development	S1PRONU	Metric	Metric
	Facilitating prosocial behaviour	S1PRSOC	Metric	Metric
	Behavioural support	S1SOCIAL	Metric	Configural
	Sources of work stress	S1STRES	Metric	Metric

Source: OECD, TALIS Starting Strong 2018 database.

In addition, recommendations for analysis based on the different levels of invariance are provided below. The recommendations are specific to cross-country invariance (i.e. within a single population) but may also be applied to cross-population invariance within a single participating country. Different levels of measurement invariance provide different potentials for the analysis of data and reporting. The proposed analyses for each level of measurement invariance are:

- Cross-country analysis of scales with only a configural level of invariance:
  - *Recommendation for analysis:* At the cross-national level, the only comparisons that are statistically justified are qualitative (descriptive) comparisons (e.g. "associations between staff-beliefs and staff education is positive in countries A and B, whereas there is no significant association in country C"). Presentation of these comparisons should include the limitations influencing interpretation of the results, in particular the differences between participating countries with respect to the meaning of the construct.
  - *Limitation:* If a scale reaches only configural invariance, it means that the scale score has been constructed in a way that allows the factor loadings and intercepts to vary across participating countries. Further analysis aimed at cross-country comparisons can only be conducted at the conceptual level, meaning that *no* statistical methods of comparison (such as *t*-tests) are applicable in these cases.

- Cross-country analysis of scales with a metric level of invariance:
  - *Recommendation for analysis:* The strength of the association between the scale and items in scales with a metric level of invariance is comparable across participating countries, and statistical methods such as correlation and regression are applicable. Comparisons of associations between participating countries are justified (e.g. "The association between staff-beliefs and level of education is significantly stronger in countries A and B than in country C"). The metric level of measurement invariance means scales can be used for analysis based on correlation and linear regression.
  - *Limitation:* If a scale reaches only the metric level of invariance, it means the score of the scale was created respectively with equal factor loadings but with intercepts allowed to vary across the participating countries. As such, a mean score between participating countries is not possible.
- Cross-country analysis of scales with a scalar level of invariance:
  - *Recommendation for analysis:* With these scales, the comparison of the mean score of the scale is meaningful across groups. The scale's mean can be compared across participating countries (e.g. "Staff-beliefs are *significantly higher* in country A compared to countries B and C").

Violating these assumptions or ignoring these recommendations may or will lead to biased results and/or interpretations. For instance, if scale scores based on the assumption of metric invariance are created and used to compare the country means of that scale, the differences between participating countries will most likely be biased. Thus, it is recommended that data users apply statistical procedures to analyse scale scores according to the scale's reached level of invariance.

11.4. Results from scale evaluation

The remainder of this chapter presents the evaluation results of the scales. While the results presented in tables for the first scale on *Staff need for professional development in pedagogical content knowledge* are commented on in the text in great detail, the following scales' results contain short remarks on the most important results, including complementary information. The scales are grouped by questionnaire (staff and leader).

These scale-specific sections are further divided into several parts that present detailed results for each scale. Each scale description begins with a presentation of the items used as indicators to measure the latent constructs. These include item names, item wording, and response categories (the presented items are copied from the ISCED level 02 versions of the questionnaires). This part of the chapter also includes special item notes that state whether certain items were reverse coded for the purpose of the scale construction, were removed from the scale, or were not administered by design to the U3 population. The reason for reverse coding items was to align the responses to the continuum of the scale; in other words, if the response to the item was positive (coded as a higher integer numerically), then it corresponded to a positive association with the latent construct.[24] Removing an item from a scale construct was applied in specific cases. Initially, the scale items were chosen on the basis of theoretical considerations (see the section *Scale development and methods* in this chapter and also Chapter 3 of this report) and with the expectation that a latent construct measured by selected items would have statistical properties that supported the construct (e.g. sufficiently high factor loadings). However, empirical examination of the

scales showed that, occasionally, certain items did not exhibit the expected properties, thereby indicating that the empirical data did not support the conceptual development of the construct. These items were removed, and the model was evaluated without them.

Each scale description also contains information on improvements (where applicable) to the implemented model. Improvements either encompassed all participating countries or were country specific. They typically occurred in those instances where the initial model did not fit the data well, but where a minor improvement (such as a correlation between two items) resulted in an acceptable model fit. These results corresponded to results obtained during the reliability analyses of scale evaluation (see the section *Scale development and methods* in this chapter for details on scale evaluation).

While all model fit calculations and tests conducted during all further steps included these improvements, there were several instances where additional improvements were made to the models during the testing for measurement invariance. Changes were made programmatically so that the models could be estimated in a way that neither biased the results nor changed the content of the scale construct. These changes included, for example, fixing the negative residual variance for items in some scales so that it was positive and close to zero.

The reliabilities for each scale follow the scale's description. The omega statistic was used for scales based on continuous response options, the Cronbach's alpha coefficient for scales based on categorical response options, and the stratified Cronbach's alpha for the multidimensional scales. Table 11.1 presents the criteria guiding interpretation of each scale's reliability.

The information on reliability is followed with a description of the scale-specific results from the model estimation and comparisons. Confirmatory factor analysis (CFA) was used to evaluate the theoretically developed scale models reflecting the study's latent constructs, while the model fit indices presented at this point were those obtained from the model analysis conducted at the country level for each scale (i.e. the CFA conducted separately for each of the analysed populations).[25] The criteria used to evaluate and select the models are described in the first part of this chapter (see Table 11.1). The model fit indices for scales based on three items are not reported.[26] This information is followed by the results of the measurement invariance testing across countries for each population from the multigroup CFA (MGCFA). A summary of the results from the measurement invariance testing across populations within each country can be found in Annex I. (Tables I.2 and I.3). The results of the model comparisons during the measurement invariance testing were used to establish the invariance levels of each scale.

The final section in the scale description contains the international unstandardised model parameters that were estimated in the scales' final models. Of note, no scale reached scalar invariance. Only the factor loadings are reported for the metric invariant scales. The unstandardised item parameters are not reported for scales that reached configural invariance because the parameters are country specific. Furthermore, the standardised factor loadings and unstandardised intercepts are reported for each country per population in a separate table. This is due to the fact that in the MGCFA final models the factor loadings were standardised at the country level, meaning that these standardised parameters would differ at the country level, even for the scalar invariant scales. The standardised factor loadings indicate the strength of the relationship between each item and the latent scale. The values were interpreted as suggested by the cut-off criteria provided in Table 11.1. The unstandardised intercepts indicate the predicted values for an item when the value of the latent trait is zero.

Three multidimensional scales were constructed from the staff questionnaire, while none were constructed from the leader questionnaire. The multidimensional scales consist of several subscales. Because the results of the subscales depended on the multidimensional or hierarchical structure of the overall scale,[10] fit indices and measurement invariance testing results do not appear for these scales individually. However, to ensure the quality of each subscale, unidimensional models were tested for model fit. The tests found that all subscales were sufficient for inclusion in the multidimensional model. The standardised latent covariances (S1PROLD and S1PRDEV) and first order factor loadings on the general factor (S1STRES) are also presented for these scales.

Denmark did not meet the requirements determined in the TALIS Starting Strong 2018 technical standards (the adjudication process rated their data as "insufficient"; for details, see Chapter 10). The international parameters from the final scale models (fixed factor loadings in the ISCED level 02 metric models; fixed factor loadings and intercepts in the ISCED level 02 scalar models) were consequently used to estimate their scale scores. Countries were sometimes excluded from the scale evaluation because of model non-convergence, instability or some other issue. They were excluded during the CFA analysis at the country level from further evaluation and did not contribute to the final scale model. The scale scores for all excluded countries were derived using fixed parameters (for details, see the section of this chapter titled *Final scale modelling*). Table 11.6 presents the populations that were excluded from the scales and the reasons for their exclusion. In some cases the fixed parameter models failed and for this reason no scales scores could be produced for a country. These scales and countries are noted in Table 11.6.

Table 11.6. Excluded populations from the estimation of scales' international parameters

Scale	Excluded country	Reason for exclusion	Received a scale score
Staff			
S1PDIV	Turkey (ISCED level 02)	Administered two items, and the use of a phantom indicator resulted in complications with model estimation during measurement invariance testing	No
S1PROLD	Germany (U3)	Poor model fit	Yes
S1SOCIAL	Korea (ISCED level 02)	Poor model fit	Yes
S1STRES	Israel (U3)	Complications with model estimation	Yes
Leader			
S1LDIVB	Israel (ISCED level 02)	Complications with model estimation	No
S1LNIGHB	Korea (ISCED level 02)	Poor model fit	Yes
S1LPANE	Israel (U3)	Poor model fit	Yes
	Norway (U3)	Complications with model estimation	Yes
S1LPART	Denmark (U3)	Model complications[1]	No
	Denmark (ISCED level 02)	Model complications[1]	No
	Germany (ISCED level 02)	Complications with model estimation	No
	Korea (ISCED level 02)	Complications with model estimation	No
	Norway (U3)	Complications with model estimation	No

Note: Denmark was excluded from all the scales because it did not reach the determined TALIS Starting Strong 2018 technical standards and because the data were rated as insufficient during the adjudication process.

[1]Complications with model estimation occurred due to a variety of factors such as non-convergence regardless of the maximum number of iterations allowed, or when continually occurring negative residual variances in one item after a previous negative residual variance was constrained to a positive value.

11.4.1. Scales derived from the staff questionnaire

Staff need for professional development in pedagogical content knowledge (S1PDPE)

Items

The items included in the scale *Staff need for professional development in pedagogical content knowledge* (S1PDPE) represented different areas of professional development and required respondents to assess the level of need for each. The items are presented in Table 11.7.

Table 11.7. Item wording for the scale S1PDPE

S1PDPE: Staff need for professional development in pedagogical content knowledge	
SS1G16: For each of the areas listed below, please indicate the extent to which you currently need professional development.	
Response options: no need at present (1), low level of need (2), moderate level of need (3), high level of need (4)	
SS1G16H	Learning theories (e.g. socio-cultural, behavioural, cognitive, constructivist)
SS1G16I	Facilitating learning in literacy and oral language
SS1G16J	Facilitating learning in mathematics/numeracy
SS1G16K	Facilitating learning in science and technology
SS1G16L	Facilitating learning in arts

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

Improvements were added to the model for S1PDPE for Israel. Correlations between items SS1G16H and SS1G16I and between SS1G16J and SS1G16K were first included in the reliability analyses of the country level models and persisted through the MGCFA and final models where applicable.

Scale reliabilities and model fits

Table 11.8 presents the reliabilities and model fits for all populations for the scale S1PDPE. The omega coefficients for all participating countries are above 0.81. Overall, the results indicate that the scale has very high reliability in all participating countries for both populations, ISCED level 02 and U3. Notably, the reliabilities in Denmark (which did not contribute to the model parameters) are sufficiently high in both populations.

The CFI for the scale S1PDPE is above the cut-off criterion (acceptable) for all countries in both populations, while the TLI is below the cut-off criterion in Chile, Germany, Japan and Korea for ISCED level 02 and in Germany for U3. The RMSEA is acceptable only in Korea, Norway and Turkey for ISCED level 02. Finally, except for Denmark, the SRMR is acceptable in all participating countries for both populations. Overall, when analysed together, the fit indices indicate an acceptable to good model fit in most of the countries.

Table 11.8. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1PDPE

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.904	0.944	0.888	0.119	0.036
Germany	0.859	0.923	0.846	0.131	0.047
Iceland	0.867	0.954	0.909	0.090	0.032
Israel	0.899	0.970	0.901	0.105	0.022
Japan	0.878	0.944	0.888	0.097	0.032
Korea	0.966	0.942	0.885	0.077	0.026
Norway	0.882	0.982	0.963	0.064	0.021
Turkey	0.953	0.980	0.960	0.057	0.015
Denmark	0.895	0.906	0.906	0.102	0.148
U3					
Germany	0.884	0.918	0.835	0.132	0.053
Israel	0.812	0.972	0.906	0.099	0.017
Norway	0.876	0.964	0.928	0.093	0.028
Denmark	0.897	0.906	0.906	0.101	0.113

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.9 provides the results from the invariance testing for S1PDPE for each population (see the section of this chapter titled *Measurement invariance testing: across countries and across populations* for a refresher on invariance testing procedures).

The configural model shows a good fit, as it meets the criteria for CFI, TLI and SRMR in the ISCED level 02 population and CFI and SRMR in the U3 population. Because the degree of change in the TLI and RMSEA between the configural and metric models for ISCED level 02 is within the cut-off criteria, S1PDPE is considered metric invariant for ISCED level 02. The degree of change for the U3 population in the TLI and RMSEA between the configural and metric models is within the cut-off criteria, as is the degree of change for U3 in the TLI, RMSEA and SRMR between the metric and scalar models. Thus, S1PDPE reached the scalar invariance level for the U3 population.

Table 11.9. Invariance test results for the scale S1PDPE

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.955	0.905	0.089	0.030				
Metric	0.942	0.930	0.076	0.052	0.013	-0.025	0.013	-0.022
Scalar	0.864	0.884	0.098	0.101	0.078	0.046	-0.022	-0.049
Invariance level for U3								
Configural	0.950	0.884	0.112	0.036				
Metric	0.936	0.908	0.100	0.055	0.014	-0.024	0.012	-0.019
Scalar	0.911	0.908	0.100	0.058	0.025	0.000	0.000	-0.003

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

As mentioned above, as S1PDPE is metric invariant for ISCED level 02, the unstandardised factor loadings are the same across populations. However, the intercepts need to be estimated separately for each population. Table 11.10 presents the unstandardised factor loadings of the scale for both populations.

Table 11.10. Unstandardised factor loadings for the scale S1PDPE

Items	Factor loadings
SS1G16H	0.610
SS1G16I	0.761
SS1G16J	0.864
SS1G16K	0.812
SS1G16L	0.724

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.11 reports the standardised factor loadings and unstandardised intercepts parameters for the scale S1PDPE. The standardised factor loadings indicate the strength of the relationship between each item and the scale construct (see the section *Results from scale evaluation* in this chapter).

The standardised factor loadings for the scale S1PDPE in the table are all above 0.45, which indicates at least a moderate relationship between the items and the latent construct. More specifically, all values for items SS1G16I, SS1G16J, SS1G16K and SS1G16L are above 0.60, while the lowest item loading found for item SS1G16H is still above 0.45.

Table 11.11. Unstandardised intercepts and standardised factor loadings for the scale S1PDPE

	Standardised factor loadings					Unstandardised intercepts				
Participating country	SS1G16H	SS1G16I	SS1G16J	SS1G16K	SS1G16L	SS1G16H	SS1G16I	SS1G16J	SS1G16K	SS1G16L
ISCED level 02										
Chile	0.634	0.789	0.883	0.832	0.746	2.455	2.522	2.527	2.604	2.444
Germany	0.564	0.661	0.837	0.787	0.681	2.342	2.475	2.199	2.279	2.254
Iceland	0.563	0.740	0.849	0.759	0.682	2.301	2.672	2.614	2.649	2.502
Israel	0.640	0.805	0.900	0.806	0.741	2.342	2.445	2.351	2.476	2.494
Japan	0.556	0.767	0.863	0.774	0.698	2.499	3.056	2.875	2.777	3.052
Korea	0.636	0.873	0.970	0.928	0.822	2.601	2.707	2.666	2.681	2.765
Norway	0.626	0.724	0.846	0.808	0.748	2.327	2.649	2.349	2.393	2.376
Turkey	0.733	0.882	0.954	0.873	0.809	1.899	1.759	1.810	1.941	1.710
Denmark	0.678	0.704	0.829	0.848	0.815	2.124	2.350	1.964	2.063	2.060
U3										
Germany	0.580	0.664	0.876	0.825	0.685	2.342	2.475	2.199	2.279	2.254
Israel	0.581	0.709	0.814	0.758	0.719	2.342	2.475	2.199	2.279	2.254
Norway	0.610	0.713	0.841	0.824	0.699	2.342	2.475	2.199	2.279	2.254
Denmark	0.643	0.734	0.833	0.860	0.798	2.214	2.277	1.953	1.958	2.119

Source: OECD, TALIS Starting Strong 2018 database.

Staff need for professional development for dealing with diversity (S1PDIV)

Items

The items included in the scale *Staff need for professional development for dealing with diversity* (S1PDIV) (see Table 11.12) represent different areas of professional development related to dealing with diversity and required respondents to assess the level of need for each of the three areas.

Table 11.12. Item wording for the scale S1PDIV

S1PDIV: Staff need for professional development for dealing with diversity	
SS1G16: For each of the areas listed below, please indicate the extent to which you currently need professional development.	
Response options: no need at present (1), low level of need (2), moderate level of need (3), high level of need (4)	
SS1G16M	Working with children with special needs
SS1G16P	Working with children from diverse backgrounds (e.g. multicultural, economically disadvantaged, religious)
SS1G16Q	Working with <dual/second language learners>

Source: OECD, TALIS Starting Strong 2018 database.

Scale reliabilities and model fits

Table 11.13, which presents the reliability coefficients for both populations for the scale S1PDIV, shows a high omega coefficient (above 0.70) for both populations in all participating countries.

The model fit indices for the scale S1PDIV are not reported, as the scale was based on only three items.[26]

Table 11.13. Reliability coefficients for the scale S1PDIV

Participating country	Omega
ISCED level 02	
Chile	0.716
Germany	0.850
Iceland	0.885
Israel	0.835
Japan	0.776
Korea	0.891
Norway	0.899
Turkey*	-
Denmark	0.897
U3	
Germany	0.834
Israel	0.828
Norway	0.895
Denmark	0.916

Note: *Turkey administered only two items, and the use of a phantom indicator led to errors in model estimation. Therefore, the results are not available.
Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

The invariance testing results for the scale S1PDIV are presented in Table 11.14. The scale reached configural invariance for the ISCED level 02 population and scalar invariance for the U3 population.

Table 11.14. Invariance test results for the scale S1PDIV

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	1.000	1.000	0.000	0.000				
Metric	0.966	0.940	0.091	0.124	.034	.060	-.091	-.124
Scalar	0.468	0.535	0.255	0.451	.498	.405	-.164	-.327
Invariance level for U3								
Configural	1.000	1.000	0.000	0.000				
Metric	0.996	0.990	0.038	0.033	0.004	0.010	-0.038	-0.033
Scalar	0.988	0.986	0.046	0.042	0.008	0.004	-0.008	-0.009

Note: An adjusted number of degrees of freedom was used to compute the CFI, TLI and RMSEA for ISCED level 02 (see the subsection *Treating missing items in scales* in this chapter).
Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

There are no unstandardised parameters for the S1PDIV scale because it reached only a configural level of invariance for the ISCED level 02 population and the parameters were country specific. Table 11.15 presents the item parameters. The factor loadings for items SS1G16M, SS1G16P and SS1G16Q are generally high and mostly above 0.45. However, item SS1G16M shows lower factor loadings in many countries and a loading below 0.45 in Japan at the ISCED level 02, indicating a weak relationship between this item and the scale construct in most of the participating countries.

Table 11.15. Unstandardised intercepts and standardised factor loadings for the scale S1PDIV

	Standardised factor loadings			Unstandardised intercepts		
Participating country	SS1G16M	SS1G16P	SS1G16Q	SS1G16M	SS1G16P	SS1G16Q
ISCED level 02						
Chile	0.569	0.765	0.623	3.131	2.578	2.893
Germany	0.476	0.878	0.816	2.714	2.472	2.536
Iceland	0.605	0.868	0.897	2.926	2.702	2.823
Israel	0.544	0.873	0.767	2.714	2.472	2.536
Japan	0.380	0.841	0.683	3.778	3.050	2.674
Korea	0.718	0.909	0.839	2.907	2.816	2.750
Norway	0.549	0.914	0.878	2.714	2.472	2.536
Turkey*	-	-	-	-	-	-
Denmark	0.482	0.926	0.835	2.575	2.040	2.231
U3						
Germany	0.461	0.864	0.801	2.714	2.472	2.536
Israel	0.516	0.858	0.792	2.714	2.472	2.536
Norway	0.571	0.919	0.852	2.714	2.472	2.536
Denmark	0.542	0.937	0.876	2.454	2.098	2.224

Note: *Turkey administered only two items, and the use of a phantom indicator led to errors in model estimation. Therefore, the results are not available.
Source: OECD, TALIS Starting Strong 2018 database.

Multidimensional scale – Facilitating socio-emotional development (overall) (S1PRDEV), with the following subscales: Facilitating child initiative (S1PRINT); Facilitating prosocial behaviour (S1PRSOC); and Facilitating emotional development (S1PREM)

Items

The overall scale *Facilitating socio-emotional development* (S1PRDEV) was based on items from two questions. Respondents had to indicate the extent to which the statements applied to their ECEC centre. The first group of items was about facilitating child-based initiatives and together these items formed the subscale S1PRINT. The second group of items, which focused on facilitating prosocial behaviour, formed the subscale S1PRSOC. The last group of items was about facilitating emotional development, and it formed the subscale S1PREM. Table 11.16 presents the wording of these items and their response options.

Table 11.16. Item wording for the overall scale S1PRDEV and subscales S1PRINT, S1PRSOC and S1PREM

S1PRDEV: Facilitating socio-emotional development (overall)	
S1PRINT: Facilitating child initiative (subscale)	
SS1G29: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G29A	If invited, <ECEC staff> join in with the children's play
SS1G29B	When <ECEC staff> play with children, the children are allowed to take the lead
SS1G29C	<ECEC staff> show enjoyment when joining the children's play
SS1G29D*	<ECEC staff> arrange settings to allow extensive play amongst groups of children
S1PRSOC: Facilitating prosocial behaviour (subscale)	
SS1G29: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G29E	<ECEC staff> encourage sharing amongst children
SS1G29F	<ECEC staff> encourage children to help each other
SS1G29G*	<ECEC staff> encourage children playing in small groups to include other children
SS1G29H	<ECEC staff> encourage children if they comfort each other
S1PREM: Facilitating emotional development (subscale)	
SS1G30: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G30A	<ECEC staff> hug the children
SS1G30B	<ECEC staff> talk with children about feelings
SS1G30C	<ECEC staff> help children to talk about what makes them happy
SS1G30D	<ECEC staff> help children to talk about what makes them sad

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

For the multidimensional scale S1PRDEV, correlations between SS1G29A and SS1G29C and SS1G29E and SS1G29F were added for all countries.

Scale reliabilities and model fits

Table 11.17 presents the reliability coefficients (omega for the subscales, stratified Cronbach's alpha for the overall scale for S1PRDEV) for all populations for each (sub)scale. In general, the subscales S1PRINT, S1PRSOC and S1PREM have high reliabilities in most countries in both populations. Reliability coefficients are slightly lower, below 0.70 in Germany and Japan (ISCED level 02) and Denmark, Germany and Israel (U3) for the scale S1PRINT. The composite scale S1PRDEV has very high reliabilities in all countries in both populations.

Table 11.18 presents the model fits for S1PRDEV. The indices CFI, TLI, RMSEA and SRMR for this scale meet the cut-off criteria in all countries for both populations, while CFI, TLI and SRMR are below or above the cut-off value for both populations in Denmark.

Table 11.17. Reliability coefficients for the multidimensional scale S1PRDEV and subscales S1PRINT, S1PRSOC and S1PREM

	Stratified alpha	Omega		
Participating country	S1PRDEV	S1PRINT	S1PRSOC	S1PREM
ISCED level 02				
Chile	0.937	0.821	0.826	0.908
Germany	0.928	0.726	0.774	0.916
Iceland	0.915	0.716	0.780	0.867
Israel	0.911	0.753	0.773	0.828
Japan	0.918	0.697	0.769	0.927
Korea	0.945	0.832	0.854	0.916
Norway	0.932	0.814	0.837	0.897
Turkey	0.944	0.835	0.857	0.897
Denmark	0.946	0.812	0.876	0.945
U3				
Germany	0.925	0.686	0.780	0.914
Israel	0.912	0.714	0.776	0.819
Norway	0.939	0.769	0.889	0.893
Denmark	0.934	0.817	0.852	0.918

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.18. Confirmatory factor analysis of the model–data fit for the scale S1PRDEV

Participating country	CFI	TLI	RMSEA	SRMR
ISCED level 02				
Chile	0.973	0.964	0.042	0.033
Germany	0.945	0.926	0.057	0.041
Iceland	0.983	0.978	0.025	0.024
Israel	0.969	0.958	0.028	0.032
Japan	0.915	0.885	0.057	0.052
Korea	0.952	0.936	0.052	0.047
Norway	0.992	0.989	0.024	0.025
Turkey	0.958	0.943	0.030	0.037
Denmark	0.889	0.880	0.065	0.487
U3				
Germany	0.938	0.907	0.069	0.047
Israel	0.955	0.933	0.042	0.034
Norway	0.985	0.977	0.034	0.031
Denmark	0.819	0.797	0.072	0.509

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.19 provides the measurement invariance testing results for the scale S1PRDEV. The configural models exhibit acceptable fit for both populations, with the difference between the metric and configural models below the cut-off criteria and the difference between the scalar and metric models greater than the acceptable cut-off criteria. The scale can therefore be considered metric invariant for both populations.

Table 11.19. Invariance test results for the scale S1PRDEV

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.959	0.944	0.041	0.038				
Metric	0.944	0.935	0.045	0.095	0.015	0.009	-0.004	-0.057
Scalar	0.858	0.856	0.066	0.129	0.086	0.079	-0.021	-0.034
Invariance level for U3								
Configural	0.962	0.954	0.038	0.035				
Metric	0.954	0.949	0.040	0.074	0.008	0.005	-0.002	-0.039
Scalar	0.886	0.882	0.061	0.112	0.068	0.066	-0.021	-0.038

Note: An adjusted number of degrees of freedom was used to compute the CFI, TLI and RMSEA for U3 (see the subsection *Treating missing items in scales* in this chapter).
Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.20 presents the unstandardised item parameters for the scale S1PRDEV.

Table 11.20. Unstandardised factor loadings for the scale S1PRDEV

Items	Factor loadings
SS1G30A	0.274
SS1G30B	0.417
SS1G30C	0.507
SS1G30D	0.526
SS1G29A	0.402
SS1G29B	0.385
SS1G29C	0.452
SS1G29D	0.421*
SS1G29E	0.365
SS1G29F	0.374
SS1G29G	0.396*
SS1G29H	0.390

Note: *This item was not administered to the U3 population.
Source: OECD, TALIS Starting Strong 2018 database.

The standardised factor loadings, unstandardised intercepts and standardised latent covariances for S1PRDEV are presented in Table 11.21, Table 11.22 and Table 11.23, respectively. As is evident from the first two tables, the factor loadings of items are moderate to high for all items in almost all populations. Exceptions include item SS1G29A for the Japan ISCED level 02 and Israel U3 populations, and item SS1G30A for the Chile ISCED level 02 population. The factor loadings and intercepts for the items with phantom indicators in the U3 populations in both tables are zero. As the scale S1PRDEV was modelled to have a multidimensional structure, Table 11.23 provides the standardised latent covariances for all pairs of subscales. The covariances range from 0.48 to 0.70 for S1PRINT and S1PREM, 0.45 to 0.71 for S1PRSOC and S1PREM, and 0.63 to 0.84 for S1PRSOC and S1PRINT.

Table 11.21. Standardised factor loadings for the scale S1PRDEV

	Standardised factor loadings											
Participating country	SS1G29A	SS1G29B	SS1G29C	SS1G29D	SS1G29E	SS1G29F	SS1G29G	SS1G29H	SS1G30A	SS1G30B	SS1G30C	SS1G30D
ISCED level 02												
Chile	0.662	0.553	0.786	0.670	0.763	0.816	0.629	0.636	0.441	0.742	0.904	0.883
Germany	0.494	0.504	0.604	0.560	0.653	0.716	0.584	0.666	0.480	0.752	0.897	0.909
Iceland	0.559	0.564	0.595	0.528	0.551	0.723	0.497	0.718	0.533	0.718	0.872	0.759
Israel	0.479	0.531	0.691	0.634	0.669	0.767	0.479	0.630	0.455	0.725	0.837	0.636
Japan	0.377	0.488	0.665	0.530	0.666	0.734	0.566	0.675	0.443	0.707	0.909	0.931
Korea	0.640	0.610	0.746	0.672	0.748	0.800	0.618	0.803	0.550	0.776	0.889	0.909
Norway	0.669	0.702	0.747	0.694	0.757	0.808	0.575	0.764	0.462	0.777	0.903	0.828
Turkey	0.535	0.543	0.816	0.629	0.852	0.851	0.519	0.593	0.332	0.779	0.925	0.514
Denmark	0.606	0.600	0.776	0.646	0.674	0.884	0.576	0.640	0.463	0.835	0.927	0.942
U3												
Germany	0.523	0.529	0.640	-	0.644	0.753	-	0.758	0.517	0.754	0.888	0.909
Israel	0.414	0.595	0.708	-	0.624	0.735	-	0.712	0.620	0.717	0.791	0.577
Norway	0.615	0.652	0.738	-	0.802	0.849	-	0.874	0.489	0.783	0.900	0.794
Denmark	0.621	0.584	0.842	-	0.670	0.875	-	0.603	0.599	0.826	0.901	0.887

Note: The U3 population was given phantom indicators for the non-administered items SS1G29D and SS1G29G. Therefore, factor loadings were fixed to 0 (denoted by "-" in the table).
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.22. Unstandardised intercepts for the scale S1PRDEV

	Unstandardised intercepts											
Participating country	SS1G29A	SS1G29B	SS1G29C	SS1G29D	SS1G29E	SS1G29F	SS1G29G	SS1G29H	SS1G30A	SS1G30B	SS1G30C	SS1G30D
ISCED level 02												
Chile	3.313	3.132	3.469	3.286	3.773	3.786	3.512	3.472	3.552	3.677	3.704	3.635
Germany	3.159	3.018	3.341	3.417	3.430	3.596	3.134	3.503	3.465	3.460	3.334	3.343
Iceland	3.282	3.522	3.269	3.287	3.700	3.819	3.546	3.795	3.760	3.663	3.742	3.625
Israel	3.308	3.530	3.657	3.650	3.882	3.915	3.669	3.859	3.780	3.784	3.799	3.648
Japan	3.422	3.451	3.775	3.527	3.460	3.564	2.944	3.344	3.605	3.691	3.617	3.650
Korea	3.443	3.536	3.522	3.338	3.540	3.605	3.275	3.603	3.754	3.724	3.518	3.552
Norway	2.990	3.217	3.159	3.003	3.570	3.597	3.164	3.610	3.444	3.478	3.430	3.244
Turkey	3.741	3.773	3.880	3.825	3.948	3.946	3.772	3.848	3.720	3.901	3.915	3.766
Denmark	3.229	3.364	3.723	3.464	3.668	3.875	3.405	3.575	3.732	3.835	3.764	3.772
U3												
Germany	3.159	3.018	3.341	-	3.430	3.596	-	3.503	3.465	3.460	3.334	3.343
Israel	3.259	3.634	3.680	-	3.812	3.854	-	3.859	3.881	3.701	3.703	3.402
Norway	2.990	3.217	3.159	-	3.570	3.597	-	3.610	3.444	3.478	3.430	3.244
Denmark	3.444	3.346	3.784	-	3.674	3.863	-	3.563	3.879	3.822	3.719	3.738

Note: The U3 population was given phantom indicators for the non-administered items SS1G29D and SS1G29G. Therefore, factor loadings were fixed to 0 (denoted by "-" in the table).
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.23. Standardised latent covariances for the scale S1PRDEV

Participating country	S1PRINT & S1PREM	S1PRSOC & S1PREM	S1PRSOC & S1PRINT
ISCED level 02			
Chile	0.552	0.599	0.773
Germany	0.614	0.552	0.751
Iceland	0.606	0.594	0.693
Israel	0.599	0.542	0.715
Japan	0.486	0.457	0.676
Korea	0.551	0.514	0.849
Norway	0.522	0.492	0.638
Turkey	0.700	0.663	0.822
Denmark	0.662	0.634	0.706
U3			
Germany	0.605	0.582	0.656
Israel	0.608	0.712	0.710
Norway	0.547	0.556	0.720
Denmark	0.494	0.575	0.633

Source: OECD, TALIS Starting Strong 2018 database.

Multidimensional scale – Facilitating learning and development (overall) (S1PROLD), with the following subscales: Facilitating language development (S1PRLAN); Facilitating literacy development (S1PROLI); and Facilitating numeracy development (S1PRONU)

Items

The overall scale *Facilitating learning and development* (S1PROLD) was based on items from three questions, with respondents having to indicate the extent to which the statements applied to the staff at their centre. The first group of items, which was about facilitating language development, formed the subscale S1PRLAN. The second group of items, which focused on facilitating literacy development, formed the subscale S1PROLI, and the last group, on facilitating numeracy development, formed the subscale S1PRONU. The wording of these three sets of items is presented in Table 11.24.

Table 11.24. Item wording for the overall scale S1PROLD and subscales S1PRLAN, S1PROLI and S1PRONU

S1PROLD: Facilitating learning and development (overall)	
S1PRLAN: Facilitating language development (subscale)	
SS1G31: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G31A	<ECEC staff> encourage children to talk to each other
SS1G31B*+	<ECEC staff> encourage children to lead conversations
SS1G31C*+	<ECEC staff> support children in giving longer explanations by asking supportive questions
SS1G31D	<ECEC staff> position themselves at the children's height when talking or listening
SS1G31E	<ECEC staff> rephrase or recite statements to make sure children have been understood
SS1G31F	<ECEC staff> model the correct word rather than correcting the child directly
S1PROLI: Facilitating literacy development (subscale)	
SS1G32: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G32C	<ECEC staff> play word games with the children
SS1G32D	<ECEC staff> play with letters with the children
SS1G32E	<ECEC staff> sing songs or rhymes with the children
S1PRONU: Facilitating numeracy development (subscale)	
SS1G33: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G33A	<ECEC staff> use sorting activities by shape or colour
SS1G33B	<ECEC staff> play number games (e.g. dice)
SS1G33C	<ECEC staff> sing songs about numbers
SS1G33D	<ECEC staff> help children to use numbers or to count
SS1G33E	<ECEC staff> refer to groups of objects by the size of the group (i.e. number of objects in the group)

Notes: This scale was modelled with several items that also modelled the scale S1ACTIV. Therefore, analysis should avoid using these two scales together due to possible collinearity of the scales.
*This item was not included in the U3 version of the questionnaire.
+This item was removed and is not included in any of the results presented for this scale.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

Items SS1G31B and SS1G31C were dropped from the multidimensional scale S1PROLD and from the respective subscale S1PRLAN. In addition, a cross-loading of SS1G32E with the subscale S1PRONU was added for Turkey in the multidimensional scale.

Scale reliabilities and model fits

Table 11.25 provides the scale reliabilities for the scale S1PROLD and the subscales S1PRLAN, S1PROLI and S1PRONU. The reliability coefficients for the subscale S1PRLAN are high in all countries in both populations except in Denmark (both populations). The coefficients for the subscale S1PROLI are likewise high in all countries except for Germany (both populations), while the coefficients for the subscale S1PRONU indicate high reliability. The composite scale S1PROLD has very high reliabilities in all countries in both populations.

Table 11.25. Reliability coefficients for the multidimensional scale S1PROLD and subscales S1PRLAN, S1PROLI AND S1PRONU

	Stratified alpha		Omega	
Participating country	S1PROLD	S1PRLAN	S1PROLI	S1PRONU
ISCED level 02				
Chile	0.933	0.821	0.828	0.876
Germany	0.914	0.783	0.759	0.810
Iceland	0.926	0.746	0.878	0.852
Israel	0.926	0.746	0.843	0.863
Japan	0.912	0.755	0.794	0.846
Korea	0.947	0.852	0.895	0.870
Norway	0.929	0.797	0.850	0.854
Turkey	0.942	0.774	0.960	0.859
Denmark	0.906	0.654	0.794	0.835
U3				
Germany	0.917	0.785	0.803	0.808
Israel	0.924	0.760	0.805	0.826
Norway	0.928	0.814	0.817	0.837
Denmark	0.886	0.621	0.785	0.843

Source: OECD, TALIS Starting Strong 2018 database.

The model fits for S1PROLD are presented in Table 11.26. The model fit is good in all countries at ISCED level 02 (except for the TLI for Denmark and Germany).

Table 11.26. Confirmatory factor analysis of the model–data fit for the scale S1PROLD

Participating country	CFI	TLI	RMSEA	SRMR
ISCED level 02				
Chile	0.936	0.917	0.052	0.045
Germany	0.917	0.892	0.060	0.047
Iceland	0.966	0.957	0.038	0.036
Israel	0.943	0.926	0.044	0.044
Japan	0.953	0.939	0.041	0.040
Korea	0.938	0.920	0.053	0.058
Norway	0.965	0.955	0.044	0.039
Turkey	0.938	0.918	0.030	0.037
Denmark	0.918	0.893	0.060	0.048
U3				
Germany	0.833	0.784	0.088	0.072
Israel	0.871	0.833	0.067	0.062
Norway	0.907	0.880	0.068	0.055
Denmark	0.811	0.756	0.085	0.085

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

The invariance testing results for S1PROLD, presented in Table 11.27, show that the scale model reached the configural invariance level in both populations. The table shows a good model fit for nearly all of the ISCED level 02 populations. Although the CFI and TLI values

for the U3 population are below the cut-off criteria, they approach the criteria, a result that led to the decision to accept this scale at the configural level for the U3 population.

Table 11.27. Invariance test results for the scale S1PROLD

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.942	0.925	0.045	0.043				
Metric	0.918	0.908	0.050	0.104	0.024	0.017	-0.005	-0.061
Scalar	0.789	0.792	0.075	0.153	0.129	0.116	-0.025	-0.049
Invariance level for U3								
Configural	0.889	0.856	0.068	0.059				
Metric	0.877	0.854	0.068	0.077	0.012	0.002	0.000	-0.018
Scalar	0.764	0.740	0.091	0.114	0.113	0.114	-0.023	-0.037

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

The unstandardised item parameters for the S1PROLD subscales are not presented, as the scale reached configural invariance and the parameters are country specific.

Table 11.28, Table 11.29 and Table 11.30 present the standardised factor loadings, unstandardised intercepts and standardised latent covariances, respectively, for S1PROLD. The factor loadings are generally high (mostly above 0.60) for all items with the exception of item SS1G32D, where the item standardised factor loadings are below 0.60 and even below 0.45 in half of the participating countries, indicating a weak relationship between the item and the latent factor.

As the scale was modelled to have a multidimensional structure, Table 11.30 provides the standardised latent covariances for all pairs of subscales. The covariances range between 0.19 and 0.80 for S1PROLI and S1PRLAN, between 0.30 and 0.72 for S1PRONU and S1PRLAN, and between 0.45 and 0.91 for S1PRONU and S1PROLI.

Table 11.28. Standardised factor loadings for the scale S1PROLD

	Standardised factor loadings											
Participating country	SS1G31A	SS1G31D	SS1G31E	SS1G31F	SS1G32C	SS1G32D	SS1G32E	SS1G33A	SS1G33B	SS1G33C	SS1G33D	SS1G33E
ISCED level 02												
Chile	0.629	0.618	0.804	0.717	0.744	0.804	0.542	0.647	0.794	0.710	0.769	0.758
Germany	0.532	0.661	0.747	0.669	0.671	0.567	0.501	0.657	0.678	0.596	0.633	0.599
Iceland	0.462	0.664	0.716	0.581	0.856	0.875	0.441	0.658	0.757	0.684	0.674	0.677
Israel	0.547	0.594	0.591	0.705	0.835	0.785	0.544	0.614	0.665	0.688	0.798	0.715
Japan	0.491	0.596	0.731	0.645	0.782	0.713	0.569	0.709	0.794	0.654	0.653	0.662
Korea	0.672	0.770	0.805	0.717	0.839	0.912	0.387	0.759	0.718	0.709	0.717	0.702
Norway	0.585	0.687	0.788	0.611	0.822	0.821	0.415	0.709	0.644	0.684	0.707	0.672
Turkey	0.541	0.766	0.541	0.449	0.979	0.601	0.075 0.623*	0.745	0.585	0.742	0.629	0.714
Denmark	0.387	0.536	0.603	0.476	0.765	0.726	0.491	0.721	0.746	0.652	0.591	0.666
U3												
Germany	0.507	0.664	0.775	0.650	0.581	0.562	0.448	0.514	0.631	0.630	0.711	0.618
Israel	0.469	0.670	0.628	0.640	0.553	0.502	0.581	0.647	0.529	0.735	0.663	0.665
Norway	0.523	0.714	0.813	0.628	0.780	0.752	0.426	0.668	0.566	0.682	0.681	0.670
Denmark	0.259	0.562	0.642	0.492	0.695	0.797	0.155	0.682	0.773	0.640	0.679	0.618

*Note:**Because of improvements to this model, this item has more than one parameter estimate. For details, see the subsection titled *Model improvements* for this scale.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.29.Unstandardised intercepts for the scale S1PROLD

	Unstandardised Intercepts											
Participating country	SS1G31A	SS1G31D	SS1G31E	SS1G31F	SS1G32C	SS1G32D	SS1G32E	SS1G33A	SS1G33B	SS1G33C	SS1G33D	SS1G33E
ISCED level 02												
Chile	3.707	3.559	3.709	3.614	3.581	3.591	3.765	3.663	3.523	3.517	3.770	3.674
Germany	3.632	3.401	3.344	3.503	3.138	2.530	3.638	3.042	3.178	2.569	3.434	3.343
Iceland	3.827	3.375	3.475	3.540	3.473	3.468	3.867	3.366	3.227	3.300	3.692	3.278
Israel	3.792	3.621	3.414	3.581	3.572	3.492	3.782	3.489	3.556	3.260	3.749	3.636
Japan	3.569	3.809	3.635	3.434	3.410	3.047	3.482	3.073	2.860	2.755	3.340	2.950
Korea	3.593	3.657	3.621	3.516	3.257	3.290	3.755	3.474	3.422	3.328	3.639	3.386
Norway	3.470	3.122	3.130	3.341	2.769	2.700	3.450	2.701	2.892	2.601	3.163	2.818
Turkey	3.892	3.889	3.815	3.750	3.737	3.454	3.964 3.964*	3.911	3.786	3.900	3.863	3.920
Denmark	3.819	3.540	3.538	3.722	3.326	3.019	3.760	2.758	2.784	2.925	3.474	3.177
U3												
Germany	3.632	3.507	3.420	3.554	3.131	2.296	3.744	3.152	2.746	2.609	3.230	3.300
Israel	3.615	3.687	3.475	3.481	3.144	2.346	3.876	3.421	2.313	3.030	3.307	3.437
Norway	3.380	3.215	3.230	3.409	2.802	2.568	3.698	2.752	2.357	2.740	3.208	2.834
Denmark	3.685	3.647	3.558	3.739	3.228	2.609	3.954	2.733	2.403	3.031	3.294	2.888

Note: *Because of improvements to this model, this item has more than one parameter estimate. For details, see the subsection titled *Model improvements* for this scale.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.30. Standardised latent covariances for the scale S1PROLD

Participating country	S1PROLI & S1PRLAN	S1PRONU & S1PRLAN	S1PRONU & S1PROLI
ISCED level 02			
Chile	0.593	0.438	0.746
Germany	0.606	0.526	0.834
Iceland	0.448	0.510	0.754
Israel	0.513	0.586	0.772
Japan	0.557	0.417	0.645
Korea	0.540	0.606	0.738
Norway	0.454	0.496	0.821
Turkey	0.464	0.721	0.455
Denmark	0.621	0.485	0.726
U3			
Germany	0.620	0.414	0.919
Israel	0.809	0.608	0.872
Norway	0.528	0.527	0.835
Denmark	0.194	0.305	0.769

Source: OECD, TALIS Starting Strong database.

Satisfaction with work environment or working conditions (S1JOBSA)

Items

The scale *Satisfaction with work environment or working conditions* (S1JOBSA) was based on items from one question. Respondents had to indicate their (dis)agreement with statements about different aspects of their work. Table 11.31 presents the wording of these items and their response options.

Table 11.31. Item wording for the scale S1JOBSA

S1JOBSA: Satisfaction with work environment or working conditions	
SS1G44: We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SS1G44B	I am satisfied with the salary I receive for my work
SS1G44C	Apart from my salary, I am satisfied with the terms of my <contract/employment> (e.g. benefits, work schedule)
SS1G44D	I enjoy working at this <ECEC centre>
SS1G44E	I am satisfied with the support that I receive from parents or guardians in this <ECEC centre>
SS1G44F*	I need more support from my <ECEC centre leader>
SS1G44H	I would recommend this <ECEC centre> as a good place to work
SS1G44K	All in all, I am satisfied with my job

Note: *This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between SS1G44B and SS1G44C was added for all countries.

Scale reliabilities and model fits

The reliability coefficients and the model fit indices for the scale S1JOBSA, presented in Table 11.32, show that the coefficients are above the cut-off criterion of 0.70 in all countries in both populations, while the model fit indices indicate a good model fit. However, the SRMR index is above the cut-off value in Denmark for both populations.

Table 11.32. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1JOBSA

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.834	0.975	0.960	0.046	0.025
Germany	0.841	0.951	0.921	0.070	0.034
Iceland	0.843	0.971	0.953	0.048	0.027
Israel	0.806	0.948	0.917	0.053	0.038
Japan	0.846	0.970	0.951	0.050	0.033
Korea	0.839	0.946	0.912	0.059	0.051
Norway	0.863	0.942	0.907	0.075	0.043
Turkey	0.839	0.943	0.909	0.048	0.044
Denmark	0.863	0.917	0.913	0.052	0.231
U3					
Germany	0.834	0.957	0.931	0.061	0.032
Israel	0.769	0.980	0.968	0.032	0.026
Norway	0.852	0.965	0.944	0.056	0.030
Denmark	0.899	0.938	0.935	0.054	0.197

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.33 presents the invariance testing results for the S1JOBSA scale. The scale reached metric invariance for ISCED level 02 and configural invariance for U3.

Table 11.33. Invariance test results for the scale S1JOBSA

S1JOBSA	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.956	0.929	0.054	0.037				
Metric	0.926	0.915	0.060	0.093	0.030	0.014	-0.006	-0.056
Scalar	0.787	0.809	0.089	0.126	0.139	0.106	-0.029	-0.033
Invariance level for U3								
Configural	0.966	0.944	0.051	0.030				
Metric	0.930	0.913	0.063	0.087	0.036	0.031	-0.012	-0.057
Scalar	0.787	0.787	0.099	0.109	0.143	0.126	-0.036	-0.022

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

The unstandardised factor loadings for S1JOBSA are presented in Table 11.34. The standardised factor loadings for this scale can be found in Table 11.35 and the unstandardised intercepts in Table 11.36.

Table 11.34. Unstandardised factor loadings for the scale S1JOBSA

Items	Factor loadings
SS1G44B	0.207
SS1G44C	0.341
SS1G44D	0.534
SS1G44E	0.331
SS1G44F	0.302
SS1G44H	0.609
SS1G44K	0.468

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.35. Standardised factor loadings for the scale S1JOBSA

	Standardised factor loadings						
Participating country	SS1G44B	SS1G44C	SS1G44D	SS1G44E	SS1G44F	SS1G44H	SS1G44K
ISCED level 02							
Chile	0.212	0.375	0.802	0.403	0.290	0.781	0.733
Germany	0.243	0.431	0.819	0.427	0.337	0.792	0.706
Iceland	0.244	0.406	0.804	0.465	0.317	0.802	0.722
Israel	0.177	0.314	0.781	0.391	0.231	0.759	0.688
Japan	0.292	0.454	0.814	0.538	0.424	0.764	0.741
Korea	0.258	0.433	0.826	0.503	0.377	0.795	0.617
Norway	0.231	0.405	0.852	0.503	0.306	0.810	0.720
Turkey	0.240	0.394	0.785	0.462	0.341	0.812	0.720
Denmark	0.277	0.482	0.831	0.515	0.377	0.785	0.786
U3							
Germany	0.228	0.438	0.807	0.427	0.328	0.770	0.726
Israel	0.192	0.278	0.753	0.399	0.238	0.671	0.679
Norway	0.229	0.390	0.838	0.481	0.297	0.806	0.705
Denmark	0.269	0.482	0.885	0.513	0.406	0.855	0.785

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.36. Unstandardised intercepts for the scale S1JOBSA

	Unstandardised intercepts						
Participating country	SS1G44B	SS1G44C	SS1G44D	SS1G44E	SS1G44F	SS1G44H	SS1G44K
ISCED level 02							
Chile	2.085	2.885	3.628	3.124	2.489	3.430	3.592
Germany	2.021	2.938	3.568	2.962	2.874	3.348	3.358
Iceland	1.544	2.886	3.504	3.257	2.946	3.446	3.419
Israel	2.119	2.929	3.706	3.421	2.957	3.629	3.670
Japan	1.921	2.360	3.056	2.859	2.571	2.760	2.953
Korea	2.190	2.380	2.966	2.937	2.135	2.857	2.935
Norway	2.108	2.778	3.596	3.399	2.780	3.469	3.448
Turkey	2.257	2.628	3.300	3.020	2.593	3.270	3.475
Denmark	2.234	2.958	3.574	3.225	2.921	3.447	3.483
U3							
Germany	2.021	2.938	3.568	2.962	2.874	3.348	3.358
Israel	1.765	2.465	3.625	3.333	2.556	3.440	3.518
Norway	2.108	2.778	3.596	3.399	2.780	3.469	3.448
Denmark	2.243	2.927	3.591	3.221	2.918	3.499	3.507

Source: OECD, TALIS Starting Strong 2018 database.

Multidimensional scale: Sources of work stress (S1STRES)

Items

The *Sources of work stress* (S1STRES) scale was based on items from one question (see Table 11.37). Respondents had to indicate the level of intensity for different sources of stress as they perceived them in their work.

Table 11.37. Item wording for the scale S1STRES

S1STRES: Sources of work stress	
SS1G45: Thinking about your job at this <ECEC centre>, to what extent are the following sources of stress in your work?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G45A	Having too much preparation work for children's activities
SS1G45B	Having too much work related to documenting children's development
SS1G45C	Having too much administrative work to do (e.g. filling out forms)
SS1G45D	Having extra duties due to absent <ECEC staff>
SS1G45E	Being held responsible for children's development, well-being and learning
SS1G45F	Managing <classroom/playgroup/group> behaviour
SS1G45G	Keeping up with changing requirements from <local, municipality/regional, state, or national/federal> authorities
SS1G45H	Addressing parent or guardian concerns
SS1G45I	Accommodating children with special needs
SS1G45J	Having too many children in my <classroom/playgroup/group>
SS1G45K	Having too many additional duties (e.g. cleaning)
SS1G45L	A lack of resources (e.g. financial support, material resources, <ECEC staff>)

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

Testing of this scale revealed a multidimensional structure. The subsequent scale model therefore has a hierarchical structure, in which the overall latent construct S1STRES is indicated by three subscales named, for simplicity, S1STRESa, S1STRESb, and S1STRESc. The items for S1STRESa are SS1G45A, SS1G45B, and SS1G45C; for S1STRESb are SS1G45D, SS1G45J, SS1G45K, and SS1G45L; and for S1STRESc are SS1G45E, SS1G45F, SS1G45G, SS1G45H, and SS1G45I.

In addition to the development of a hierarchical structure for this scale, several improvements were added for various countries. A cross-loading of item SS1G45A with the subscale S1STRESc was added to the model for Chile. A cross-loading of item SS1G45A with the subscale S1STRESb was added to the model for Iceland as was a cross-loading of item SS1G45G with the second-order, overall latent construct S1STRES. Finally, a cross-loading of item SS1G45G with the subscale S1STRESa was added to the U3 population for Norway.

Scale reliabilities and model fits

Table 11.38, which presents the reliability coefficients for the scale S1STRES, shows high reliability coefficients in all participating countries except Germany and Norway at ISCED level 02. The model fit indices indicate a good model fit in most countries, with the exception of Korea's ISCED level 02 population and Germany's and Israel's U3 populations.

Table 11.38. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1STRES

Participating country	Omega*	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.702	0.939	0.920	0.057	0.044
Germany	0.663	0.905	0.878	0.059	0.052
Iceland	0.702	0.956	0.941	0.050	0.037
Israel	0.776	0.924	0.901	0.070	0.057
Japan	0.778	0.930	0.909	0.060	0.060
Korea	0.769	0.840	0.793	0.081	0.072
Norway	0.694	0.955	0.942	0.052	0.050
Turkey	0.785	0.884	0.849	0.063	0.063
Denmark	0.738	0.901	0.901	0.069	0.105
U3					
Germany	0.712	0.877	0.841	0.072	0.056
Israel	0.723	0.806	0.806	0.083	0.185
Norway	0.785	0.933	0.911	0.055	0.048
Denmark	0.740	0.894	0.894	0.062	0.100

Note: *The omega coefficients are based on the hierarchical model, where the variances of specific factors were considered as an error (in contrast to the stratified Cronbach's alpha).
Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

As is evident from the results of the invariance testing presented in Table 11.39, the model fits and the differences between them indicate that the scale reached the metric level of invariance in both populations.

Table 11.39. Invariance test results for the scale S1STRES

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.914	0.888	0.065	0.062				
Metric	0.883	0.878	0.068	0.121	-0.031	-0.010	0.003	0.059
Scalar	0.722	0.75	0.097	0.162	-0.161	-0.128	0.029	0.041
Invariance level for U3								
Configural	0.904	0.874	0.065	0.053				
Metric	0.888	0.873	0.065	0.076	-0.016	-0.001	0.000	0.023
Scalar	0.808	0.800	0.081	0.100	-0.080	-0.073	0.016	0.024

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.40 presents the unstandardised factor loadings for the S1STRES scale. The standardised factor loadings, unstandardised intercepts and first order factor loadings for the scale S1STRES are presented in Table 11.41, Table 11.42 and Table 11.43, respectively.

The standardised factor loadings for all items in all participating countries for both populations are above 0.45, and many are above 0.60, thus indicating a strong relationship between all the items and the factor in all participating countries.

As the scale was modelled to have a hierarchical structure, the factor loadings of the subscales (first order factors) on the composite scale (second order factor) were examined. The loadings for S1STRESa range between 0.59 and 0.81, for S1STRESb between 0.71 and 0.92 and for S1STRESc between 0.69 and 0.89.

Table 11.40. Unstandardised factor loadings for the scale S1STRES

Items	Factor loadings
SS1G45A	0.420
SS1G45B	0.643
SS1G45C	0.606
SS1G45D	0.328
SS1G45J	0.419
SS1G45K	0.465
SS1G45L	0.457
SS1G45E	0.443
SS1G45F	0.426
SS1G45G	0.357
SS1G45H	0.410
SS1G45I	0.388
SS1G45A	0.104
S1STRESA	0.947
S1STRESB	1.368
S1STRESC	1.282

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.41. Standardised factor loadings for the scale S1STRES

	Standardised factor loadings											
Participating country	SS1G45A	SS1G45B	SS1G45C	SS1G45D	SS1G45E	SS1G45F	SS1G45G	SS1G45H	SS1G45I	SS1G45J	SS1G45K	SS1G45L
ISCED level 02												
Chile	0.637 0.182*	0.935	0.818	0.475	0.749	0.721	0.640	0.742	0.654	0.557	0.699	0.610
Germany	0.574	0.759	0.642	0.537	0.654	0.622	0.500	0.666	0.471	0.596	0.712	0.635
Iceland	0.450 0.228*	0.868	0.829	0.496	0.751	0.635	0.530 0.231*	0.754	0.656	0.635	0.745	0.731
Israel	0.708	0.932	0.800	0.614	0.771	0.756	0.633	0.742	0.658	0.664	0.755	0.756
Japan	0.661	0.906	0.839	0.558	0.782	0.773	0.608	0.766	0.687	0.633	0.733	0.643
Korea	0.651	0.859	0.769	0.525	0.713	0.663	0.502	0.594	0.514	0.659	0.808	0.754
Norway	0.676	0.943	0.784	0.551	0.722	0.645	0.561	0.727	0.582	0.697	0.713	0.745
Turkey	0.632	0.884	0.780	0.569	0.769	0.727	0.647	0.710	0.596	0.613	0.725	0.667
Denmark	0.637	0.903	0.833	0.608	0.771	0.729	0.577	0.771	0.657	0.664	0.755	0.757
U3												
Germany	0.583	0.738	0.608	0.553	0.687	0.661	0.522	0.680	0.517	0.571	0.703	0.636
Israel	0.688	0.889	0.812	0.594	0.735	0.706	0.558	0.715	0.568	0.630	0.743	0.666
Norway	0.709	0.921	0.820	0.578	0.753	0.722	0.529 0.238*	0.759	0.646	0.719	0.754	0.762
Denmark	0.678	0.894	0.790	0.600	0.763	0.771	0.598	0.782	0.645	0.675	0.773	0.731

Note: *Because of improvements to this model, this item has more than one parameter estimate. For details, see the subsection titled *Model improvements* for this scale.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.42. Unstandardised intercepts for the scale S1STRES

	Unstandardised intercepts											
Participating country	SS1G45A	SS1G45B	SS1G45C	SS1G45D	SS1G45E	SS1G45F	SS1G45G	SS1G45H	SS1G45I	SS1G45J	SS1G45K	SS1G45L
ISCED level 02												
Chile	2.333 2.333*	2.305	2.214	2.034	2.198	2.231	1.972	2.126	2.023	2.216	1.786	2.285
Germany	2.278	2.913	2.550	3.005	2.352	2.309	2.423	2.370	2.080	2.743	2.648	2.771
Iceland	2.409 2.409*	1.752	1.657	2.112	1.873	2.254	2.054 2.054*	1.994	2.113	2.852	2.169	2.605
Israel	2.134	2.182	2.024	1.939	2.356	2.585	2.393	2.142	2.102	2.639	2.115	2.382
Japan	2.209	2.271	2.484	1.780	2.424	2.210	2.026	2.363	2.288	1.902	2.016	2.542
Korea	2.411	2.761	2.962	1.950	2.367	2.546	2.615	2.792	2.090	2.636	2.673	2.746
Norway	2.095	2.105	2.074	2.650	1.937	2.146	2.216	1.963	2.113	2.823	2.370	2.702
Turkey	2.313	2.470	2.282	1.662	2.325	2.205	2.092	2.517	1.995	2.108	1.858	2.381
Denmark	2.202	2.611	2.543	2.669	2.048	2.313	2.309	1.962	2.524	2.543	2.132	2.923
U3												
Germany	2.278	2.913	2.550	3.005	2.352	2.309	2.423	2.370	2.080	2.743	2.648	2.771
Israel	1.807	1.784	1.431	2.048	2.339	2.518	2.058	2.141	1.735	2.205	2.395	2.657
Norway	2.095	2.105	2.074	2.650	1.937	2.146	2.216 2.216*	1.963	2.113	2.823	2.370	2.702
Denmark	2.156	2.534	2.481	2.623	2.073	2.226	2.197	1.921	2.194	2.602	2.289	2.775

Note: *Because of improvements to this model, this item has more than one parameter estimate. For details, see the subsection titled *Model improvements* for this scale.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.43. Second-order factor loadings on the first-order factors for the scale S1STRES

Participating country	S1STRES & S1STRESa	S1STRES & S1STRESb	S1STRES & S1STRESc
ISCED level 02			
Chile	0.592	0.890	0.691
Germany	0.721	0.710	0.774
Iceland	0.703	0.719	0.721
Israel	0.697	0.820	0.842
Japan	0.686	0.925	0.766
Korea	0.728	0.784	0.898
Norway	0.626	0.710	0.803
Turkey	0.779	0.869	0.824
Denmark	0.670	0.832	0.755
U3			
Germany	0.801	0.775	0.787
Israel	0.670	0.832	0.755
Norway	0.715	0.787	0.875
Denmark	0.670	0.832	0.755

Source: OECD, TALIS Starting Strong 2018 database.

Staff participation in collaborative professional development (S1PDNET)

Items

The scale *Staff participation in collaborative professional development* (S1PDNET) was based on items from one question. Respondents were requested to indicate if they had participated in any of the stated professional development activities. The wording of the items and their response options are presented in Table 11.44.

Table 11.44. Item wording for the scale S1PDNET

S1PDNET: Staff participation in collaborative PD	
SS1G13: During the last 12 months, did you participate in any of the following professional development activities?	
Response options: yes (1), no (2)	
SS1G13E*	Observation visits to other <ECEC centres>
SS1G13F*	Peer and/or self-observation and coaching as part of a formal arrangement
SS1G13H*	Participation in a network of professionals working with children
SS1G13I*	Induction or mentoring activities

Note: *This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Scale reliabilities and model fits

Table 11.45 presents the reliability coefficients and model fits for the scale S1PDNET, which was based on items with categorical response options (yes/no). The reliabilities are lower than those for the scales with items based on ordinal response options and are mostly above 0.50 (the only exception is Norway, ISCED level 02). Almost all model fits meet the cut-off criteria, thus showing a good model fit in all participating countries.

Table 11.45. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1PDNET

Participating country	Cronbach's alpha	CFI	TLI	RMSEA	WRMR
ISCED level 02					
Chile	0.554	0.985	0.956	0.049	0.807
Germany	0.540	0.992	0.975	0.036	0.629
Iceland	0.526	0.975	0.926	0.054	0.764
Israel	0.767	0.994	0.983	0.064	0.999
Japan	0.718	1.000	1.002	0.000	0.281
Korea	0.638	1.000	1.007	0.000	0.237
Norway	0.340	1.000	1.073	0.000	0.224
Turkey	0.707	1.000	0.999	0.010	0.384
Denmark	0.555	1.000	1.019	0.000	0.171
U3					
Germany	0.548	0.999	0.996	0.014	0.452
Israel	0.686	0.984	0.951	0.085	1.152
Norway	0.476	0.983	0.949	0.040	0.656
Denmark	0.496	0.976	0.929	0.052	0.617

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.46 contains the results from the invariance testing analyses for S1PDNET. As the scale was based on items with categorical response options, there is no metric model. The configural models show a good fit, but the difference in model fits between the configural and scalar models is above the cut-off criteria, meaning the scale is configural invariant in both populations.

Table 11.46. Invariance test results for the scale S1PDNET

	CFI	TLI	RMSEA	WRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ WRMR
Invariance level for ISCED level 02								
Configural	0.983	0.950	0.064	1.610				
Scalar	0.839	0.711	0.154	4.687	0.144	0.239	-0.090	-3.077
Invariance level for U3								
Configural	0.996	0.987	0.036	1.754				
Scalar	0.883	0.813	0.137	7.794	0.113	0.174	-0.101	-6.040

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Unstandardised parameters are not available for S1PDNET because the scale reached configural invariance. The item parameters presented in Table 11.47 show that most factor loadings are above 0.45 for all items (except for items SS1G13E and SS1G13F in Norway), thus indicating a strong relationship between them and the latent factor.

Table 11.47. Unstandardised thresholds and standardised factor loadings for the scale S1PDNET

	Standardised factor loadings				Unstandardised thresholds			
Participating country	SS1G13E	SS1G13F	SS1G13H	SS1G13I	SS1G13E	SS1G13F	SS1G13H	SS1G13I
ISCED level 02								
Chile	0.577	0.658	0.696	0.630	0.888	0.516	0.271	0.942
Germany	0.726	0.650	0.620	0.602	0.667	0.975	0.869	1.190
Iceland	0.453	0.700	0.733	0.533	-0.082	0.714	0.452	0.572
Israel	0.771	0.810	0.806	0.880	0.428	0.292	0.264	0.058
Japan	0.790	0.899	0.718	0.643	0.120	0.280	0.450	0.113
Korea	0.767	0.772	0.563	0.721	0.155	-0.287	0.851	-0.250
Norway	0.378	0.354	0.592	0.651	1.112	0.337	0.380	1.291
Turkey	0.745	0.885	0.684	0.801	0.231	0.731	0.608	0.826
Denmark	0.680	0.750	0.702	0.713	1.295	0.376	-0.057	1.342
U3								
Germany	0.763	0.731	0.602	0.570	0.754	1.179	0.907	1.211
Israel	0.684	0.744	0.842	0.841	1.126	0.795	0.815	0.164
Norway	0.513	0.692	0.526	0.727	1.039	0.306	0.454	1.362
Denmark	0.679	0.689	0.739	0.445	1.434	0.440	0.342	1.401

Source: OECD, TALIS Starting Strong 2018 database.

Engagement in collaborative professional practices (S1COLL)

Items

The scale *Engagement in collaborative professional practices* (S1COLL) was based on items from one question that required respondents to indicate how often they engaged in several listed activities at their centre. The items and their response options are presented in Table 11.48.

Table 11.48. Item wording for the scale S1COLL

S1COLL: Engagement in collaborative professional practices	
SS1G23: On average, how often do you engage in the following activities in this <ECEC centre>?	
Response options: never (1), less than monthly (2), monthly (3), weekly (4), daily (5)	
SS1G23A	Provide feedback to other <ECEC staff> about their practice
SS1G23B	Engage in discussions about approaches to children's development, well-being and learning
SS1G23C	Engage in joint activities across different groups, including age groups
SS1G23D	Exchange learning or pedagogical materials with colleagues
SS1G23E	Engage in discussions about planned activities
SS1G23F	Engage in discussions about the development or needs of specific children
SS1G23G	Work with other <ECEC staff> to discuss the evaluation of children's development and well-being

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SS1G23F and SS1G23G was added for all countries, and a correlation between items SS1G23D and SS1G23E was added for the ISCED level 02 populations of Israel, Japan and Korea.

Scale reliabilities and model fits

The reliabilities and model fits for the scale S1COLL are presented in Table 11.49, where the coefficients are high in all participating countries for both populations. The model fit indices indicate a good model fit for all participating countries, except Denmark (both populations).

Table 11.49. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1COLL

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.869	0.945	0.912	0.084	0.037
Germany	0.801	0.944	0.909	0.083	0.040
Iceland	0.841	0.956	0.928	0.068	0.035
Israel	0.850	0.906	0.798	0.120	0.048
Japan	0.814	0.959	0.928	0.058	0.035
Korea	0.880	0.956	0.923	0.070	0.043
Norway	0.821	0.914	0.861	0.094	0.048
Turkey	0.904	0.975	0.960	0.043	0.023
Denmark	0.918	0.661	0.644	0.133	0.650
U3					
Germany	0.799	0.951	0.921	0.069	0.035
Israel	0.839	0.967	0.938	0.062	0.033
Norway	0.794	0.942	0.906	0.072	0.039
Denmark	0.906	0.753	0.741	0.104	0.466

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

The scale S1COLL reached the metric invariance level in both populations, as can be seen from Table 11.50.

Table 11.50. Invariance test results for the scale S1COLL

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance models for ISCED level 02								
Configural	0.924	0.882	0.030	0.044				
Metric	0.909	0.899	0.028	0.071	0.015	-0.016	0.002	-0.027
Scalar	0.762	0.794	0.039	0.126	0.146	0.104	-0.011	-0.055
Invariance models for U3								
Configural	0.925	0.887	0.081	0.042				
Metric	0.911	0.897	0.078	0.062	0.013	-0.010	0.004	-0.020
Scalar	0.788	0.798	0.109	0.104	0.124	0.099	-0.031	-0.042

Note: An adjusted number of degrees of freedom was used to compute the CFI, TLI and RMSEA for the ISCED level 02 and U3 populations (see the subsection *Treating missing items in scales* in this chapter).
Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.51 presents the unstandardised factor loadings for the scale S1COLL.

Table 11.51. Unstandardised factor loadings for the scale S1COLL

Items	Factor loadings
SS1G23A	0.787
SS1G23B	0.956
SS1G23C	0.734
SS1G23D	0.852
SS1G23E	0.897
SS1G23F	0.900
SS1G23G	0.869

Source: OECD, TALIS Starting Strong 2018 database.

The standardised factor loadings and unstandardised intercepts are presented in Table 11.52 and Table 11.53, respectively. The factor loadings are moderate to high for S1COLL in all participating countries. Exceptions include Germany, Israel and Norway for the U3 populations, where the factor loadings for item SS1G23C are below 0.45.

Table 11.52. Standardised factor loadings for the scale S1COLL

	Standardised factor loadings						
Participating country	SS1G23A	SS1G23B	SS1G23C	SS1G23D	SS1G23E	SS1G23F	SS1G23G
ISCED level 02							
Chile	0.601	0.764	0.609	0.635	0.745	0.752	0.721
Germany	0.476	0.654	0.492	0.593	0.721	0.690	0.593
Iceland	0.493	0.735	0.484	0.668	0.750	0.749	0.606
Israel*	0.543	0.764	-	0.644	0.733	0.754	0.650
Japan	0.525	0.728	0.476	0.565	0.686	0.735	0.680
Korea	0.660	0.829	0.646	0.735	0.758	0.725	0.743
Norway	0.492	0.742	0.472	0.526	0.702	0.723	0.644
Turkey	0.656	0.821	0.631	0.728	0.804	0.826	0.779
Denmark	0.653	0.889	0.572	0.712	0.832	0.848	0.689
U3							
Germany	0.479	0.692	0.444	0.605	0.686	0.678	0.592
Israel*	0.514	0.750	-	0.639	0.755	0.725	0.650
Norway	0.496	0.696	0.424	0.511	0.666	0.685	0.624
Denmark	0.636	0.862	0.577	0.682	0.819	0.814	0.691

Note: *There is no factor loading for item SS1G23C for Israel because it did not administer this item.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.53. Unstandardised intercepts for the scale S1COLL

	Unstandardised intercepts						
Participating country	SS1G23A	SS1G23B	SS1G23C	SS1G23D	SS1G23E	SS1G23F	SS1G23G
ISCED level 02							
Chile	2.973	3.134	2.926	3.373	3.385	3.527	3.169
Germany	2.980	3.612	3.752	3.315	3.956	4.054	3.452
Iceland	3.129	3.824	3.531	3.535	4.010	4.234	3.809
Israel*	3.039	3.104	-	2.958	3.418	3.349	2.823
Japan	2.495	3.721	3.444	3.228	3.755	3.928	3.792
Korea	2.867	3.060	3.000	3.176	3.514	3.614	3.250
Norway	3.872	4.183	3.723	3.122	3.938	4.026	3.914
Turkey	3.508	3.597	3.051	3.861	3.448	3.559	3.639
Denmark	3.626	4.483	3.312	2.950	4.115	4.395	3.485
U3							
Germany	2.980	3.612	3.752	3.315	3.956	4.054	3.452
Israel*	3.039	3.104	-	2.958	3.418	3.349	2.823
Norway	3.872	4.183	3.723	3.122	3.938	4.026	3.914
Denmark	3.455	4.299	3.034	3.020	3.923	4.166	3.432

Note: *There is no factor loading for item SS1G23C for Israel because it did not administer this item.
Source: OECD, TALIS Starting Strong 2018 database.

Activities to enhance the development of children's abilities and skills: Staff pedagogical practices (match to the needs) (S1PRACT)

Items

The wording of items included in the scale S1PRACT and their response options are presented in Table 11.54. Respondents had to indicate the frequency with which they performed specified activities with the children for a specific period of the day with a specific class of children (i.e. the "target group") in their centre.

Table 11.54. Item wording for the scale S1PRACT

S1PRACT: Activities to enhance the development of children's abilities and skills; Staff pedagogical practices (Match to the needs)	
SS1G41: Thinking about your work with the <target group>, how often do you do the following?	
Response options: never or almost never (1), occasionally (2), frequently (3), always or almost always (4)	
SS1G41A	I set daily goals for the children
SS1G41B	I explain how a new activity relates to children's lives
SS1G41C	I give different activities to suit different children's interests
SS1G41D	I give different activities to suit different children's level of development
SS1G41E	I adapt my activities to differences in children's cultural background

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

Correlations between SS1G41A and SS1G41B, and SS1G41C and SS1G41D were added for all countries.

Scale reliabilities and model fits

Table 11.55 presents the reliabilities and the model fits for S1PRACT. High reliability coefficients are observed in all countries except Germany (ISCED level 02 and U3) and Norway (U3). The CFI is acceptable or close to the acceptable value of 0.90 in most countries. The TLI and RMSEA do not meet the cut-off criteria in most countries but the SRMR shows a good fit in nearly all the countries.

Table 11.55. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1PRACT

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.740	0.910	0.820	0.114	0.051
Germany	0.608	0.862	0.725	0.142	0.068
Iceland	0.726	0.838	0.675	0.157	0.061
Israel	0.832	0.962	0.925	0.095	0.031
Japan	0.843	0.948	0.896	0.123	0.039
Korea	0.846	0.887	0.774	0.168	0.050
Norway	0.716	0.889	0.777	0.145	0.059
Turkey	0.828	0.958	0.916	0.069	0.032
Denmark	0.781	0.855	0.819	0.115	0.297
U3					
Germany	0.612	0.796	0.591	0.201	0.091
Israel	0.736	0.891	0.783	0.130	0.055
Norway	0.684	0.864	0.727	0.165	0.068
Denmark	0.771	0.817	0.772	0.114	0.343

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.56 presents the invariance testing results for the scale S1PRACT, which was found to be metric invariant for both populations.

Table 11.56. Invariance test results for the scale S1PRACT

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.989	0.965	0.058	0.015				
Metric	0.973	0.959	0.063	0.069	0.016	0.006	-0.005	-0.054
Scalar	0.821	0.821	0.130	0.151	0.152	0.138	-0.067	-0.082
Invariance level for U3								
Configural	0.995	0.984	0.038	0.012				
Metric	0.995	0.991	0.028	0.024	0.000	-0.007	0.010	-0.012
Scalar	0.915	0.898	0.096	0.082	0.080	0.093	-0.068	-0.058

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.57 presents the unstandardised factor loadings for the scale S1PRACT.

Table 11.57. Unstandardised factor loadings for the scale S1PRACT

Items	Factor loadings
SS1G41A	0.553
SS1G41B	0.615
SS1G41C	0.583
SS1G41D	0.577
SS1G41E	0.585

Source: OECD, TALIS Starting Strong 2018 database.

The standardised factor loadings and unstandardised intercepts for S1PRACT are presented in Table 11.58. Here, most of the factor loadings for all items in all participating countries are above 0.60 but are a bit lower for item SS1G41E (mostly between 0.45 and 0.60).

Table 11.58. Unstandardised intercepts and standardised factor loadings for the scale S1PRACT

	Standardised factor loadings					Unstandardised intercepts				
Participating country	SS1G41A	SS1G41B	SS1G41C	SS1G41D	SS1G41E	SS1G41A	SS1G41B	SS1G41C	SS1G41D	SS1G41E
ISCED level 02										
Chile	0.611	0.676	0.701	0.652	0.542	3.443	3.287	3.377	3.347	2.976
Germany	0.457	0.527	0.599	0.581	0.474	2.532	2.454	3.143	3.243	2.177
Iceland	0.531	0.625	0.681	0.675	0.541	2.375	2.618	2.924	3.181	2.278
Israel	0.702	0.784	0.791	0.751	0.579	3.171	3.077	3.270	3.259	2.239
Japan	0.675	0.765	0.821	0.819	0.672	2.907	2.927	3.189	3.168	1.990
Korea	0.680	0.770	0.798	0.809	0.728	3.132	2.973	3.239	3.239	2.882
Norway	0.484	0.622	0.714	0.678	0.526	2.568	2.291	3.170	3.125	2.101
Turkey	0.687	0.794	0.771	0.731	0.575	3.569	3.478	3.453	3.473	3.119
Denmark	0.605	0.657	0.750	0.740	0.565	2.807	2.551	3.086	3.377	2.189
U3										
Germany	0.472	0.542	0.609	0.619	0.463	2.532	2.454	3.143	3.243	2.177
Israel	0.594	0.642	0.738	0.678	0.495	3.192	2.971	3.358	3.378	1.989
Norway	0.493	0.579	0.697	0.659	0.513	2.568	2.291	3.170	3.125	2.101
Denmark	0.578	0.658	0.745	0.744	0.545	2.821	2.349	3.133	3.446	2.086

Source: OECD, TALIS Starting Strong 2018 database.

Behavioural support (S1SOCIAL)

Items

Table 11.59 shows the item wording and responses for the scale S1SOCIAL. Respondents had to indicate the frequency with which they exhibited certain behaviours while working with the "target group" children in their centre.

Table 11.59. Item wording for the scale S1SOCIAL

S1SOCIAL: Behavioural support	
SS1G41: Thinking about your work with the <target group>, how often do you do the following?	
Response options: never or almost never (1), occasionally (2), frequently (3), always or almost always (4)	
SS1G41G	I help children to follow the rules
SS1G41H	I calm children who are upset
SS1G41I	When the activities begin, I ask children to quieten down
SS1G41J*	I address children's disruptive behaviour that slows down other children's learning
SS1G41L	I help children understand the consequences if they do not follow the rules

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between SS1G41G and SS1G41H was added for all countries. However, because this improvement caused a non-convergence problem in the final multigroup model, it was not used for factor score estimation.

Scale reliabilities and model fits

Table 11.60 presents the reliabilities and the model fits for the scale S1SOCIAL. The reliability coefficients are high in all participating countries for both populations, and the model fits are good for all countries except Korea for ISCED level 02 and Denmark for U3.

Table 11.60. Reliability coefficients and confirmatory factor analysis of model–data fit for the scale S1SOCIAL

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.832	0.983	0.958	0.062	0.020
Germany	0.797	0.965	0.913	0.078	0.026
Iceland	0.850	0.985	0.964	0.052	0.015
Israel	0.828	0.996	0.990	0.027	0.009
Japan	0.882	0.975	0.937	0.064	0.022
Korea	0.841	0.874	0.860	0.107	0.206
Norway	0.826	0.989	0.974	0.053	0.013
Turkey	0.843	0.989	0.973	0.040	0.019
Denmark	0.857	0.951	0.951	0.062	0.188
U3					
Germany	0.728	0.978	0.871	0.105	0.013
Israel	0.721	0.948	0.686	0.118	0.021
Norway	0.774	1.000	1.007	0.000	0.006
Denmark	0.734	0.883	0.883	0.096	0.276

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

The invariance testing results presented in Table 11.61 show that S1SOCIAL is metric invariant for ISCED level 02 and configural invariant for U3.

Table 11.61. Invariance test results for the scale S1SOCIAL

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.984	0.960	0.055	0.019				
Metric	0.971	0.961	0.054	0.101	0.013	-0.001	0.001	-0.082
Scalar	0.888	0.898	0.088	0.153	0.083	0.063	-0.034	-0.052
Invariance level for U3								
Configural	0.916	0.871	0.077	0.039				
Metric	0.896	0.882	0.074	0.060	0.020	-0.011	0.003	-0.021
Scalar	0.798	0.819	0.091	0.081	0.097	0.063	-0.018	-0.021

Note: An adjusted number of degrees of freedom was used to compute the CFI, TLI and RMSEA for U3 (see the subsection *Treating missing items in scales* in this chapter).
Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.62 contains the unstandardised factor loadings for the S1SOCIAL scale, while Table 11.63 provides the standardised factor loadings and unstandardised intercepts for this scale. Most of the factor loadings for all items in all participating countries in Table 11.63 are high. The only exception is item SS1G41H in Israel (U3).

Table 11.62. Unstandardised factor loadings for the scale S1SOCIAL

Items	Factor loadings
SS1G41G	0.385
SS1G41H	0.391
SS1G41I	0.552
SS1G41J	0.608
SS1G41L	0.589

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.63. Unstandardised intercepts and standardised factor loadings for the scale S1SOCIAL

	Standardised factor loadings					Unstandardised Intercepts				
Participating country	SS1G41G	SS1G41H	SS1G41I	SS1G41J	SS1G41L	SS1G41G	SS1G41H	SS1G41I	SS1G41J	SS1G41L
ISCED level 02										
Chile	0.661	0.643	0.700	0.715	0.773	3.713	3.704	3.543	3.321	3.530
Germany	0.588	0.544	0.625	0.764	0.698	3.674	3.658	3.423	3.486	3.418
Iceland	0.628	0.641	0.559	0.845	0.759	3.678	3.732	3.299	3.621	3.491
Israel	0.677	0.659	0.612	0.755	0.751	3.745	3.695	3.177	3.331	3.377
Japan	0.650	0.653	0.732	0.837	0.847	3.509	3.523	3.375	3.379	3.400
Korea	0.669	0.602	0.598	0.686	0.865	3.592	3.412	2.679	2.829	3.320
Norway	0.606	0.615	0.630	0.793	0.743	3.420	3.477	2.945	3.221	3.140
Turkey	0.613	0.621	0.604	0.709	0.848	3.662	3.663	3.275	3.388	3.584
Denmark	0.631	0.688	0.677	0.847	0.722	3.573	3.748	3.328	3.491	3.240
U3*										
Germany	0.645	0.624	0.590	-	0.666	3.680	3.713	3.290	-	3.220
Israel	0.501	0.446	0.711	-	0.701	3.708	3.752	3.185	-	3.038
Norway	0.688	0.557	0.566	-	0.787	3.306	3.436	2.752	-	2.972
Denmark	0.593	0.721	0.601	-	0.601	3.414	3.804	3.020	-	2.659

Note: *The U3 population was given a phantom indicator for the non-administered item SS1G41J. Therefore, factor loadings and intercepts were fixed to 0 (denoted by "-" in the table).
Source: OECD, TALIS Starting Strong 2018 database.

Facilitating parent/guardian engagement (S1PAREN)

Items

The wording and response options for the items forming the scale S1PAREN are presented in Table 11.64. Respondents had to indicate how well the items described the engagement of the parents in their centre.

Table 11.64. Item wording for the scale S1PAREN

S1PAREN: Facilitating parent/guardian engagement	
SS1G43: How well do these statements describe how you engage parents or guardians at this <ECEC centre>?	
Response options: not at all (1), somewhat (2), well (3), very well (4)	
SS1G43A	Parents or guardians can get in touch with <ECEC staff> easily
SS1G43B	Parents or guardians are informed about the development, well-being, and learning of their children on a regular basis
SS1G43C	Parents or guardians are informed about daily activities on a regular basis
SS1G43E	Parents or guardians are encouraged by <ECEC staff> to do play and learning activities with their children at home

Source: OECD, TALIS Starting Strong database.

Scale reliabilities and model fits

Table 11.65 provides the reliability coefficients and the model fit indices for the S1PAREN scale. The reliability coefficients for all countries are above 0.70 except for Germany for both populations. In addition, at least two of the model fit indices show a good model fit in most of the participating countries. The only exception is Denmark for both populations.

Table 11.65. Reliability coefficients and confirmatory factor analysis of model–data fit for the scale S1PAREN

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.785	0.960	0.879	0.093	0.033
Germany	0.679	0.974	0.922	0.066	0.023
Iceland	0.753	0.982	0.947	0.077	0.022
Israel	0.766	0.965	0.895	0.085	0.027
Japan	0.852	0.979	0.937	0.086	0.028
Korea	0.895	0.998	0.993	0.033	0.009
Norway	0.740	0.995	0.986	0.033	0.013
Turkey	0.867	0.937	0.812	0.104	0.037
Denmark	0.834	0.474	0.474	0.177	0.870
U3					
Germany	0.672	0.987	0.960	0.044	0.017
Israel	0.817	0.970	0.911	0.074	0.023
Norway	0.819	0.997	0.992	0.030	0.015
Denmark	0.874	0.770	0.770	0.119	1.194

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.66 presents the invariance testing results for S1PAREN. Here it can be seen that the scale is metric invariant in both populations.

Table 11.66. Invariance test results for the scale S1PAREN

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.969	0.908	0.087	0.027				
Metric	0.959	0.947	0.067	0.079	0.010	-0.039	0.020	-0.052
Scalar	0.571	0.645	0.172	0.249	0.388	0.302	-0.105	-0.170
Invariance level for U3								
Configural	0.986	0.958	0.054	0.019				
Metric	0.967	0.951	0.058	0.082	0.019	0.007	-0.004	-0.063
Scalar	0.863	0.863	0.097	0.103	0.104	0.088	-0.039	-0.021

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

The unstandardised factor loadings for the scale S1PAREN are presented in Table 11.67.

Table 11.67. Unstandardised factor loadings for the scale S1PAREN

Items	Factor loadings
SS1G43E	0.440
SS1G43A	0.385
SS1G43B	0.569
SS1G43C	0.535

Source: OECD, TALIS Starting Strong 2018 database.

The standardised factor loadings and unstandardised intercepts for the scale S1PAREN presented in Table 11.68 show high factor loadings for all the items in all participating countries. The only exception is item SS1G43E, the values for which are below 0.45 in some countries, namely Germany, Iceland, Norway for ISCED level 02, and Germany and Israel for U3.

Table 11.68. Unstandardised intercepts and standardised factor loadings for the scale S1PAREN

	Standardised factor loadings				Unstandardised intercepts			
Participating country	SS1G43A	SS1G43B	SS1G43C	SS1G43E	SS1G43A	SS1G43B	SS1G43C	SS1G43E
ISCED level 02								
Chile	0.532	0.824	0.655	0.534	3.663	3.628	3.454	3.453
Germany	0.567	0.712	0.533	0.420	3.654	3.465	3.201	2.732
Iceland	0.644	0.761	0.657	0.421	3.743	3.366	3.332	2.586
Israel	0.559	0.802	0.637	0.489	3.546	3.448	3.143	3.143
Japan	0.588	0.868	0.790	0.571	2.981	3.166	3.266	2.636
Korea	0.733	0.903	0.823	0.688	3.367	3.173	3.192	2.994
Norway	0.594	0.747	0.674	0.404	3.687	3.369	3.396	2.384
Turkey	0.659	0.891	0.709	0.682	3.650	3.569	3.367	3.562
Denmark	0.768	0.790	0.772	0.552	3.842	3.469	3.617	2.969
U3								
Germany	0.551	0.709	0.538	0.403	3.654	3.465	3.201	2.732
Israel	0.528	0.862	0.697	0.440	3.620	3.589	3.434	2.953
Norway	0.639	0.839	0.737	0.455	3.687	3.369	3.396	2.384
Denmark	0.813	0.835	0.835	0.524	3.884	3.625	3.758	2.871

Source: OECD, TALIS Starting Strong 2018 database.

Activities to enhance the development of children's abilities and skills (diverse kinds of activities) (S1ACTIV)

Items

The wording and the response options for the items in the scale S1ACTIV are presented in Table 11.69. These items were a combination of items from three questions that asked respondents to indicate the extent to which the stated activities/behaviour applied to the staff in their centre.

Table 11.69. Item wording for the scale S1ACTIV

S1ACTIV: Activities to enhance the development of children's abilities and skills (diverse kinds of activities)	
SS1G31: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G31F	<ECEC staff> position themselves at the children's height when talking or listening
SS1G32: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G32C	<ECEC staff> play word games with the children
SS1G32D	<ECEC staff> play with letters with the children
SS1G33: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G33B	<ECEC staff> play number games (e.g. dice)
SS1G33C	<ECEC staff> sing songs about numbers
SS1G33D	<ECEC staff> help children to use numbers or to count
SS1G33E	<ECEC staff> refer to groups of objects by the size of the group (i.e. number of objects in the group)

Note: This scale was modelled with several items that also modelled the scale S1PROLD. Therefore, analysis should avoid using these two scales together due to possible collinearity of the scales.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

The model was improved through the addition of a correlation between items SS1G32C and SS1G32D for all countries.

Scale reliabilities and model fits

Table 11.70 presents the reliability coefficients and the model fit indices for the S1ACTIV scale. The coefficients are above the cut-off criterion of 0.70 in all countries in both populations, and the model fit indices indicate a good model fit in both populations in all participating countries except the U3 populations in Germany and Norway.

Table 11.70. Reliability coefficients and confirmatory factor analysis of model–data fit for the scale S1ACTIV

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.863	0.972	0.956	0.049	0.032
Germany	0.771	0.934	0.894	0.068	0.039
Iceland	0.828	0.983	0.973	0.042	0.026
Israel	0.845	0.969	0.950	0.049	0.026
Japan	0.808	0.945	0.911	0.060	0.039
Korea	0.841	0.970	0.952	0.044	0.039
Norway	0.817	0.981	0.969	0.046	0.025
Turkey	0.803	0.940	0.903	0.040	0.034
Denmark	0.799	0.957	0.955	0.047	0.130
U3					
Germany	0.792	0.872	0.794	0.103	0.053
Israel	0.808	0.915	0.862	0.079	0.052
Norway	0.806	0.895	0.830	0.106	0.051
Denmark	0.766	0.825	0.816	0.101	0.220

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.71 presents the invariance testing results for the scale *S1ACTIV*. Those results show the scale is metric invariant at ISCED level 02 and configural invariant for the U3 population.

Table 11.71. Invariance test results for the scale S1ACTIV

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.963	0.940	0.050	0.033				
Metric	0.951	0.943	0.048	0.097	0.012	-0.003	0.002	-0.064
Scalar	0.778	0.801	0.090	0.167	0.173	0.142	-0.042	-0.070
Invariance level for U3								
Configural	0.897	0.833	0.094	0.052				
Metric	0.881	0.854	0.088	0.076	0.016	-0.021	0.006	-0.024
Scalar	0.709	0.709	0.124	0.117	0.172	0.145	-0.036	-0.041

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.72 presents the unstandardised factor loadings for the S1ACTIV scale, while Table 11.73 and Table 11.74 present the standardised factor loadings and unstandardised intercepts for this scale, respectively. Most of the standardised factor loadings for all the items in all participating countries are high. The only exceptions are items SS1G32C (where the majority of the loadings are between 0.45 and 0.60) and SS1G31F (where the factor loadings in all participating countries are below the cut-off value).

Table 11.72. Unstandardised factor loadings for the scale S1ACTIV

Items	Factor loadings
SS1G31F	0.230
SS1G32C	0.387
SS1G32D	0.459
SS1G33B	0.515
SS1G33C	0.515
SS1G33D	0.426
SS1G33E	0.449

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.73. Standardised factor loadings for the scale S1ACTIV

	Standardised factor loadings						
Participating country	SS1G31F	SS1G32C	SS1G32D	SS1G33B	SS1G33C	SS1G33D	SS1G33E
ISCED level 02							
Chile	0.370	0.580	0.675	0.748	0.707	0.788	0.754
Germany	0.356	0.499	0.528	0.639	0.569	0.668	0.630
Iceland	0.366	0.583	0.672	0.682	0.673	0.762	0.642
Israel	0.364	0.594	0.584	0.698	0.613	0.811	0.725
Japan	0.387	0.607	0.591	0.674	0.654	0.695	0.641
Korea	0.374	0.515	0.634	0.747	0.709	0.755	0.676
Norway	0.374	0.561	0.620	0.678	0.695	0.712	0.655
Turkey	0.214	0.385	0.303	0.535	0.771	0.597	0.768
Denmark	0.425	0.548	0.569	0.686	0.654	0.666	0.632
U3							
Germany	0.381	0.505	0.587	0.610	0.614	0.661	0.674
Israel	0.405	0.551	0.552	0.604	0.703	0.686	0.693
Norway	0.401	0.579	0.621	0.602	0.704	0.683	0.647
Denmark	0.423	0.507	0.520	0.616	0.650	0.655	0.555

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.74. Unstandardised intercepts for the scale S1ACTIV

	Unstandardised intercepts						
Participating country	SS1G31F	SS1G32C	SS1G32D	SS1G33B	SS1G33C	SS1G33D	SS1G33E
ISCED level 02							
Chile	3.614	3.581	3.590	3.523	3.517	3.770	3.674
Germany	3.503	3.139	2.531	3.178	2.569	3.434	3.343
Iceland	3.540	3.472	3.468	3.226	3.299	3.692	3.277
Israel	3.582	3.572	3.492	3.557	3.260	3.749	3.636
Japan	3.434	3.409	3.047	2.860	2.755	3.339	2.950
Korea	3.516	3.257	3.290	3.423	3.328	3.639	3.386
Norway	3.341	2.769	2.699	2.891	2.600	3.162	2.818
Turkey	3.750	3.736	3.454	3.786	3.899	3.864	3.920
Denmark	3.722	3.326	3.019	2.785	2.925	3.474	3.178
U3							
Germany	3.555	3.132	2.295	2.747	2.609	3.230	3.300
Israel	4.644	3.391	2.146	2.058	3.139	4.042	4.019
Norway	5.334	3.753	3.115	2.471	3.357	4.617	3.661
Denmark	6.877	4.228	2.955	2.872	3.826	5.066	3.570

Source: OECD, TALIS Starting Strong 2018 database.

Excluded scales

Staff beliefs of self-efficacy (S1SELF)

This scale exhibited poor model fits in most of the participating countries for both populations (see Table 11.75). Added model improvements were not successful and attempts to find a meaningful multifactor solution failed. The scale was consequently excluded from further analysis.

Table 11.75. Item wording for the scale S1SELF

S1SELF: Staff beliefs of self-efficacy	
SS1G24: In your work with children, to what extent do you feel that you can do the following?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G24B	Adapt your work to individual child needs
SS1G24C	Help children develop their capacity to learn independently
SS1G24D*	Help children prepare for starting <ISCED 2011 level 1> school
SS1G24E	Stimulate children's interest in cultural differences and commonalities
SS1G24F	Help children to interact with each other and show good social behaviour (e.g. sharing, helping others)
SS1G24G	Calm a child who is upset
SS1G24H	Monitor and observe children's development
SS1G24I	Help children to develop self-confidence
SS1G24J	Help children develop creativity and problem solving
SS1G24K	Use digital technology (e.g. computers, tablets, smart boards) to support children's learning
SS1G24L	Provide all children with a feeling of security

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Satisfaction with profession (S1PROSA)

This scale, the items of which are depicted in Table 11.76, was excluded due to non-scalar invariance within the countries that participated with both populations. The scale could not be improved because it consisted of only two items for the U3 population.

Table 11.76. Item wording for the scale S1PROSA

S1PROSA: Satisfaction with profession	
SS1G44: We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SS1G44A	I think that <ECEC staff> are valued in society
SS1G44I*	The children value me as an <ECEC staff>
SS1G44J	Parents or guardians value me as an <ECEC staff>

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Behavioural management; facilitating play and child-initiated activities; language stimulation and support for literacy learning (S1PRBEH)

This scale, depicted in Table 11.77, was excluded due to non-scalar invariance within the countries that participated with both populations. The scale could not be improved because it consisted of only two items for the U3 population.

Table 11.77. Item wording for the scale S1PRBEH

S1PRBEH: Behavioural management; facilitating play and child-initiated activities; language stimulation and support for literacy learning staff	
SS1G29: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G29A	If invited, <ECEC staff> join in with the children´s play
SS1G29B	When <ECEC staff> play with children, the children are allowed to take the lead
SS1G31: To what extent do the following apply to <ECEC staff> at this <ECEC centre>?	
Response options: not at all (1), very little (2), to some extent (3), a lot (4)	
SS1G31B*	ECEC staff> encourage children to lead conversations

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Situational judgement; facilitating pro-social behaviour (S1SJSOC)

This scale (see Table 11.78) was evaluated within the CFA and IRT framework. The intended unidimensional structure could not be confirmed, an outcome that was also supported by the low reliability coefficients for most of the participating countries. The poor model results in both populations led to the scale being excluded.

Table 11.78. Item wording for the scale S1SJSOC

S1SJSOC: Situational judgement, facilitating pro-social behaviour	
SS1G25: Suppose that you notice that two three-year-old children are independently playing with building blocks. Child A has taken almost all the building blocks and is building things. Child B is shy, looks a bit sad and is struggling with his/her construction. What would you do?	
Response options: I would definitely do this (1), I would probably do this (2), I would probably not do this (3), I would definitely not do this (4)	
SS1G25A	I would divide the building blocks in two equal piles, so that both children have an equal number of building blocks
SS1G25B	I would help child B in building a construction
SS1G25C	I would encourage them to build something together
SS1G25D	I would talk to child A to try to make him/her aware of child B's feelings
SS1G25E	I would encourage child A to share with child B

Source: OECD, TALIS Starting Strong 2018 database.

Situational judgement, facilitating play and child-initiated activities (S1SJINT)

This scale, depicted in Table 11.79, was also evaluated within the CFA and IRT framework. The lack of confirmation for the scale's intended unidimensional structure was supported by the low reliability coefficients for most of the participating countries. The scale was subsequently excluded because of the poor model results for both populations.

Table 11.79. Item wording for the scale S1SJINT

S1SJINT: Situational judgement, facilitating play and child-initiated activities	
SS1G26: Suppose that five three-year-old children are playing with different toys of their choosing. In an ideal situation where you could choose what to do during this time, what would you do? For each suggestion, mark the option that best describes what you would do.	
Response options: I would definitely do this (1), I would probably do this (2), I would probably not do this (3), I would definitely not do this (4)	
SS1G26A	I would play with the children by following their lead
SS1G26B	I would let children play by themselves and only intervene when they request it
SS1G26C	I would contribute to children's play by asking questions or providing explanations
SS1G26D	I would encourage children to play together rather than joining in their play
SS1G26E	I would contribute to children's play by providing new ideas or materials

Source: OECD, TALIS Starting Strong 2018 database.

11.4.2. Complex scale indices from the leader questionnaire

Needs for professional development for instructional leadership (S1LDEV)

Items

The items included in the *Needs for professional development for instructional leadership* (S1LDEV) scale and presented in Table 11.80 represent different areas of professional development. Respondents were asked to assess the extent to which they thought they needed professional development in each of the designated areas.

Table 11.80. Item wording for the scale S1LDEV

S1LDEV: Needs for professional development for instructional leadership	
SL1G09: For each of the areas listed below, please indicate the extent to which you currently need professional development.	
Response options: no need at present (1), low level of need (2), moderate level of need (3), high level of need (4)	
SL1G09C	Using data for improving the quality of the <ECEC centre> (e.g. feedback from <ECEC staff> and parents or guardians)
SL1G09D	Designing the <ECEC centre's> goals for children's development, well-being and learning
SL1G09G	Observing <ECEC staff's> practices and staff–child interactions
SL1G09H	Providing effective feedback

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between SL1G09G and SL1G09H was added for all countries.

Scale reliabilities and model fits

Table 11.81 presents the reliability coefficients and the model fit indices for the scale S1LDEV. The table shows that the coefficients are above the cut-off criterion of 0.70 for both populations in all countries, except Germany. The model fit indices indicate a good model fit in all participating countries but Denmark for ISCED level 02.

Table 11.81. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LDEV

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.874	0.994	0.962	0.098	0.010
Germany	0.669	0.996	0.976	0.050	0.013
Iceland	0.714	0.994	0.963	0.057	0.026
Israel	0.810	1.000	1.004	0.000	0.006
Japan	0.746	1.000	1.021	0.000	0.007
Korea	0.904	1.000	1.025	0.000	0.004
Norway	0.766	0.986	0.914	0.107	0.019
Turkey	0.869	1.000	1.021	0.000	0.002
Denmark	0.803	0.885	0.862	0.121	0.213
U3					
Germany	0.676	1.000	1.023	0.000	0.008
Israel	0.848	0.992	0.953	0.066	0.018
Norway	0.794	1.000	1.053	0.000	0.002
Denmark	0.783	0.992	0.991	0.031	0.160

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.82 presents the invariance testing results for the scale S1LDEV, which proved to be metric invariant for ISCED level 02 and configural invariant for U3.

Table 11.82. Invariance test results for the scale S1LDEV

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.999	0.996	0.024	0.012				
Metric	0.997	0.994	0.028	0.056	0.002	0.002	-0.004	-0.044
Scalar	0.930	0.933	0.099	0.098	0.067	0.061	-0.071	-0.042
Invariance level for U3								
Configural	0.998	0.985	0.038	0.012				
Metric	0.970	0.939	0.077	0.099	0.028	0.046	-0.039	-0.087
Scalar	0.938	0.926	0.085	0.131	0.032	0.013	-0.008	-0.032

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Table 11.83 presents the unstandardised factor loadings for S1LDEV.

Table 11.83. Unstandardised factor loadings for the scale S1LDEV

Items	Factor loadings
SL1G09C	0.663
SL1G09D	0.722
SL1G09G	0.621
SL1G09H	0.612

Source: OECD, TALIS Starting Strong 2018 database.

The standardised factor loadings and unstandardised intercepts for the scale S1LDEV are presented in Table 11.84. All of the factor loadings for all the items in all participating countries are higher than 0.60 in both populations, indicating a strong relationship between the items and the latent factor.

Table 11.84. Unstandardised intercepts and standardised factor loadings for the scale S1LDEV

	Standardised factor loadings				Unstandardised intercepts			
Participating country	SL1G09C	SL1G09D	SL1G09G	SL1G09H	SL1G09C	SL1G09D	SL1G09G	SL1G09H
ISCED level 02								
Chile	0.834	0.856	0.760	0.751	2.810	2.599	2.636	2.804
Germany	0.570	0.671	0.573	0.547	2.705	2.882	2.743	2.741
Iceland	0.620	0.707	0.575	0.612	2.710	2.495	2.711	2.919
Israel	0.777	0.763	0.692	0.707	2.482	2.397	2.312	2.290
Japan	0.709	0.694	0.655	0.628	3.426	3.456	3.530	3.359
Korea	0.892	0.865	0.808	0.775	3.208	3.070	3.231	3.258
Norway	0.720	0.738	0.633	0.628	2.778	2.647	2.766	2.593
Turkey	0.799	0.887	0.699	0.674	2.365	2.297	2.059	2.079
Denmark	0.696	0.812	0.648	0.681	2.264	1.996	2.029	2.304
U3								
Germany	0.612	0.649	0.561	0.579	2.705	2.882	2.743	2.741
Israel	0.770	0.845	0.706	0.744	2.482	2.397	2.312	2.290
Norway	0.731	0.808	0.630	0.629	2.778	2.647	2.766	2.593
Denmark	0.649	0.776	0.708	0.717	2.296	2.256	2.141	2.328

Source: OECD, TALIS Starting Strong 2018 database.

Environment of the neighbourhood (S1LNIGHB)

Items

The items from one question formed the scale *Environment of the neighbourhood* (S1LNIGHB). When answering this question, centre leaders needed to state their (dis)agreement with items describing their centre's neighbourhood. The wording of the items and their response options are provided in Table 11.85.

Table 11.85. Item wording for the scale S1LNIGHB

S1LNIGHB: Environment of the neighbourhood	
SL1G12: Thinking about the neighbourhood where this <ECEC centre> is located, how strongly do you agree or disagree with the following statements?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SL1G12A	This is a good neighbourhood to bring up children
SL1G12B*	There is <litter> lying around
SL1G12C*	There is vandalism and deliberate damage to property
SL1G12D*	People experience insults or attacks to do with ethnic or cultural background
SL1G12E	There are public places where children can play safely
SL1G12F*	There is drug-related crime

Note: *This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SL1G12A and SL1G12E was included for all countries. A correlation between SL1G12B and SL1G12C was also added for the ISCED level 02 populations in Iceland, Israel and Norway as well as the U3 population in Israel. Finally, a correlation between SL1G12C and SL1G12D was added for Turkey (ISCED level 2).

Scale reliabilities and model fits

Table 11.86 presents the reliabilities and the model fits for S1LNIGHB. Except for Norway ISCED level 2, the reliability coefficients are high in all participating countries for both populations. The model fits show a good fit for all countries except Korea at ISCED level 02.

Table 11.86. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LNIGHB

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.882	0.997	0.994	0.027	0.021
Germany	0.880	0.979	0.961	0.069	0.024
Iceland	0.731	1.000	1.026	0.000	0.021
Israel	0.801	0.929	0.847	0.107	0.052
Japan	0.874	1.000	1.019	0.000	0.027
Korea	0.857	0.864	0.746	0.120	0.062
Norway	0.627	1.000	1.085	0.000	0.045
Turkey	0.783	0.947	0.887	0.064	0.048
Denmark	0.850	1.000	1.024	0.000	0.030
U3					
Germany	0.897	0.987	0.976	0.057	0.019
Israel	0.821	1.000	1.011	0.000	0.032
Norway	0.839	0.960	0.925	0.071	0.041
Denmark	0.867	0.987	0.976	0.051	0.037

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

The invariance testing results for S1NIGHB presented in Table 11.87 show that the scale reached configural invariance for ISCED level 02 and metric invariance for U3.

Table 11.87. Invariance test results for the scale S1NIGHB

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.974	0.948	0.058	0.039				
Metric	0.937	0.919	0.072	0.125	0.037	0.029	-0.014	-0.086
Scalar	0.767	0.782	0.118	0.175	0.170	0.137	-0.046	-0.050
Invariance level for U3								
Configural	0.986	0.973	0.046	0.030				
Metric	0.986	0.982	0.038	0.090	0.000	-0.009	0.008	-0.060
Scalar	0.811	0.802	0.124	0.117	0.175	0.180	-0.086	-0.027

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

As the S1NIGHB scale reached only the configural invariance level, there are no unstandardised item parameters. The standardised factor loadings and unstandardised intercepts for the scale are presented in Table 11.88 and

Source: OECD, TALIS Starting Strong 2018 database

Table 11.89, respectively. The standardised factor loadings are acceptable, except in the ISCED level 02 populations of Iceland, Japan, Korea, Norway and Denmark for item SL1G12A, and all populations for item SL1G12E.

Table 11.88. Standardised factor loadings for the scale S1NIGHB

	Standardised factor loadings					
Participating country	SL1G12A	SL1G12B	SL1G12C	SL1G12D	SL1G12E	SL1G12F
ISCED level 02						
Chile	0.611	0.772	0.836	0.786	0.327	0.767
Germany	0.618	0.761	0.855	0.744	0.283	0.757
Iceland	0.338	0.512	0.500	0.762	0.130	0.647
Israel	0.568	0.614	0.712	0.712	0.194	0.743
Japan	0.092	0.660	0.862	0.821	0.182	0.730
Korea	0.356	0.660	0.882	0.762	0.332	0.372
Norway	0.427	0.554	0.598	0.489	0.176	0.514
Turkey	0.651	0.778	0.713	0.529	0.206	0.470
Denmark	0.349	0.728	0.786	0.711	0.269	0.797
U3						
Germany	0.652	0.806	0.858	0.793	0.260	0.781
Israel	0.532	0.608	0.757	0.725	0.161	0.783
Norway	0.584	0.751	0.817	0.615	0.203	0.653
Denmark	0.689	0.718	0.777	0.829	0.372	0.662

Source: OECD, TALIS Starting Strong 2018 database

Table 11.89. Unstandardised intercepts for the scale S1NIGHB

	Unstandardised intercepts					
Participating country	SL1G12A	SL1G12B	SL1G12C	SL1G12D	SL1G12E	SL1G12F
ISCED level 02						
Chile	2.935	2.792	2.805	3.170	2.640	2.724
Germany	3.548	3.311	3.165	3.375	3.338	3.231
Iceland	3.659	3.149	2.942	3.646	3.370	3.463
Israel	3.389	3.095	3.272	3.533	3.060	3.476
Japan	3.204	3.517	3.508	3.536	3.034	3.568
Korea	2.979	2.967	3.347	3.392	2.917	3.831
Norway	3.705	3.517	3.451	3.390	3.379	3.271
Turkey	2.962	3.156	3.295	3.452	2.989	3.371
Denmark	3.362	3.321	3.136	3.417	3.252	3.157
U3						
Germany	3.548	3.311	3.165	3.375	3.338	3.231
Israel	3.389	3.095	3.272	3.533	3.060	3.476
Norway	3.705	3.517	3.451	3.390	3.379	3.271
Denmark	3.483	3.484	3.099	3.429	3.419	3.198

Source: OECD, TALIS Starting Strong 2018 database.

Leader support for pedagogical learning (S1LEADS)

Items

The items forming the scale *Leader support for pedagogical learning* (S1LEADS) represent different areas of leadership within an ECEC facility. Table 11.90 provides the wording of the items and their response options.

Table 11.90. Item wording for the scale S1LEADS

S1LEADS: Leader support for pedagogical learning	
SL1G23: Please indicate how frequently you engaged in the following activities in this <ECEC centre> during the last 12 months.	
Response options: never (1), less than monthly (2), monthly (3), weekly (4), daily (5)	
SL1G23A	I collaborated with <ECEC staff> to improve how children play together
SL1G23D	I took actions to support co-operation among <ECEC staff> to develop new approaches to ECEC practices
SL1G23E	I took actions to ensure that <ECEC staff> take responsibility for improving their skills in working with children
SL1G23F	I took actions to ensure that <ECEC staff> feel responsible for children's development, well-being and learning
SL1G23G	I worked on developing a <vision> for this <ECEC centre>
SL1G23I	I facilitated the use of the indoor or outdoor space for learning

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between SL1G23E and SL1G23F was added for all countries. In addition, a correlation between SL1G23A and SL1G23G was included for the ISCED level 02 population of Germany, while a correlation between SL1G23G and SL1G23I was added for the U3 population in Israel.

Scale reliabilities and model fits

The reliability coefficients for the S1LEADS scale presented in Table 11.91 are high in all participating countries for both populations, while the model fits are good in all countries except for the Denmark U3 population.

Table 11.91. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LEADS

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.870	0.947	0.901	0.106	0.047
Germany	0.750	0.932	0.854	0.106	0.042
Iceland	0.706	0.960	0.925	0.080	0.042
Israel	0.790	0.991	0.984	0.039	0.025
Japan	0.878	0.976	0.955	0.065	0.038
Korea	0.846	1.000	1.021	0.000	0.027
Norway	0.706	0.951	0.909	0.065	0.050
Turkey	0.897	0.951	0.907	0.085	0.039
Denmark	0.876	0.915	0.909	0.122	0.129
U3					
Germany	0.733	0.983	0.969	0.046	0.034
Israel	0.806	0.945	0.883	0.068	0.042
Norway	0.719	0.975	0.953	0.043	0.038
Denmark	0.865	0.854	0.844	0.113	0.362

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.92 presents the invariance testing results for S1LEADS, which reached metric invariance for both populations.

Table 11.92. Invariance test results for the scale S1LEADS

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.968	0.940	0.070	0.038				
Metric	0.958	0.948	0.065	0.078	0.010	-0.008	0.005	-0.040
Scalar	0.676	0.708	0.155	0.208	0.282	0.240	-0.090	-0.130
Invariance level for U3								
Configural	0.963	0.927	0.058	0.038				
Metric	0.943	0.923	0.060	0.078	0.020	0.004	-0.002	-0.040
Scalar	0.702	0.688	0.121	0.140	0.241	0.235	-0.061	-0.062

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

As the S1LDEV scale reached metric invariance, the unstandardised factor loadings for the scale are presented in Table 11.93. The standardised loadings appear in Table 11.94, while the unstandardised intercepts can be found in Table 11.95.

Although many of the factor loadings in Table 11.94 are high to moderate, low factor loadings are evident for item SL1G32A in the ISCED level 02 populations of Germany, Iceland and Norway, and in the U3 populations of Germany and Norway. In addition, low factor loadings can be observed for item SL1G32G in the ISCED level 02 populations of Iceland and Norway and the U3 population of Norway. Finally, the factor loadings for item SL1G32I are low in the ISCED level 02 populations of Germany, Iceland and Norway, as well as the U3 populations of Germany, Israel and Norway.

Table 11.93. Unstandardised factor loadings for the scale S1LEADS

Items	Factor loadings
SL1G23D	0.748
SL1G23E	0.798
SL1G23F	0.755
SL1G23A	0.593
SL1G23G	0.610
SL1G23I	0.513

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.94. Standardised factor loadings for the scale S1LEADS

	Standardised factor loadings					
Participating country	SL1G23A	SL1G23D	SL1G23E	SL1G23F	SL1G23G	SL1G23I
ISCED level 02						
Chile	0.495	0.828	0.847	0.795	0.582	0.464
Germany	0.402	0.711	0.700	0.592	0.504	0.358
Iceland	0.433	0.670	0.668	0.652	0.430	0.366
Israel	0.599	0.719	0.711	0.730	0.507	0.468
Japan	0.584	0.764	0.866	0.814	0.698	0.536
Korea	0.609	0.733	0.776	0.801	0.687	0.512
Norway	0.445	0.640	0.648	0.662	0.358	0.359
Turkey	0.674	0.785	0.880	0.844	0.617	0.545
Denmark	0.590	0.835	0.871	0.779	0.571	0.496
U3						
Germany	0.370	0.703	0.676	0.601	0.509	0.366
Israel	0.619	0.703	0.765	0.724	0.510	0.430
Norway	0.390	0.675	0.680	0.566	0.386	0.367
Denmark	0.577	0.836	0.854	0.730	0.588	0.523

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.95. Unstandardised intercepts for the scale S1LEADS

	Unstandardised intercepts					
Participating country	SL1G23A	SL1G23D	SL1G23E	SL1G23F	SL1G23G	SL1G23I
ISCED level 02						
Chile	3.706	3.534	3.588	3.701	3.263	3.952
Germany	3.730	2.971	2.929	3.214	2.643	3.378
Iceland	2.859	3.250	3.426	3.558	3.146	1.903
Israel	4.275	3.410	3.714	4.082	3.339	3.835
Japan	3.273	2.972	2.941	3.035	2.460	2.907
Korea	3.569	2.921	3.589	3.568	3.277	3.364
Norway	3.443	3.558	3.350	3.886	2.699	2.923
Turkey	3.193	3.211	3.470	3.588	3.433	3.969
Denmark	3.589	3.441	3.144	3.324	2.780	3.365
U3						
Germany	3.730	2.971	2.929	3.214	2.643	3.378
Israel	3.592	2.861	3.379	3.756	2.701	2.882
Norway	3.443	3.558	3.350	3.886	2.699	2.923
Denmark	3.373	3.331	3.102	3.141	2.754	3.149

Source: OECD, TALIS Starting Strong 2018 database.

Approaches to multicultural and gender diversity (S1LDIVB)

Items

The items for the scale *Approaches to multicultural and gender diversity* (S1LDIVB) provide statements about cultural and gender diversity in ECEC, as depicted in Table 11.96.

Table 11.96. Item wording for the scale S1LDIVB

S1LDIVB: Approaches to multicultural and gender diversity	
SL1G29: In your view, approximately how many <ECEC staff> in this <ECEC centre> would agree with the following statements?	
Response options: none or almost none (1), some of them (2), many (3), all or almost all (4)	
SL1G29A	It is important to be responsive to differences in children's cultural backgrounds
SL1G29B	It is important for children to learn that people from other cultures can have different values
SL1G29C	Respecting other cultures is something that children should learn as early as possible
SL1G29D	It is important to recognise the different interests of children regardless of gender

Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SL1G29A and SL1G29B was added for the ISCED level 02 populations of Chile, Denmark, Germany, Israel and Turkey, and the U3 populations of Germany and Denmark.

Scale reliabilities and model fits

Table 11.97 presents the reliabilities and the model fits for the S1LDIVB scale. The table shows high reliability coefficients in all participating countries for both populations. It also shows good model fits for all countries, with the exception of Denmark's U3 population.

Table 11.97. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LDIVB

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.901	1.000	1.017	0.000	0.003
Germany	0.857	1.000	1.012	0.000	0.005
Iceland	0.929	1.000	1.032	0.000	0.009
Israel [1]	-	-	-	-	-
Japan	0.910	0.991	0.972	0.066	0.017
Korea	0.939	0.991	0.972	0.067	0.018
Norway	0.927	1.000	1.018	0.000	0.023
Turkey	0.891	0.972	0.834	0.076	0.013
Denmark	0.865	0.984	0.906	0.115	0.019
U3					
Germany	0.837	1.000	1.001	0.000	0.007
Israel	0.925	0.996	0.989	0.030	0.027
Norway	0.814	1.000	1.057	0.000	0.008
Denmark [2, 3]	-	0.671	-	0.427	0.054

1. Israel ISCED level 02 was excluded from the scale because of complications with model estimation.
2. This participating country's reliability coefficient estimation failed in the final scale model due to a negative residual variance for one or more items that could not be corrected and therefore has untrustworthy scale scores for the corresponding population.

3. The poor fit of the model affected the TLI calculation, which is not reported.
Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.98 presents the invariance testing results for the scale S1LDIVB. It shows the scale reached configural invariance for ISCED level 02 and scalar invariance for U3.

Table 11.98. Invariance test results for the scale S1LDIVB

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.996	0.986	0.037	0.013				
Metric	0.967	0.952	0.068	0.239	0.029	0.034	-0.031	-0.226
Scalar	0.911	0.920	0.088	0.258	0.056	0.032	-0.020	-0.019
Invariance level for U3								
Configural	1.000	1.004	0.000	0.017				
Metric	1.000	1.009	0.000	0.098	0.000	-0.005	0.000	-0.081
Scalar	0.959	0.957	0.068	0.122	0.041	0.052	-0.068	-0.024

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Unstandardised item parameters are not reported because the scale S1LDIVB reached only the configural invariance level for the ISCED level 02 population. The standardised factor loadings are reported in Table 11.99 and the unstandardised intercepts in Table 11.100. Most of the factor loadings in Table 11.99 are above 0.60, while those below this cut-off criterion remain moderate, with values above 0.45.

Table 11.99. Unstandardised intercepts and standardised factor loadings for the scale S1LDIVB

	Standardised factor loadings				Unstandardised intercepts			
Participating country	SL1G29A	SL1G29B	SL1G29C	SL1G29D	SL1G29A	SL1G29B	SL1G29C	SL1G29D
ISCED level 02								
Chile	0.731	0.836	0.887	0.850	3.643	3.647	3.733	3.721
Germany	0.706	0.826	0.861	0.641	3.598	3.531	3.585	3.737
Iceland	0.778	0.942	0.869	0.604	3.758	3.718	3.700	3.788
Israel *	-	-	-	-	-	-	-	-
Japan	0.841	0.893	0.820	0.800	2.853	3.034	2.854	3.205
Korea	0.879	0.944	0.814	0.792	3.232	3.332	3.239	3.499
Norway	0.868	0.937	0.797	0.637	3.598	3.531	3.585	3.737
Turkey	0.625	0.775	0.865	0.887	3.655	3.652	3.625	3.719
Denmark	0.766	0.754	0.876	0.729	3.826	3.724	3.785	3.902
U3								
Germany	0.675	0.763	0.857	0.691	3.598	3.531	3.585	3.737
Israel	0.837	0.941	0.785	0.655	3.598	3.531	3.585	3.737
Norway	0.775	0.751	0.751	0.459	3.598	3.531	3.585	3.737
Denmark	0.911	0.910	0.745	0.662	3.672	3.604	3.469	3.809

Note: *Israel ISCED level 02 was excluded from the scale because of complications with model estimation.
Source: OECD, TALIS Starting Strong 2018 database.

Parent/guardian engagement and relationship with the community (S1LPART)

Items

The items for the scale *Parent/guardian engagement and relationship with the community* (S1LPART) presented in Table 11.100 provide statements about how the ECEC centres that participated in the TALIS Starting Strong reached out to parents.

Table 11.100. Item wording for the scale S1LPART

S1LPART: Parent engagement/guardian and relationship with the community	
SL1G25: During the last 12 months, did this <ECEC centre> either alone or in conjunction with another organisation provide any of the following to parents or guardians?	
Response options: yes (1), no (2)	
SL1G25A+	Workshops or courses for parents or guardians regarding child rearing or child development
SL1G25B+	Support for parents' or guardians' involvement with the operation of the <ECEC centre> (e.g. fund raising, cleaning of the <ECEC centre>)
SL1G25C*+	Meetings to allow parents or guardians to contribute to <ECEC centre> management decisions
SL1G25D+	Setting up events for families and prospective parents or guardians to visit the <ECEC centre>
SL1G33: How strongly do you agree or disagree with these statements as applied to this <ECEC centre>?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SL1G33B	This <ECEC centre> provides parents or guardians with opportunities to actively participate in <ECEC centre> decisions
SL1G33C	This <ECEC centre> provides children with opportunities to actively participate in decisions

Notes:
*This item was removed and is not included in any of the results presented for this scale.
+This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SL1G33B and SL1G33C was added for all countries.

Scale reliabilities and model fits

Table 11.101 includes the reliabilities and model fits for the scale S1LPART. The reliability coefficients are between 0.50 and 0.70 for all countries except Norway, while the model fits exhibit acceptable fit for all countries, with the exception of Japan's ISCED level 02 population.

Table 11.101. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LPART

Participating country	Cronbach's alpha based on standardised items	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.566	0.985	0.963	0.035	0.449
Germany*	-	-	-	-	-
Iceland	0.557	1.000	1.039	0.000	0.348
Israel	0.617	1.000	1.045	0.000	0.222
Japan	0.632	0.930	0.826	0.068	0.585
Korea[1]	-	-	-	-	-
Norway	0.314	1.000	1.078	0.000	0.344
Turkey	0.570	1.000	1.029	0.000	0.368
Denmark*	-	-	-	-	-
U3					
Germany	0.597	0.986	0.966	0.038	0.486
Israel	0.675	1.000	1.105	0.000	0.237
Norway*	-	-	-	-	-
Denmark*	-	-	-	-	-

Note: *These populations were removed from the parameter estimation due to model instability (the parameter estimates could not be generated for them because the models did not converge). Without these populations, the scale exhibited acceptable fit. However, the scale model was not a strong one.
Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Error! Not a valid bookmark self-reference., which presents the invariance testing results for S1LPART, shows the scale reached configural invariance for both populations.

Table 11.102. Invariance test results for the scale S1LPART

	CFI	TLI	RMSEA	WRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ WRMR
Invariance level for ISCED level 02								
Configural	0.988	0.971	0.029	1.413				
Metric	0.890	0.868	0.060	2.862	-0.098	-0.103	0.031	1.449
Scalar	0.768	0.723	0.088	3.324	-0.122	-0.145	0.028	0.462
Invariance level for U3								
Configural	1.000	1.020	0.000	0.698				
Metric	0.834	0.774	0.070	1.713	-0.166	-0.246	0.070	1.015
Scalar	0.490	0.305	0.123	2.351	-0.344	-0.469	0.053	0.638

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

As the scale S1LPART reached configural invariance, there are no unstandardised item parameters to report. The standardised factor loadings can be found in Table 11.103 and the unstandardised intercepts in Table 11.104. Many of the factor loadings in Table 11.103 are only moderate, while others fall below 0.45, including the loading for item SL1G25B for the U3 population of Israel, item SL1G33B for the ISCED level 02 populations of Chile, Iceland, Japan, Norway and Turkey, and item SL1G33C for the ISCED level 02

populations of Iceland, Israel, Japan, Norway and Turkey, and the U3 populations of Germany and Israel.

Table 11.103. Standardised factor loadings for the scale S1LPART

	Standardised factor loadings				
Participating country	SL1G25A	SL1G25B	SL1G25D	SL1G33B	SL1G33C
ISCED level 02					
Chile	0.468	0.468	0.555	0.315	0.514
Germany*	-	-	-	-	-
Iceland	0.667	0.564	0.635	0.248	0.228
Israel	0.554	0.483	0.464	0.534	0.379
Japan	0.613	0.669	0.797	0.228	0.107
Korea[1]	-	-	-	-	-
Norway	0.694	0.536	-0.635	0.286	-0.324
Turkey	0.600	0.868	0.503	0.224	0.117
Denmark*	-	-	-	-	-
U3					
Germany	0.558	0.540	0.793	0.480	0.173
Israel	0.572	0.242	0.667	0.747	0.378
Norway*	-	-	-	-	-
Denmark*	-	-	-	-	-

Note: *These populations were removed from the parameter estimations due to model instability.
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.104. Unstandardised thresholds and intercepts for the scale S1LPART

	Unstandardised thresholds and intercepts				
Participating country	SL1G25A	SL1G25B	SL1G25D	SL1G33B	SL1G33C
ISCED level 02					
Chile	-1.124	-0.604	-0.649	3.118	3.196
Germany*	-	-	-	-	-
Iceland	0.024	0.724	-1.020	2.888	2.972
Israel	-0.117	-0.034	0.125	2.480	2.875
Japan	-0.383	-0.587	-0.927	2.764	2.719
Korea[1]	-	-	-	-	-
Norway	-0.545	0.506	-1.798	3.184	3.202
Turkey	-0.303	-0.466	-1.135	3.185	3.105
Denmark*	-	-	-	-	-
U3					
Germany	-0.009	-0.083	-0.984	2.912	3.291
Israel	0.131	2.566	-0.218	2.349	2.186
Norway*	-	-	-	-	-
Denmark*	-	-	-	-	-

Note: *These populations were removed from the parameter estimation due to model instability.
Source: OECD, TALIS Starting Strong 2018 database.

Networking with parents, other ECEC leaders, ISCED level 1 teachers, and related institutions (S1LPANE)

Items

The items for the scale *Networking with parents, other ECEC leaders, ISCED level 1 teachers, and related institutions* (S1LPANE) were used to gauge communication and co-operation with other partners both within and outside the ECEC settings. The wording of these items and their response options are set out in Table 11.105.

Table 11.105. Item wording for the scale S1LPANE

S1LPANE: Networking with parents, other ECEC leaders, ISCED 1 teachers, and related institutions	
SL1G26: How often do the following activities take place in this <ECEC centre>?	
Response options: never (1), less than monthly (2), monthly (3), weekly (4), daily (5)	
SL1G26A	Informal communication with parents or guardians (e.g. informal conversations on children's development or activities)
SL1G26B	Formal communication with parents or guardians (e.g. attending parent-staff meetings)
SL1G26C	Communication with staff and/or leaders from other <ECEC centres>
SL1G26D*	Communication with <ISCED 2011 level 1> school teachers
SL1G26E	Cooperation with child, family or social services (e.g. child protection agency, family support services)
SL1G26F	Collaboration with health-related services (e.g. child health services providing screening and support regarding child or family health)
SL1G26G	Consultation with child development specialists

Note: *This item was not included in the U3 version of the questionnaire.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SL1G26D and SL1G26F was included for all countries. Several specific improvements were also added. These included the addition of correlations between the following items: SL1G26B and SL1G26C for the U3 population of Germany; SL1G26B and SL1G26G for the ISCED level 02 population of Israel; SL1G26A and SL1G26D for the ISCED level 2 population of Korea; and SL1G26D and SL1G26G for Turkey.

Item SL1G26D was not administered to the U3 populations in all of the participating countries, which meant the correlation between SL1G26D and SL1G26F was not added to the country-level models for those populations. However, additional correlations were added as follows: SL1G26C and SL1G26G in Denmark and SL1G26A and SL1G26C in Israel.

Scale reliabilities and model fits

The reliabilities and model fits for the scale S1LPANE are included in Table 11.106. Here it can be seen that the reliability coefficients are high, with the exceptions of Norway's ISCED level 02 population and Israel's U3 population. The model fits exhibit some low fit statistics for several countries, including the ISCED level 02 populations of Iceland, Israel, Korea, Norway, Turkey and Denmark, and all U3 populations.

Table 11.106. Reliability coefficients and confirmatory factor analysis of model–data fit for the scale S1LPANE

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.762	0.918	0.877	0.065	0.049
Germany	0.623	0.906	0.859	0.044	0.049
Iceland	0.616	0.829	0.743	0.080	0.058
Israel	0.697	0.885	0.827	0.069	0.045
Japan	0.618	0.924	0.886	0.042	0.045
Korea	0.792	0.890	0.834	0.081	0.067
Norway	0.446	0.707	0.560	0.081	0.062
Turkey	0.771	0.754	0.631	0.092	0.068
Denmark	0.794	0.820	0.811	0.106	0.158
U3					
Germany	0.745	0.891	0.819	0.088	0.048
Israel	0.462	0.701	0.654	0.112	0.139
Norway	0.792	0.437	0.437	0.085	0.299
Denmark	0.769	0.811	0.797	0.060	0.180

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.107 presents the invariance testing results for the scale S1LPANE, which reached metric invariance for the ISCED level 02 populations.

Table 11.107. Invariance test results for the scale S1LPANE

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.921	0.869	0.057	0.048				
Metric	0.900	0.883	0.054	0.076	0.021	-0.014	0.003	-0.028
Scalar	0.000	-0.192	0.172	0.376	0.900	1.075	-0.118	-0.300

Note: The Israel and Norway U3 populations were excluded due to model complications, while Denmark was excluded from all models because of not meeting technical standards. Measurement invariance testing was therefore not possible.
Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

As the S1LPANE scale reached metric invariance, the unstandardised factor loadings could be reported (see Table 11.108). Table 11.109 presents the standardised factor loadings and Table 11.110 reports the unstandardised intercepts. Most of the unstandardised factors are low; the only moderate factor loading observed is that for item SL1G26E. The same outcome is evident for many of the standardised factor loadings.

Table 11.108. Unstandardised factor loadings for the scale S1LPANE

Items	Factor loadings
SL1G26A	0.149
SL1G26B	0.193
SL1G26C	0.337
SL1G26D	0.376
SL1G26E	0.500
SL1G26F	0.438
SL1G26G	0.432

Source: OECD, TALIS Starting Strong 2018 database.

Table 11.109. Standardised factor loadings for the scale S1LPANE

	Standardised factor loadings						
Participating country	SL1G26A	SL1G26B	SL1G26C	SL1G26D	SL1G26E	SL1G26F	SL1G26G
ISCED level 02							
Chile	0.191	0.373	0.431	0.370	0.707	0.714	0.600
Germany	0.079	0.155	0.305	0.304	0.484	0.663	0.434
Iceland	0.114	0.335	0.316	0.276	0.551	0.629	0.443
Israel	0.126	0.240	0.280	0.413	0.673	0.678	0.441
Japan	0.099	0.279	0.388	0.345	0.571	0.509	0.500
Korea	0.135	0.252	0.363	0.368	0.736	0.721	0.712
Norway	0.167	0.219	0.200	0.330	0.451	0.425	0.357
Turkey	0.152	0.333	0.364	0.289	0.677	0.766	0.556
Denmark	0.174	0.234	0.362	0.423	0.801	0.723	0.650
U3							
Germany	0.132	0.228	0.432	-	0.715	0.714	0.600
Israel	0.111	0.195	0.260	-	0.407	0.525	0.481
Norway	0.285	0.431	0.430	-	0.748	0.771	0.588
Denmark	0.210	0.234	0.354	-	0.666	0.689	0.702

Note: The U3 population was given phantom indicators for the non-administered items SS1G29D and SS1G29G. Therefore, factor loadings were fixed to 0 (denoted by "-" in the table).
Source: OECD, TALIS Starting Strong 2018 database.

Table 11.110. Unstandardised intercepts for the scale S1LPANE

	Unstandardised intercepts						
Participating country	SL1G26A	SL1G26B	SL1G26C	SL1G26D	SL1G26E	SL1G26F	SL1G26G
ISCED level 02							
Chile	4.312	3.196	2.897	2.566	2.382	2.519	2.167
Germany	4.174	2.425	2.551	2.198	2.286	2.013	2.044
Iceland	4.386	2.112	2.738	2.393	2.274	2.203	2.748
Israel	3.643	2.656	3.396	1.584	1.603	1.381	2.153
Japan	4.438	3.099	2.778	2.048	1.956	1.860	2.176
Korea	3.741	2.460	2.819	1.736	1.822	2.125	1.736
Norway	4.808	2.163	3.198	2.025	2.296	2.269	2.264
Turkey	3.648	2.648	2.761	3.478	1.703	2.212	2.001
Denmark	4.602	2.739	3.122	2.079	2.624	2.349	2.848
U3							
Germany	4.174	2.425	2.551	-	2.286	2.013	2.044
Israel	3.527	2.122	3.391	-	2.400	1.661	2.087
Norway	4.773	2.130	3.029	-	2.302	2.271	2.192
Denmark	4.701	2.599	3.179	-	2.660	2.276	2.793

Note: The U3 population was given phantom indicators for the non-administered item SL1G26D. Therefore, factor loadings were fixed to 0 (denoted by "-" in the table).
Source: OECD, TALIS Starting Strong 2018 database.

Distributed leadership (S1LEADP)

Items

The items for the scale *Distributed leadership* (S1LEADP) include statements about shared leadership in ECEC. Table 11.111 provides the wording of and the response options for these items.

Table 11.111. Item wording for the scale S1LEADP

S1LEADP: Distributed leadership	
SL1G33: How strongly do you agree or disagree with these statements as applied to this <ECEC centre>?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SL1G33A	This <ECEC centre> provides <ECEC staff> with opportunities to actively participate in <ECEC centre> decisions
SL1G33B	This <ECEC centre> provides parents or guardians with opportunities to actively participate in <ECEC centre> decisions
SL1G33C	This <ECEC centre> provides children with opportunities to actively participate in decisions
SL1G33D	This <ECEC centre> has a culture of shared responsibility for <ECEC centre> issues
SL1G33E*	I make the important decisions on my own

Note: *This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Model improvements

A correlation between items SL1G33A and SL1G33C was added for the ISCED level 02 populations of Chile and Korea, the U3 population of Norway and both populations in Denmark. A correlation between items SL1G33B and SL1G33C was added for both the ISCED level 02 and U3 populations in Israel. A correlation between items SL1G33B and SL1G33D for Norway's ISCED level 02 population was also included in the model.

Scale reliabilities and model fits

Table 11.112 presents the reliabilities and model fits for the S1LEADP scale. The reliability coefficients are high, while the model fit statistics are acceptable in all countries for each population with the exception of Denmark's ISCED level 02 population.

Table 11.112. Reliability coefficients and confirmatory factor analysis of the model–data fit for the scale S1LEADP

Participating country	Omega	CFI	TLI	RMSEA	SRMR
ISCED level 02					
Chile	0.943	0.987	0.968	0.058	0.027
Germany	0.686	0.955	0.911	0.071	0.033
Iceland	0.734	0.975	0.938	0.065	0.033
Israel	0.632	0.971	0.928	0.059	0.031
Japan	0.799	1.000	1.056	0.000	0.021
Korea	0.895	0.950	0.874	0.073	0.040
Norway	0.799	0.949	0.874	0.089	0.046
Turkey	0.828	0.987	0.973	0.035	0.028
Denmark	0.627	0.893	0.733	0.076	0.044
U3					
Germany	0.712	0.969	0.939	0.055	0.036
Israel	0.632	0.952	0.880	0.068	0.044
Norway	0.716	0.962	0.905	0.066	0.040
Denmark	0.832	1.000	1.411	0.000	0.032

Source: OECD, TALIS Starting Strong 2018 database.

Invariance testing

Table 11.113 presents the invariance testing results for the scale S1LEADP. The scale is configural invariant for the ISCED level 02 populations and metric invariant for the U3 populations.

Table 11.113. Invariance test results for the scale S1LEADP

	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Invariance level for ISCED level 02								
Configural	0.982	0.958	0.048	0.032				
Metric	0.928	0.909	0.070	0.109	0.054	0.049	-0.022	-0.077
Scalar	0.227	0.320	0.192	0.279	0.701	0.589	-0.122	-0.170
Invariance level for U3								
Configural	0.959	0.906	0.063	0.040				
Metric	0.961	0.945	0.049	0.075	-0.002	-0.039	0.014	-0.035
Scalar	0.583	0.569	0.136	0.193	0.378	0.376	-0.087	-0.118

Source: OECD, TALIS Starting Strong 2018 database.

Item parameters

Standardised item parameters are not presented because the scale S1LEADP reached the configural invariance level for the ISCED level 02 population. Table 11.114 includes the standardised factor loadings and the unstandardised intercepts. The factor loadings in the first of these two tables are mostly moderate to high except for the following items:

SL1G33B for the ISCED level 02 population of Israel and the U3 population of Denmark; SL1G33D for the U3 population of Denmark; and SL1G33E for all countries for both populations.

Table 11.114. Unstandardised intercepts and standardised factor loadings for the scale S1LEADP

	Standardised factor loadings					Unstandardised intercepts				
Participating country	SL1G33A	SL1G33B	SL1G33C	SL1G33D	SL1G33E	SL1G33A	SL1G33B	SL1G33C	SL1G33D	SL1G33E
ISCED level 02										
Chile	0.693	0.711	0.895	0.582	0.002	3.551	3.118	3.196	3.249	2.527
Germany	0.665	0.490	0.560	0.605	0.198	3.641	2.979	3.257	3.454	3.272
Iceland	0.586	0.652	0.718	0.554	0.039	3.455	2.888	2.972	3.215	2.171
Israel	0.550	0.428	0.569	0.644	0.138	3.311	2.479	2.876	3.189	2.009
Japan	0.597	0.786	0.775	0.470	-0.146	3.227	2.760	2.719	2.907	1.938
Korea	0.714	0.759	0.763	0.635	0.271	3.436	3.243	3.098	3.123	3.250
Norway	0.588	0.668	0.664	0.638	0.174	3.571	3.197	3.225	3.190	2.857
Turkey	0.728	0.730	0.662	0.780	0.398	3.441	3.172	3.092	3.399	3.337
Denmark	0.497	0.640	0.475	0.496	0.129	3.543	2.817	2.865	3.160	2.594
U3										
Germany	0.683	0.480	0.615	0.628	0.199	3.641	2.979	3.257	3.454	3.272
Israel	0.629	0.545	0.503	0.572	0.224	3.119	2.349	2.186	2.759	2.279
Norway	0.611	0.575	0.615	0.593	0.168	3.571	3.197	3.225	3.190	2.857
Denmark	0.777	0.371	0.615	0.427	0.371	3.372	2.766	2.786	3.191	2.494

Source: OECD, TALIS Starting Strong 2018 database.

Excluded scales

Satisfaction with ECEC centre, professions and working conditions (S1LJOBSA)

Despite improvements to the model, this scale still exhibited poor model fits in Germany, Israel and Norway (both populations) and Korea and Turkey ISCED level 02. As the scale potentially worked in no more than half of the populations, it was excluded from further analysis.

Table 11.115 presents the wording of and response options for the items in this scale.

Table 11.115. Item wording for the scale S1LJOBSA

S1LJOBSA: Satisfaction with ECEC centre, professions, and working conditions	
SL1G36: How strongly do you agree or disagree with the following statements?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SL1G36B	I am satisfied with the salary I receive from my work
SL1G36C	Apart from my salary, I am satisfied with the terms of my <ECEC centre leader> <contract/ employment> (e.g. benefits, work schedule)
SL1G36F	I am satisfied with the support that I receive from parents or guardians in this <ECEC centre>
SL1G36G	I am satisfied with the support that I receive from the staff in this <ECEC centre>
SL1G36H*	I need more support from <local, municipality/ regional, state, or national/federal> authorities

Note: *This item was reverse coded.
Source: OECD, TALIS Starting Strong 2018 database.

Sources of work stress (S1LSTRES)

This scale, the items of which are set out in Table 11.116, exhibited poor model fit indices based on the data from the pooled model. Added model improvements were not successful and attempts to find a meaningful multifactor solution failed.

Table 11.116. Item wording for the scale S1LSTRES

S1LSTRES: Sources of work stress	
SL1G37: Thinking about your job at this <ECEC centre>, to what extent are the following sources of stress in your work?	
Response options: strongly disagree (1), disagree (2), agree (3), strongly agree (4)	
SL1G37A	Having too much administrative work to do (e.g. filling out forms)
SL1G37B	Having extra duties due to absent <ECEC staff>
SL1G37C	Being held responsible for children's development, well-being and learning
SL1G37D	Keeping up with changing requirements from <local, municipality/regional, state, or national/federal> authorities
SL1G37E	Managing <ECEC staff>
SL1G37F	Addressing parent or guardian concerns
SL1G37G	Accommodating children with special needs
SL1G37H	A lack of support from local authorities or government
SL1G37I	A lack of resources (e.g. financial support and material resources)

Source: OECD, TALIS Starting Strong 2018 database.

Notes

1. The models used in the scaling procedure account for random measurement error, but also recognise that variables are measured with certain but not complete precision.

2. A pooled sample is a dataset where all the participating countries from ISCED level 02 were put together and analysed as one group. All factor loadings and item intercepts were freely estimated and the latent variances were fixed to one.

3. Constrained models were used for "insufficient" populations (see Chapter 10 for more details). These models were based on the parameters of the final scale models (see the subsection *Scale score estimation* of this chapter for further details).

4. For practical reasons, the two most recent versions of Mplus were used (versions 8 and 7.3).

5. The approach taken for continuous models followed the procedure implemented for the scale evaluation in the TALIS 2018 main study. This decision during TALIS 2018 was based on the analysis of the field trial data based on continuous and categorical models so as to assess the comparability of the results. Although the majority of TALIS items are ordinal, in both previous cycles (TALIS 2008 and 2013), linear models were used to estimate complex indices. During TALIS 2018, both models were used to evaluate the scales, followed by a comparison of their results. The results were very similar, although there were minor differences in terms of the performance and the measurement invariance level of the scale. Practical challenges related to a possible change in approach between TALIS cycles (i.e. from continuous CFA to categorical CFA modelling) resulted in the decision to implement the linear measurement model also for the TALIS 2018 scaling. This approach is supported by simulation studies (Van de Vijver et al., 2019[18]). The same procedure was used to evaluate the scales derived from the TALIS Starting Strong 2018 main study data.

6. Weighted least square parameter estimates using a diagonal weight matrix with standard errors and a mean- and variance-adjusted chi-square test statistic that uses a full weight matrix (Muthén, L. K. and B. O. Muthén, 1998–2017[21]).

7. This procedure meant that each participating country made an equal contribution to the estimates (Gonzalez, 2012[32]). The staff and leader weights for each participating country in each population were rescaled so that the total sum of the weights (i.e. the number of staff and leaders in the population) was equal to 3,000 and 500 respectively. The SPSS macro provided by Gonzalez (Gonzalez, 2012[32]) was used for rescaling.

8. A more detailed examination of the model parameters (factor loadings, residual variances) was part of further model analyses.

9. Principal axis factoring (PAF) is a form of EFA that is commonly adopted to examine the internal factor structure of constructs. Unlike principal component analysis (PCA), which is a linear combination of variables, PAF is a measurement model of latent constructs. Oblimin rotation was chosen over varimax rotation due to the assumption that the extracted factors within the constructs in TALIS Starting Strong 2018 correlated with one another (Brown, 2006[7]; Chen, P. P., T. J. Cleary and A. M. Lui, 2014[36]).

10. In the context of this report, a multidimensional scale is one that acknowledges covariance relationships between several factors. A hierarchical scale is one that includes a second-order factor with factor loadings on the first-order factors.

11. Participating countries that did not meet the technical standards were not included in this stage of the analysis. In addition, certain countries were excluded for specific scales. For a detailed explanation, see the section of this chapter titled *Results from scale evaluation* and Table 11.6.

12. Four countries participated in the U3 population and nine countries in the ISCED level 02 population. However, two populations, one from each ISCED level 02 and U3, were excluded, as they did not meet the technical standards for the study.

13. The measurement parameters describe measurement characteristics of observed variables (items). Factor loadings indicate the strength of the relationship between the item and the latent factor, intercepts indicate the predicted values for an item when the value of the latent factor is zero, and residual variances present the portion of item variance that is not attributable to variance in the latent factor.

14. In cases where one or more items were not administered for some groups, and thus one or more phantom indicators were used, a partial measurement invariance approach was employed. In such cases, both the intercept and the factor loading were fixed to zero for each phantom variable in the group(s) concerned.

15. For these models, a minimum number of parameters were constrained for reasons of model identification. During the analysis, the latent variances were set to one for each group, thus allowing for free estimation of factor loadings and intercepts.

16. In strict invariance, not only the factor loadings and item intercepts but also the residual variances of the items have to be equal across groups. This requirement means that the portion of the item variance not attributable to variance in the latent construct is the same across groups. However, this assumption is very hard to meet in practice, especially in large-scale assessments where many groups are compared. In line with previous TALIS cycles, the current cycle did not test these models.

17. To be specific, for a certain population, p CFA models were created, where p is the number of countries within that population.

18. This analysis was conducted only for those countries whose participation encompassed both the ISCED level 02 and U3 populations.

19. Configural, metric and scalar models are in essence nested models: the scalar model is nested in the metric model, and the metric model is nested in the configural model. Therefore, the χ^2 difference test can be adopted to evaluate which model fits the data best. If the χ^2 difference value is significant, the less restrictive model (the model with more freely estimated parameters) fits the data better than the nested, more restrictive model (the model with fixed/constrained parameters). If the χ^2 difference value is not significant, both models fit the data equally well. However, because χ^2 is sensitive to sample size (and the sample size for TALIS Starting Strong was considerable), changes in the model fit indices (e.g., CFI, TLI, RMSEA, SRMR/WRMR) were used to evaluate the measurement invariance of each scale.

20. For scales measured with categorical variables, the standardised factor loadings come from the STDYX standardisation of Mplus, while for scales measured with binary items they come from the STDY standardisation (Muthén, L. K. and B. O. Muthén, 1998–2017, pp. 799-802[21]).

21. The mean score does not always rise and fall in this way. Certain aggregate scores are observed much less than others. In this example, with the scale S1PDPE, aggregate scores of 1.25, 1.33, 1.50, 1.67, 1.75, 2.25, 2.33, 2.50, 2.67, 2.75, 3.25, 3.33, 3.50, 3.67 and 3.75 are less common, resulting in a smaller number of observations for these values. Therefore, the variance is not very large, and an outcome may result in a mean that is lower or greater than expected. However, the correlation between the simple average of item scores and factor scores is still generally strong.

22. *Facilitating socioemotional development* is composed of the following subscales: facilitating child initiative, facilitating prosocial behaviour and facilitating emotional development. *Facilitating learning and development* is composed of these subscales: facilitating language development, facilitating literacy development and facilitating numeracy development. Although *Sources of work stress* is a multidimensional scale, it is the second-order scale in a hierarchical model, which means the scale scores are directly computable. For further information, see endnote 10.

23. This is known as a *formative measurement model* approach. Formative models seek to fully cover the construct's domain (Diamantopoulos, 2001[41], 2008[42]).

24. For example, in relation to satisfaction with work environment or working conditions, a positive response of "strongly agree" to the item "I enjoy working at this <ECEC centre>" is also positive for the latent construct of satisfaction with work environment or working conditions. However, a positive response of "strongly agree" to the item "I need more support from my <ECEC centre leader>" is negatively related to the latent construct, meaning this item would be reverse coded. As such, a response of "strongly disagree" would now be coded (i.e. assigned a higher integer) to align with a positive association with the latent construct for this particular item.

25. The purpose of conducting a CFA on a pooled sample during the evaluation process was to obtain a general overview of scale performance; the results from this step are not presented in this technical report. In addition, the model parameters for the excluded populations were fixed by using the results from the final scale modelling, meaning these countries/entities were not considered during the reliability analyses at the country level.

26. In models with three items, the number of informative parameters in the variance-covariance matrix equals the number of parameters being estimated, resulting in zero degrees of freedom for these models ($df = 0$). This offers a unique solution to parameter estimation because the models are just identified. Models based on three items can therefore not be evaluated with respect to their fit to the empirical data (because the model fit indices suggest perfect model fit).

References

Asparouhov, T. (2018), *SRMR in Mplus: Technical report*, http://www.statmodel.com/download/SRMR2.pdf. [13]

Beauducel, A. Harms, C. and N. Hilger (2016), "Reliability estimates for three factor score estimators", *International Journal of Statistics and Probability*, Vol. 5/6, p. 94, http://dx.doi.org/10.5539/ijsp.v5n6p94. [37]

Brown, J. (2009), "Choosing the right type of rotation in PCA and EFA.", *Shiken: JALT Testing & Evaluation SIG Newsletter*, Vol. 13/3, pp. 20–25, http://hosted.jalt.org/test/PDF/Brown31.pdf. [39]

Brown, T. (2015), *Confirmatory Factor Analysis for Applied Research, 2nd edition*, The Guilford Press, New York. [38]

Brown, T. (2014), *Confirmatory Factor Analysis for Applied Research*, Guilford Press, New York. [27]

Brown, T. (2006), *Confirmatory Factor Analaysis for Applied Research*, Guilford Publications, New York. [12]

Chen, P. P., T. J. Cleary and A. M. Lui (2014), "Examining parents' ratings of middle-school students' academic self-regulation using principal axis factoring analysis", *School Psychology Quarterly*, Vol. 30/3, pp. 385–397. [31]

Chen, F. (2007), "Sensitivity of goodness of fit indexes to lack of measurement invariance", *Structural Equation Modeling*, Vol. 14/3, pp. 464–504. [20]

Cheung, G. (1999), "Testing factorial invariance across groups: A reconceptualization and proposed new method", *Journal of Management*, Vol. 25/1, pp. 1–27. [15]

Cheung, G. and B. Rensvold (2002), "Evaluating goodness-of-fit indexes for testing measurement invariance", *Structural Equation Modeling*, Vol. 9/2, pp. 233–255. [3]

Cheung, G. and R. Rensvold (1998), "Cross-cultural comparisons using non-invariant measurement items", *Applied Behavioral Science Review*, Vol. 6/1, pp. 93–110. [45]

Cronbach, L. J. and P. E. Meehl (1955), "Construct validity in psychological tests", *Psychological Bulletin*, Vol. 52, pp. 281–302. [40]

Davidov, E. et al. (2014), "Measurement equivalence in cross-national research", *Annual Review of Sociology*, Vol. 40, pp. 55–75. [19]

Davidov, E. (2008), "A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey", *Survey Research Methods*, Vol. 2/1, pp. 33–46. [16]

Dempster, A. P., N. M. Laird and D. B. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39/1, pp. 1–22. [33]

Desa, D., E. Gonzalez and P. Mirazchiyski (2014), *Chapter 10: Construction of scales and indices*, OECD Publishing, Paris. [41]

DeVellis, R. (2012), *Scale development: Theory and applications*, Sage, Los Angeles. [29]

DeVellis, R. (2003), *Scale development: Theory and applications*, Sage, Los Angeles. [46]

Diamantopoulos, A. (2008), "Advancing formative measurement models", *Journal of Business Research*, Vol. 61/12, pp. 1203–1218. [35]

Diamantopoulos, A. (2001), "Index construction with formative indicators: An alternative to scale development", *Journal of Marketing Research*, Vol. 38/2, pp. 269–277. [34]

Geldhof, G. J. et al. (2013), "Orthogonalizing through residual centering: Extended applications and caveats", *Educational and Psychological Measurement*, Vol. 73/1, pp. 27–46. [14]

Gonzalez, E. (2012), "Rescaling sampling weights and selecting mini-samples from large-scale assessment databases", *IERI Monograph Series*, Vol. 5, pp. 117–134. [32]

Grice, J.W. (2001), "Computing and Evaluating Factor Scores", *Psychological Methods*, Vol. 6, pp. 430–450. [42]

Hansen, Y., K. M. Rosen and J. E. Gustavson (2006), "Measures of self-reported reading resources, attitudes and activities based on latent variable modelling", *International Journal of Research & Method in Education*, Vol. 29/2, pp. 221–237. [36]

He, Q. (2010), *Estimating the reliability of composite scores*, https://dera.ioe.ac.uk/1060/1/2010-02-01-composite-reliability.pdf. [7]

Hoyle, R. (2014), *Handbook of Structural Equation Modeling*, The Guilford Press, New York. [30]

Hu, L. and P. Bentler (1999), "Cutoff criteria for fit indexes in covariance structure snalysis: Conventional criteria versus new alternatives", *Structural Equation Modeling*, Vol. 6/1, pp. 1–55. [9]

Lüdtke, O. et al (2007), "Umgang mit fehlenden Werten in der psychologischen Forschung [Dealing with missing values in psychological texts]", *Psychologische Rundschau*, Vol. 58/2, pp. 103–117. [24]

Lyubomirsky, S. and H. Lepper (1999), "A measure of subjective happiness: Preliminary reliability and construct validation", *Social Indicators Research*, Vol. 46/2, pp. 137–155, http://dx.doi.org/10.1023/A:1006824100041. [2]

Messick, S. (1995), "Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning", *American Psychologist*, Vol. 50, No.9, pp. 741-749. [1]

Muthén, L. K. and B. O. Muthén (1998–2017), *Mplus User's Guide*, Muthén & Muthén, Los Angeles. [21]

Muthén, L.K. and B.O. Muthén (1998-2012), *Mplus User's Guide" (computer software, 7th edition), Muthén & Muthén, Los Angeles, California*. [28]

Muthén, B. (1977), *Some results on using summed raw scores and factor scores from dichotomous items in the estimation of structural equation models*, University of Uppsala, Uppsala, http://www.statmodel.com/bmuthen/articles/Muthen_Unpublished_01.pdf. [23]

OECD (2019), *Providing Quality Early Childhood Education and Care: Results from the Starting Strong Survey 2018*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/301005d1-en. [47]

OECD (2014), *TALIS Technical Report 2013*, OECD Publishing Paris. [8]

OECD (2010), *TALIS 2008 Technical Report*, OECD Publishing, Paris. [10]

Schafer, J. L. and J. W. Graham (2002), "Missing data: our view of the state of the art", *Psychological Methods*, Vol. 7/2, pp. 147–177. [25]

Schermelleh-Engel, K. (2003), "Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures", *Methods of Psychological Research Online*, Vol. 8/2, pp. 23–74. [43]

Schreiber, J. et al. (2006), "Reporting structural equation modeling and confirmatory factor analysis results: A review", *The Journal of Educational Research*, Vol. 99/6, pp. 323–338, https://doi.org/10.3200/JOER.99.6.323-338. [11]

Skrondal, A. and P. Laake (2001), "Regression among factor scores", *Psychometrika*, Vol. 66, pp. 563–575. [22]

Steenkamp, J.-B. and E. M. Baumgartner (1998), "Assessing measurement invariance in cross-national consumer research", *Journal of Consumer Research*, Vol. 25/1, pp. 78–90. [17]

Steiger, J. (1990), "Structural model evaluation and modification: An interval estimation approach", *Multivariate Behavioral Research*, Vol. 25/2, pp. 173–180. [44]

Van de Vijver, F. et al. (2019), "Invariance analyses in large-scale studies", *OECD Education Working Papers*, No. 201, OECD Publishing, Paris, https://dx.doi.org/10.1787/254738dd-en. [18]

Vandenberg, R. and C. Lance (2000), "A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research", *Organizational Research Methods*, Vol. 3/1, pp. 4–70. [4]

Viladrich, C., A. Angulo-Brunet and E. Doval (2017), "A journey around alpha and omega to estimate internal consistency reliability", *Anales de Psicología*, Vol. 33/3, pp. 755–782. [5]

Yu, C. (2002), *Evaluating cutoff criteria for model fit indices for latent variable models with binary and continous outcomes*, unpublished dissertation, University of California, Los Angeles. [26]

Zhang, Z. and K. Yuan (2016), "Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software", *Educational and Psychological Measurement*, Vol. 76/3, pp. 387–411. [6]

Chapter 12. Table production and verification, analyses

This chapter describes the process used to produce the tables for the first volume of international findings from Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) 2018. It covers the procedures and activities involved in developing the table shells up to verification of the results. The chapter also provides an overview of conducted analyses, applied quality rules, software used, and the different parties involved in ensuring high-quality results.

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

12.1. Introduction

The process and procedures presented in this chapter relate to the tables of results produced by the IEA Hamburg and the OECD for *Providing Quality Early Childhood Education and Care: Results from the Starting Strong Survey 2018* (OECD, 2019[1]). Preparation of the tables consisted of two major steps: (1) the development, review and revision of table shells; and (2) data analysis, table production and verification. After providing a brief description of the first step, this chapter focuses on the second step. The chapter also provides an overview of the procedures and methods used to estimate population parameters. It furthermore highlights relevant statistical issues using selected, but shortened,[1] tables from the international report.

12.2. Responsibilities

Responsibilities for developing and producing the tables for the TALIS Starting Strong 2018 international report were shared between the OECD and the TALIS Starting Strong Consortium, specifically the IEA and the Australian Council for Educational Research (ACER). The OECD developed the table shells and determined the content and format in line with (a) the analysis plan produced by the TALIS Starting Strong Consortium, and (b) the reporting plan developed by the OECD and agreed to by the participating countries. The table shells were then reviewed by the TALIS Starting Strong 2018 International Consortium for statistical and substantial soundness and revised by the OECD over multiple rounds.

In a subsequent step, the TALIS Starting Strong Consortium's team at the IEA Hamburg conducted the analyses, populated around 120 tables across the two study populations (ISCED level 02 and U3) in collaboration with the OECD, and verified/validated estimates in collaboration with ACER and the OECD. The IEA team members discussed methodological issues relating to these tables amongst themselves and with the OECD. During this process, both parties took into consideration advice provided by the survey's Technical Advisory Group (TAG). The participating countries also contributed, via the OECD Secretariat, to developing and reviewing the tables.

12.3. Populating the table shells: general procedures

The process of populating the table shells consisted of two major steps. First, the IEA IDB Analyzer was used to estimate, for each table, all parameters (percentages, means, etc.) and their respective standard errors.[2] Second, R packages[3] were used to transfer estimates resulting from the first step to the appropriate table shells. The code book in Annex J of this technical report includes information on (a) the variables created and delivered in the public TALIS Starting Strong 2018 dataset, and (b) the advanced recoded variables generated in the scope of table production.

All tables presented in the TALIS Starting Strong 2018 international report (OECD, 2019[1]) were structured in a manner similar to the example shown in Table 12.1, and each accorded with the following elements and principles:

1. estimates to always be presented along with their standard errors (SEs) (items 1 and 2 in Table 12.1.)

2. country-specific considerations: data from countries not meeting the requirements determined in the technical standards (see Chapter 10 of this report) to appear below the main body of the table.

Table 12.1. Results based on leader reports: Percentage of female and male ISCED level 02 leaders

	Gender			
	Female		Male	
Participating country	(1) %	(2) SE	(1) %	(2) SE
ISCED level 02				
Chile	76.8	(2.4)	23.2	(2.4)
Germany*	95.1	(1.4)	4.9	(1.4)
Iceland	97.2	(1.0)	2.8	(1.0)
Israel	98.8	(0.5)	1.2	(0.5)
Japan	64.2	(5.2)	35.8	(5.2)
Korea	87.4	(3.5)	12.6	(3.5)
Norway	93.4	(2.2)	6.6	(2.2)
Turkey	24.7	(3.0)	75.3	(3.0)
Denmark**	89.5	(2.8)	10.5	(2.8)

* Estimates for sub-groups and estimated differences between sub-groups need to be interpreted with care.
** Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

Although estimated differences between subgroups were computed for several tables, the differences themselves do not appear in the tables. However, if an estimated difference tested significantly different from zero, the values of the subgroups are set in bold in the tables. Readers can identify these tables by the annotations below them. For more information on statistical testing, see Chapter 9 of this technical report and the *TALIS 2018 and TALIS Starting Strong User Guide* (OECD, 2019[2]).

12.4. Estimating standard errors using weights

One of the main goals of TALIS Starting Strong 2018 was to generate reliable, valid and comparable estimates of population parameters based on sample data. Consequently, all estimates presented in the TALIS Starting Strong 2018 international report (OECD, 2019[1]) are weighted (see, for example, the percentages in Table 12.1) (OECD, 2019[1]). Centre weights (variable name: CNTRWGT) were used to weight the results produced with leader data, while staff weights (variable name: STAFFWGT) were used to weight the results produced with either staff data or combined staff and centre leader data (i.e. centre/leader information merged with staff records).

Because all estimates in the TALIS Starting Strong 2018 international report were based on sample data, they could only be estimated with a degree of uncertainty (OECD, 2019[1]). Thus, results from analyses of these data and the precision of the estimates must be reported together. In the tables, the degree of uncertainty of an estimate is reflected by its standard error (SE; item 2 in Table 12.1), which has the same metric as the estimate.

Fay's variant of the balanced repeated replication (BRR) technique was used to estimate the standard errors during the weighting stage. BRR estimates of sampling error can be computed with the IEA IDB Analyzer. Chapter 9 of this report provides more information

on computing sampling weights and about the BRR technique. Chapter 3 of the *TALIS 2018 and Starting Strong User Guide to the International Database* (OECD, 2019[2]) provides details on using weights for data analysis. Finally, Chapter 9 presents advice on obtaining standard errors for differences of estimates during analyses of data from different samples or from the same sample.

12.5. Estimating percentiles

Percentiles were estimated with the IEA IDB Analyzer, in which the following procedure was implemented. In a first step, the values were sorted and a vector of accumulated weighted frequencies was produced. In a second step, the first value to exceed the percentile threshold (10th, 25th, 50th, 75th, and 90th) became the respective percentile value.

"National quarters" were calculated for some of the TALIS Starting Strong 2018 tables. (Table 12.2 provides an example.) The estimates for the bottom and the top national quarters were derived by calculating the percentiles of each national sample. The rules defining the national quarters were as follows:

- *Top national quarter*: all cases with a value the same as or higher than the value of the 75th percentile to be included.
- *Middle national quarters*: all cases with a value the same as or higher than the value of the 25th percentile and all cases with a value lower than the value of the 75th percentile to be included. Thus, all cases lying between the bottom quarter and the top quarter needed to be included.
- *Bottom national quarter*: all cases with a value lower than the value of the 25th percentile to be included.

Table 12.2. Results based on U3 staff reports: percentage of U3 staff who had an ISCED level 6 degree or equivalent, by centre characteristics

Percentage of U3 staff who had an ISCED level 6 degree or equivalent, by centre size						
Participating country	Bottom quarter of centre size		Middle quarters of centre size		Top quarter of centre size	
U3	%	SE	%	SE	%	SE
All staff						
Germany*	66.2	(2.9)	68.5	(2.5)	71.4	(3.2)
Israel	18.2	(3.0)	11.5	(1.7)	11.0	(2.7)
Norway	55.0	(4.9)	49.8	(2.2)	51.3	(2.0)
Denmark**	66.1	(7.0)	67.4	(3.4)	74.7	(5.8)

* Estimates for sub-groups and estimated differences between sub-groups need to be interpreted with care.
** Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

12.6. Use of weights in tables presenting analyses of staff and leader data

Some of the TALIS Starting Strong 2018 questionnaire items were part of both the leader and the staff questionnaires, a feature of the survey that made it possible to compare the leaders' answers with answers given by staff. To populate the table that included results derived from both leader and staff data (i.e. Table 12.3), the leader data and the leader weights were used to calculate the results from the leader questionnaire, while the staff dataset and staff weights were used to calculate the results from the staff questionnaire.

The procedure of using leader weights to weight the results produced with leader data differed from the corresponding method used in TALIS 2018, where the principal data were merged with the teacher data, and the teacher weights were then used to conduct the analyses. This difference means that the results presented in the TALIS Starting Strong 2018 international report (OECD, 2019[1]) allow for inferences relating to both the staff and the centre leader populations, whereas the principal data featured in the tables of the TALIS 2018 international report have to be interpreted as a feature of teachers (OECD, 2019[3]).

Table 12.3. Results based on staff and leader reports: percentage of ECEC staff and leaders who thought the following skills and/or abilities are of high importance in preparing children for the future

	Children's oral language skills			
	Staff		Leaders	
Participating country	%	SE	%	SE
ISCED level 02				
Chile	85.1	(1.2)	99.4	(0.6)
Germany*	95.2	(0.8)	99.3	(0.7)
Iceland	87.8	(1.2)	97.0	(1.4)
Israel	84.9	(1.2)	c	c
Japan	56.4	(1.7)	62.1	(4.8)
Korea	60.0	(2.1)	64.8	(4.6)
Norway	84.4	(1.8)	96.7	(1.9)
Turkey	92.7	(0.7)	93.8	(2.0)
Denmark**	92.6	(1.3)	88.8	(4.9)
U3				
Germany*	93.1	(0.9)	97.3	(1.6)
Israel	83.5	(1.3)	97.3	(1.3)
Norway	85.7	(1.7)	91.7	(2.6)
Denmark**	92.2	(1.5)	100.0	(0.0)

Note: To calculate the results for this table, the data from centre leaders with staff duties were excluded from the dataset to avoid overestimating the correlation between staff members' and centre leaders' views.
* Estimates for sub-groups and estimated differences between sub-groups need to be interpreted with care.
** Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

12.7. Handling of filter-dependent questions, not administered items and selected subgroups

Some analyses in TALIS Starting Strong 2018 focused on a defined subgroup of the target population, either by (a) using filter and filter-dependent questions, (b) considering only respondents to whom a question was administered, or (c) focussing on a specific group for interpretative or analytical reasons, for example, by considering only centre-based settings.

The respective subgroups emanating from filter-dependent questions were identified by their responses to the respective filter questions. For example, question 11 in the staff questionnaire (variable SS1G11) asked: "How would you describe the <education or training> programme that prepared you to work with children?" This question could only be answered by staff who gave the answer "Yes" to the preceding filter question (variable SS1G10): "Have you completed an <education or training> programme that prepared you

to work with children?" Here, SS1G10 is the *filter question* and SS1G11 is the *filter-dependent question*.

Estimates involving filter-dependent questions were therefore based only on those respondents filtered in by the preceding corresponding filter question. This rule was implemented by treating cases not meeting the filter criteria (i.e. observations with *logically not applicable* codes in the dataset) as missing values. (More information about missing codes can be found in Chapter 8 of this report.) If a respondent missed a filter question but then answered the subsequent filter-dependent question or questions, his or her responses to those filter-dependent questions were included in the analyses. (This information is available in the header of each table providing the selection criteria of the filter.)

For various reasons related to the distinctive features of each of the ECEC systems in the participating countries, all items in the questionnaires were not necessarily delivered to all respondents (see Chapter 3). In this case only centre leaders and staff to whom the question was administered could be considered in the analyses. Rare exceptions were made in the case of specific recoded variables.[4]

Some of the participating countries opted to cover home-based care of children under the age of three years. Because the focus for the TALIS Starting Strong 2018 international report (OECD, 2019[1])was centre-based settings only, a filter variable, CENTRETYPE, was used to filter in centre-based settings.[5] When applicable, footnotes below the tables in the TALIS Starting Strong 2018 international report denote which group or subgroup the data pertain to in those tables.

12.8. Annotation rules related to data quality requirements (minimum number of cases, item non-response)

The TALIS 2018 and TALIS Starting Strong 2018 Consortia, in collaboration with the OECD and the TAG, were responsible for developing the annotation scheme (see Box 12.1) used to flag potential data quality issues in the tables in the TALIS Starting Strong 2018 international report (OECD, 2019[1]). The scheme was very similar to the ones used in the TALIS 2013 and PISA (Programme for International Student Assessment) 2015 international reports.

The team conducting the data quality checks first applied the *item non-response rule* and then the *minimum sample rule*. The first check (the item non-response rule) verified the percentage of respondents who answered a given question. This rule ensured that the number of respondents who gave a valid answer to a question multiplied by 100 divided by the number of participants to whom the question was applicable was never lower than 50%.

Box 12.1. Annotation rules relating to data quality requirements

a The question was *not administered in the country* because it was optional. Data were therefore deemed to be missing data.

c One or more of the following conditions applied:

- There were too few or no observations to provide reliable estimates and/or to ensure the confidentiality of respondents; that is, there were fewer than 10 centres/leaders or fewer than 30 staff with valid data.
- The item non-response rate was above 50%; that is, the ratio of missing or invalid responses to the number of respondents for whom the question was applicable was above 50%.

m Data were collected but subsequently removed for technical reasons (e.g. erroneous translation) as part of the data checking process.

w Data were withdrawn or were not collected at the request of the country concerned.

The second data quality check, applying the *minimum sample rule*, ensured that an estimate was based on at least 30 staff (for population estimates based on staff weights) and 10 leaders/centres (for population estimates based on leader weights). If a participating country did not meet these requirements, the relevant data were not analysed, an occurrence indicated by the symbol "c" in the tables. For example, in the TALIS Starting Strong 2018 dataset for Norway, fewer than 30 teachers[6] (a) gave a valid answer to the question on their education level and (b) work in centres where the proportion of children from socio-economically disadvantaged homes exceeded 30% according to the answer given by the leader of the centre (see Table 12.4, column "More than 30%").

Table 12.4. Results based on U3 staff reports: percentage of U3 teachers who had at least an ISCED level 6 degree or equivalent, by centre characteristics

	Percentage of U3 teachers who had at least an ISCED level 6 degree or equivalent by concentration of children from socio-economically disadvantaged homes[3]					
Participating countries	Less than or equal to 10%		Between 11% and 30%		More than 30%	
U3	%	SE	%	SE	%	SE
Teachers						
Germany*	76.7	(2.0)	78.9	(4.4)	74.3	(6.9)
Israel	12.9	(1.8)	11.7	(2.8)	8.3	(3.0)
Norway	97.8	(0.9)	95.2	(3.4)	c	c
Denmark**	96.6	(1.7)	96.0	(3.7)	c	c

* Estimates for sub-groups and estimated differences between sub-groups need to be interpreted with care.
** Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

12.9. Quality assurance and table verification

To ensure high quality information, the results presented in the tables of the TALIS Starting Strong 2018 international report (OECD, 2019[1]) underwent up to three review and revision rounds before publication. During each of these table releases, the TALIS Starting

Strong Consortium's team at the IEA Hamburg, along with the OECD and the participating countries, verified the plausibility of results.

ACER meanwhile validated and verified the statistical procedures that the IEA Hamburg applied by reproducing all estimates presented in the tables. ACER recalculated the table results via self-developed tools that included the replication method (Fay-BRR).

The verification steps encompassed the following:

- general plausibility checks
- different reviewers verifying the variables, subsettings and recodings several times over
- recalculation of randomly selected results from each table (usually one or two randomly selected columns of a table)
- comparison of results against the ACER-produced results.

Notes

1. Tables were shortened to highlight important statistical aspects and thus to increase readability. All footnotes not important in the context of this chapter were deleted from the tables but can be found in the international report of the study's findings.

2. The IEA IDB Analyzer is a software application developed by the IEA to perform analysis of data collected during international large-scale assessments of education. It provides, via a user interface, SPSS or SAS syntax that can be run with the respective software (see www.iea.nl/data). Version 4.0.26 of the IEA IDB Analyzer was used to produce the tables for the TALIS Starting Strong 2018.

3. The following R packages were used: data.table, openxlsx. For more information on R, go to https://www.r-project.org.

4. The exceptions are denoted respectively in the tables in Annex J with this note: "Not administered does not count as missing, only if all variables are not administered", e.g. SUMSTAFF (number of ECEC staff).

5. The filtering was implemented for U3 only because the ISCED level 02 population included only centre-based ECEC settings.

6. The variable STAFFROLE was used to identify teachers in the staff dataset.

References

OECD (2019), *Providing Quality Early Childhood Education and Care: Results from the Starting Strong Survey 2018*, TALIS, OECD Publishing, Paris, https://doi.org/10.1787/301005d1-en. [1]

OECD (2019), *TALIS 2018 and TALIS Starting Strong 2018 User Guide*, OECD, Paris, http://www.oecd.org/education/talis/TALIS_2018-TALIS_Starting_Strong_2018_User_Guide.pdf. [2]

OECD (2019), *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners*, TALIS, OECD Publishing, Paris, https://dx.doi.org/10.1787/1d0bc92a-en. [3]

Annex A. Consortium, experts and consultants

IEA Hamburg (Hamburg, Germany)

Core management team

Juliane Hencke (International Study Director)

Steffen Knoll (Deputy International Study Director)

Ralph Carstens (International Study Director TALIS)

Viktoria Gabriele Böhm (International Study Co-ordinator)

Juliane Kobelt (International Study Co-ordinator)

Marta Kostek (International Study Co-ordinator)

Malgorzata Petersen (International Study Co-ordinator Assistant)

Data management team

Alena Becker (International Data Manager)

Christine Busch (Deputy International Data Manager)

Regina Borchardt (Research Analyst)

Olesya Drozd (Research Analyst)

Wolfram Jarchow (Research Analyst)

Hannah Kowolik (Research Analyst)

Kamil Kowolik (Research Analyst)

Adeoye Babatunde Oyekan (Research Analyst)

Oriana Mora (Research Analyst)

Xiao Sui (Research Analyst)

Svenja Kalmbach (Student Assistant)

Lena Talihmanidis (Student Assistant)

Sampling and weighting team

Sabine Meinck (International Sampling Advisor)

Karsten Penon (International Sampling Manager)

Sabine Weber (Deputy International Sampling Manager)

Duygu Savaşcı (Research Analyst, Sampling)

Sabine Tieck (Research Analyst, Sampling)

Scaling, analysis and table production team

Agnes Stancel-Piątak (Senior Research Analyst, Team Leader Scaling and Analysis)

Umut Atasever (Research Analyst)

Falk Brese (Senior Research Analyst)

Minge Chen (Research Analyst)

Diego Cortes (Research Analyst)

Deana Desa (Research Analyst)

Ann-Kristin Koop (Research Analyst)

Mojca Rozman (Research Analyst)

Justin Wild (Research Analyst)

Conrad Baumgart (Student Assistant)

Isbat Hasnat (Student Assistant)

Meeting organisation

Catherine Pfeifer (Meeting Co-ordinator)

Bettina Wietzorek (Meeting Co-ordinator and SharePoint Administrator)

Software development and testing

Meng Xue (Head of Software Development Unit)

Limiao Duan (Programmer)

Anne Guttmann (Project Manager eAssessment)

Christian Harries (Programmer)

Majid Iqbal (Software Tester)

Maike Junod (Programmer)

Deepti Kalamadi (Programmer)

Ievgen Kosievtsov (Programmer)

Kevin Mo (Programmer)

Suma Padala (Software Tester)

Devi Potham Rajendra Prasath (Programmer)

Svetoslav Velkov (Software Tester)

Juan Vilas (Programmer)

Samih Al-Areqi (Requirements Engineer and Software Tester)

Rea Car (Requirements Engineer and Software Tester)

Elma Cela (Requirements Engineer and Software Tester)

Michael Jung (Requirements Engineer and Software Tester)

Lorelia Lerps (Requirements Engineer and Software Tester)

Ekaterina Mikheeva (Requirements Engineer and Software Tester)

Yasin Afana (Research Analyst)

Clara Beyer (Research Analyst)

Tim Daniel (Research Analyst)

Dirk Oehler (Research Analyst)

Darrell Gwaltney (Student Assistant)

IEA Amsterdam (Amsterdam, the Netherlands)

Andrea Netten (Director, IEA Amsterdam)

David Ebbs (Senior Research Officer)

Michelle Djekić (Research and Liaison Officer)

Sandra Dohr (Junior Research Officer)

Jan-Philipp Wagner (Junior Research Officer)

Roel Burgers (Financial Director)

Isabelle Gémin (Senior Financial Officer)

Translation verification was performed in co-operation with cApStAn Linguistic Quality Control, an independent linguistic quality control agency located in Brussels, Belgium. The IEA Amsterdam appointed, contracted and trained independent quality observers to monitor survey implementation in each participating country.

RAND Europe (Cambridge, United Kingdom)

Julie Belanger (Research Leader)

Megan Sim (Senior Analyst)

Miriam Broeks (Analyst)

Katherine Stewart (Analyst)

Statistics Canada (Ottawa, Canada)

Jean Dumais (Sampling Referee)

Australian Council for Educational Research (ACER, Melbourne, Australia)

Leigh Patterson (Analysis Coordinator)

Renee Kwong (Project Analyst)

Dulce Lay (Project Analyst)

Consultants

Hynek Cigler (Masaryk University, Czech Republic)

Eugenio J. Gonzalez (Educational Testing Service, United States)

Plamen Mirazchiyski (International Educational Research and Evaluation Institute, Slovenia)

TALIS Starting Strong Expert Groups

Questionnaire Expert Group

Julie Belanger (RAND Europe, United Kingdom, Chair)

Alejandra Cortazar (Centro de Estudios Primera Infancia, Chile)

Edward Melhuish (University of Oxford, United Kingdom)

Henrik Daae Zachrisson (University of Oslo, Norway)

Masatoshi Suzuki (Hyogo University of Teacher Education, Japan) assisted by Yumi Yodogawa (University of Tokyo, Japan) and Sakiko Sagawa (Nara University of Education, Japan)

Clara Barata (OECD, France, *ex-officio*)

Ralph Carstens (IEA Hamburg, Germany)

Jean Dumais (Statistics Canada, Canada, *ex-officio*)

Arno Engel (OECD, France, *ex-officio*)

Lynn Karoly (RAND Corporation, *ex-officio*)

Agnes Stancel-Piątak (IEA Hamburg, Germany)

Miho Taguma (OECD, France, *ex-officio*)

Fons van de Vijver (University of Tilburg, the Netherlands, *ex-officio*)

Extended Questionnaire Expert Group

Sharon Lynn Kagan (Teachers College, Columbia University, United States)

Trude Nilsen (University of Oslo, Norway)

Nirmala Rao (Hong Kong University, China)

Pauline Slot (Utrecht University, the Netherlands)

Susanne Viernickel (University of Leipzig, Germany)

Technical Advisory Group

Fons van de Vijver (University of Tilburg, the Netherlands, Chair)

Pascal Bressoux (Université Grenoble Alpes, France)

Timothy L. Kennel (US Census Bureau, United States)

Paul Leseman (Utrecht University, the Netherlands)

Bart Meulemann (University of Leuven, Belgium)

Christian Monseur (University of Liège, Belgium)

Annex B. Technical standards

You can find the full technical standards on line at:

www.oecd.org/education/talis/Annex-B-TALIS-Starting-Strong-2018-Technical-Report-Technical-Standards.pdf

Annex C. Sampling forms

Figure A C.1. Sampling form 1A

Sampling Form 1A — General Information

See chapter 2.1 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

National Sampling Manager:

1. Please indicate for which target populations your country plans to participate:

ISCED Level 0.2	*Please Select Yes or No*
U3	*Please Select Yes or No*

2. Specify the usual start and end date of the school year.

Start of school year :	End of school year:

3. Specify the expected survey administration period for the Field Trial and the Main Survey *(see chapter 3 of National Project Managers Manual)* .

Expected period for Field Trial:	Expected period for Main Survey:

4. Specify the language(s) in which the survey will be administered.

5. Describe the age and birth date rules for entering ISCED Level 1 in your country.

6. Do you wish to survey another target population (national option)?
If so, please specify your wishes.

Figure A C.2. Sampling form 1B

Sampling Form 1B — Structure of Early Childhood Education and Care System

See chapter 2.1 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

1. Please describe the structure of Early Childhood Education and Care System through ISCED Level 0 in your country.

Name of programme/setting type (national name)	English translation	Typical age range	Distinctive characteristics	ISCED Level 0.2 (yes/no)	U3 (yes/no)

Figure A C.3. Sampling form 2

Sampling Form 2 — Coverage and Exclusions

See chapter 3.3 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

1. This Sampling Form refers to:

ISCED Level 0.2	U3

2. Total enrollment in the target population: [a]
   If number of staff is not available, please fill in the number of children instead

# of centres	# of staff	# of children

3. Centre-level exclusions *(if applicable):*

	Description of exclusions	# of centres	# of staff	# of children
1.				
2.				
3.				
4.				
5.				
TOTAL:	*(Sum of exclusions - Calculated automatically)* [b]	-	-	-

	centres	staff	children
4. **Percentage of centre-level exclusions:** *(Box [b] ÷ Box [a])* **[1]**	%	%	%

5. Total enrollment after centre-level exclusions: [c]
   (Box [c] = Box [a] - Box [b])

# of centres	# of staff	# of children

6. Within-centre exclusions *(if applicable):*

	Description of exclusions	# of staff
1.		
2.		
3.		
TOTAL:	*(Sum of exclusions - Calculated automatically)* [d]	-

	centres	staff	children
7. **Expected percentage of within-centre exclusions:** *(Box [d] ÷ Box [c])* **[2]**		%	

# of centres	# of staff	# of children
8. **Expected percentage of reduced coverage and exclusions:** *(Box [1] + (1 - Box [1]) x Box [2])*		
%	%	%

Figure A C.4. Sampling form 3

Sampling Form 3 — Stratification

See chapter 3.4 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

1. This Sampling Form refers to:

ISCED Level 0.2	U3

Stratification of centres

2. List and describe the variables to be used for stratification in order of importance:
   (Please note that the choice of variables used for explicit or implicit stratification will be discussed during consultations with the Sampling Team)

Stratification Variables			
	Name	Description	# of categories
1			
2			
3			
4			
5			
6			
Include additional information if necessary:			

3. If applicable, describe additional requirements for sub-national estimates (oversampling of specific groups of the population):

Figure A C.5. Sampling form 4

Sampling Form 4 — MOS and Sampling Frame

See chapter 5.2 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

1. This Sampling Form refers to: **ISCED Level 0.2** **U3**

2. Specify the measure of size (MOS) to be used.

Specify the measure of size (MOS) to be used. (Click in box and on right arrow to see drop down menu)

If "Other", please describe:

3. Specify the year for which enrollment data will be used for the MOS:

4. If a frame other than a single-level sampling frame (list of all centres) is to be used, please provide a preliminary description of the information available to construct this frame.

Figure A C.6. Sampling form 5

Sampling Form 5 Coordination of Studies

See chapter 5.3 of SOP Unit 1 - Sampling of Centres for more details

Participating country:

1. If you plan to participate in both target populations, how should the centre sample be selected? *(Click in box and on right arrow to see drop down menu)*

   How should the centre sample be selected?

2. If you plan to participate in the Starting Strong Survey and TALIS ISCED Level 1, please answer the following questions. (Click in box and on right arrow to see drop down menu)

   Are there centres/schools that belong to both studies?

   If yes, how should the centre/school sample be selected?

Annex D. Target and survey populations

Table A D.1. Target and survey populations: services for ISCED level 02 population

Participating country	Population and coverage	Reasons for exclusions	Number of centres	Number of staff
Chile	**Target population**		**9 941**	**62 919**
		Very small centres with only one teacher (and additional staff)	115	345
		Geographic location: Easter Island, Juan Fernández Islands	4	30
	Survey population		9 822	62 544
	Coverage after exclusions (%)		98.8	99.4
Denmark	**Target population**		**Unknown**	**Unknown**
		Partly private day care	56	Unknown
		Public afterschool care centre	Unknown	Unknown
	Survey population		3 488	Unknown
	Coverage after exclusions (%)		Unknown	99.5*
Germany	**Target population**		**63 913**	**Unknown**
		Home-based settings	6 768	Unknown
	Survey population		57 145	Unknown
	Coverage after exclusions (%)		89.4	Unknown
Iceland	**Target population**		**245**	**Unknown**
	Survey population		245	Unknown
	Coverage after exclusions (%)		100.0	100.0
Israel	**Target population**		**16 025**	**Unknown**
		Ultra-orthodox centres	3 821	Unknown
	Survey population		12 204	Unknown
	Coverage after exclusions (%)		76.2	Unknown
Japan	**Target population**		**36 589**	**247 691**
	Survey population		36 589	247 691
	Coverage after exclusions (%)		100.0	100.0
Korea	**Target population**		**26 218**	**126 523**
		Home-based settings	747	747
	Survey population		25 471	125 776
	Coverage after exclusions (%)		97.2	99.4

Participating country	Population and coverage	Reasons for exclusions	Number of centres	Number of staff
Norway	**Target population**		**5 500**	**41 162**
		Home-based settings	270	569
	Survey population		5 230	40 593
	Coverage after exclusions (%)		95.1	98.6
Turkey	**Target population**		**25 531**	**74 848**
		Under the authority of the Ministry of Family and Social Issues	2 134	8 097
	Survey population		23 397	66 751
	Coverage after exclusions (%)		91.6	89.2

Note: *The coverage rate for Denmark was based on children, with the assumption that the coverage rate of staff was similar.

Table A D.2. Target and survey populations: services for U3 population

(children under 3 years of age)

Participating country	Population and coverage	Reasons for exclusions	Number of centres	Number of staff
Denmark	**Target population**		**Unknown**	**Unknown**
		Private day care	Unknown	Unknown
		Partly private day care	54	Unknown
	Survey population		3 018	Unknown
	Coverage after exclusions (%)		Unknown	99.7*
Germany	**Target population**		**77 403**	**Unknown**
		Not listed home-based settings**	24 381	Unknown
	Survey population		53 022	Unknown
	Coverage after exclusions (%)		68.5	Unknown
Israel	**Target population**		**5 791**	**Unknown**
	Survey population		5 791	Unknown
	Coverage after exclusions (%)		100.0	100.0
Norway	**Target population**		**5 794**	**41 996**
		Home-based settings	618	1 795
	Survey population		5 176	40 201
	Coverage after exclusions (%)		89.3	95.7

Notes:
*The coverage rate for Denmark was based on children, with the assumption that the coverage rate of staff was similar.
**In Germany, a private company provided a list of centres that did not include all home-based settings. The given number is the difference between the number of listed home-based settings and the number of home-based settings according to national statistics.

Annex E. Characteristics of national samples

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

Chile: ISCED level 02

In Chile, ISCED level 02 consists of kindergartens, pre-schools and schools that offer pre-school education.

Table A E.1. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		9 941	62 919
Exclusions	Very small centres with only one teacher (and additional staff)	115	345
	Geographic location: Easter Island, Juan Fernández Islands	4	30
Total exclusions		119	375
Survey population		9 822	62 544
Coverage after exclusions		98.8%	99.4%

Note: Because the number of other staff per centre was unavailable, centres with one teacher were excluded regardless of the existence of other staff in the centre.

Sample design

Sample allocation: proportional to the number of staff.

Sample sizes: 50 centres selected for the field trial; 242 centres selected for the main survey.

Method of sample selection: systematic random sample with equal probabilities.

Measure of size: not applicable.

Explicit stratification: explicit stratification was organised by sources of funding (public school; private school with public funds; private school; INTEGRA kindergarten; JUNJI kindergarten; private kindergarten; public funds, delegated administration kindergarten), resulting in a total of seven explicit strata.

Implicit stratification: implicit stratification was organised by regions (Región de Arica y Parinacota, Región de Tarapacá, Región de Antofagasta, Región de Atacama, Región de Coquimbo, Región de Valparaíso, Región Metropolitana de Santiago, Región del Libertador General Bernardo O'Higgins, Región del Maule, Región del Bío-Bío, Región de la Araucanía, Región de Los Ríos, Región de Los Lagos, Región Aysén del General Carlos Ibáñez del Campo, Región de Magallanes y de la Antártica Chilena).

Table A E.2. Participation status of centres: staff and leader populations

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Public school	64		61	1		2
Private school with public funds	82	5	71	5	1	
Private school	20	1	16	1		2
INTEGRA kindergarten	18		18			
JUNJI kindergarten	14	2	12			
Private kindergarten	16	2	12	1	1	
Public funds, delegated administration kindergarten	28		28			
Total	242	10	218	8	2	4

Note: Combined table because staff and leader populations have the same numbers.

Table A E.3. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	503	493	5		5
Assistants	525	507	5	5	8
Staff for individual children	163	153	4	1	5
Staff for special tasks	214	195	10		9
Interns	3	3			
Total	1 408	1 351	24	6	27

Data sources

The centre frame was provided by the Ministry of Education, and the data used relate to 2015 and 2016.

Denmark: ISCED level 02

In Denmark, ISCED level 02 consists of kindergartens and integrated institutions.

Table A E.4. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		Unknown	Unknown
Exclusions	Partly private day care	56	Unknown
	Public afterschool care centre	Unknown	Unknown
Total exclusions		Unknown	Unknown
Survey population		3 488	Unknown
Coverage after exclusions		Unknown	99.5%

Note: The coverage rate is based on children, with the assumption that the coverage rate of staff was similar.

Sample design

Sample allocation: proportional to the number of children, then adjusted to have a minimum of eight centres selected in the smaller strata.

Sample sizes: 32 centres selected for the field trial; 202 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to centre size; overlap control between ISCED level 02 and services for children under the age of three years (U3) by selecting subsamples for each population of interest.

Measure of size: number of ISCED level 02 children.

Explicit stratification: explicit stratification was organised by overlap categories (ISCED level 02 only, ISCED level 02 and U3) and type of centre (kindergarten, integrated institution), resulting in a total of three explicit strata.

Implicit stratification: implicit stratification was organised by management type and source of funding (public, private, self-owned).

Table A E.5. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
ISCED level 02 only, kindergarten	60	2	19	8	2	29
ISCED level 02 only, integrated institution	8	4	1	1		2
ISCED level 02 and U3, integrated institution	134	10	41	9	4	70
Total	202	16	61	18	6	101

Table A E.6. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
ISCED level 02 only, kindergarten	60	2	25	9	4	20
ISCED level 02 only, integrated institution	8	4	1	1		2
ISCED level 02 and U3, integrated institution	134	10	50	8	4	62
Total	202	16	76	18	8	84

Table A E.7. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	582	407	161	5	9
Assistants	282	163	93	7	19
Staff for individual children	21	15	6		
Interns	24	10	7	5	2
Total	909	595	267	17	30

Data sources

The centre frame was provided by Statistics Denmark, and the data used relate to 2014.

Germany: ISCED level 02

In Germany, ISCED level 02 consists of kindergartens, school kindergartens, pre-school classes and mixed-age ECEC centres. Home-based settings are also available for children of this ISCED level.

Table A E.8. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		53 995	Unknown
Exclusions	Home-based settings	6 768	Unknown
Total exclusions		6 768	Unknown
Survey population		47 227	Unknown
Coverage after exclusions		87.5%	Unknown

Sample design

Sample allocation: proportional to the number of ISCED level 02 staff on average per stratum. Adjusted to have a minimum of 10 centres selected in the smaller strata.

Sample sizes: 66 centres selected for the field trial; 398 centres selected for the main survey. The sample size was increased to account for expected ineligible centres.

Method of sample selection: systematic random sample with equal probabilities; overlap control between ISCED level 02 and services for children under the age of three years by selecting subsamples for each population of interest.

Measure of size: not applicable.

Explicit stratification: explicit stratification was organised by type of centre (centre-based ECEC settings, that is, kindergartens and mixed-aged ECEC centres (*Kindertageseinrichtung*) and school kindergartens/pre-school classes (*Schulkindergarten/Vorklassen*), and regions (west, east), resulting in a total of three explicit strata.

Implicit stratification: no implicit stratification variables used.

Table A E.9. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Centre-based ECEC settings, west	292	33	147	29	13	70
Centre-based ECEC settings, east	96	7	43	12	3	31
School kindergartens/ pre-school classes, west	10		3			7
Total	398	40	193	41	16	108

Table A E.10. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
West	302	33	145	31	13	80
East	96	7	44	11	3	31
Total	398	40	189	42	16	111

Note: "Centre-based ECEC settings, west" and "School kindergartens/pre-school classes, west" were combined for the leader population.

Table A E.11. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 400	959	433	3	5
Assistants	563	455	105	3	
Staff for individual children	69	53	16		
Staff for special tasks	42	33	9		
Interns	113	89	24		
Unknown	39	29	10		
Total	2 226	1 618	597	6	5

Data sources

The centre frame was provided by a private list broker and IT service provider, and the data used relate to 2017. In the centre frame, there was no information about the level (ISCED level 02 or services for children under the age of three years). Following the centre sampling, centres were asked whether they belonged to the population of interest they were selected for. Centres not belonging to the relevant population of interest were counted as ineligible.

Iceland: ISCED level 02

In Iceland, ISCED level 02 consists of pre-schools.

Table A E.12. Coverage and exclusions

Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population	245	unknown
Total exclusions	0	0
Survey population	245	unknown
Coverage after exclusions	100.0%	100.0%

Sample design

Sample allocation: All centres were selected.

Sample sizes: 31 centres selected for the field trial; 245 centres selected for the main survey. Field trial centres were selected again for the main survey.

Method of sample selection: systematic random sampling with equal probabilities.

Measure of size: number of ISCED level 02 children.

Explicit stratification: explicit stratification was organised by regions (Reykjavík, other capital area, other urban, rural) and sources of funding (public, private), resulting in a total of eight explicit strata.

Implicit stratification: no implicit stratification variables used.

Table A E.13. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Reykjavík, public	63	1	52			10
Reykjavík, private	12	1	10			1
Other capital area, public	50		44			6
Other capital area, private	8		5			3
Other urban, public	33	2	30			1
Other urban, private	8		7			1
Rural, public	66	5	52			9
Rural, private	5		4			1
Total	245	9	204			32

Table A E.14. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Reykjavík, public	63	1	48			14
Reykjavík, private	12	1	8			3
Other capital area, public	50		37			13
Other capital area, private	8		2			6
Other urban, public	33	2	27			4
Other urban, private	8		7			1
Rural, public	66	5	45			16
Rural, private	5		4			1
Total	245	9	178			58

Table A E.15. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 036	871	157	5	3
Assistants	406	304	92	7	3
Staff for individual children	212	190	20		2
Staff for special tasks	39	29	10		
Total	1 693	1 394	279	12	8

Data sources

The centre frame was provided by Statistics Iceland, and the data used relate to 2015.

Israel: ISCED level 02

In Israel, ISCED level 02 consists of kindergartens.

Table A E.16. Coverage and exclusions

	Exclusions	ISCED level 0.2 centres	ISCED level 0.2 staff
Target population		16 025	Unknown
Exclusions	Ultra-orthodox centres	3 821	Unknown
Total exclusions		3 821	Unknown
Survey population		12 204	Unknown
Coverage after exclusions		76.2%	Unknown

Sample design

Sample allocation: proportional to the number of ISCED level 02 centres.

Sample sizes: 80 centres selected for the field trial; 424 centres selected for the main survey.

Method of sample selection: systematic random sample with equal probabilities.

Measure of size: not applicable.

Explicit stratification: explicit stratification was organised by language and orientation (Hebrew secular, Hebrew religious, Arabic Arab, Arabic Druze, Arabic Bedouin), resulting in a total of five explicit strata.

Implicit stratification: implicit stratification was organised by region (north, south) and socio-economic status (high, medium, low).

Table A E.17. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Hebrew secular	228		220			8
Hebrew religious	98		92			6
Arabic Arab	62	1	60	1		
Arabic Druze	10		9	1		
Arabic Bedouin	26		24	2		
Total	424	1	405	4		14

Table A E.18. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Hebrew secular	228		224			4
Hebrew religious	98		95			3
Arabic Arab	62	1	60	1		
Arabic Druze	10		9	1		
Arabic Bedouin	26		24	2		
Total	424	1	412	4		7

Table A E.19. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	998	920	65	6	7
Assistants	1 193	1 061	103	12	17
Staff for individual children	263	234	22	4	3
Interns	55	53	2		
Total	2 509	2 268	192	22	27

Data sources

The centre frame was provided by the Ministry of Education, and the data used relate to 2017.

Japan: ISCED level 02

In Japan, ISCED level 02 consists of kindergartens, nursery centres and centres for early childhood and care.

Table A E.20. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		36 589	247 691
Total exclusions		0	0
Survey population		36 589	247 691
Coverage after exclusions		100.0%	100.0%

Sample design

Sample allocation: disproportional allocation to follow the country's request to select a specific number of centres per stratum to support comparisons across the strata.

Sample sizes: 30 centres selected for the field trial; 220 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to number of ISCED level 02 staff.

Measure of size: number of ISCED level 02 staff.

Explicit stratification: explicit stratification was organised by type of centre (kindergarten, nursery centre, centre for ECEC) and sources of funding (public, private), resulting in a total of six explicit strata.

Implicit stratification: implicit stratification was organised by urbanisation (major city, no major city).

Table A E.21. Participation status of centres: staff and leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Kindergarten, public	40	1	39			
Kindergarten, private	40		32	6	2	
Nursery centre, public	40		36	2	1	1
Nursery centre, private	40		33	5	1	1
Centre for ECEC, public	30		29	1		
Centre for ECEC, private	30		22	6	1	1
Total	220	1	191	20	5	3

Note: Combined table because staff and leader populations have the same numbers.

Table A E.22. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 646	1 616	7	15	8
Total	1 646	1 616	7	15	8

Data sources

The centre frame was provided by the Ministry of Education, Culture, Sports, Science and Technology (kindergartens), by the Ministry of Health, Labour and Welfare (nursery centres) and by the Cabinet Office (centres for ECEC). The data used relate to 2015 (kindergartens) and to 2016 (nursery centres and centres for ECEC).

Korea: ISCED level 02

In Korea, ISCED level 02 consists of kindergartens and childcare centres. Home-based settings are also available for children of this ISCED level.

Table A E.23. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		26 218	126 523
Exclusions	Home-based settings	747	747
Total exclusions		747	747
Survey population		25 471	125 776
Coverage after exclusions		97.2%	99.4%

Sample design

Sample allocation: proportional to the number of ISCED level 02 staff, then adjusted to have a minimum of eight centres selected in the smaller strata.

Sample sizes: 42 centres selected for the field trial; 258 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to number of ISCED level 02 staff.

Measure of size: number of ISCED level 02 staff.

Explicit stratification: explicit stratification was organised by type of centre (kindergarten, childcare centre), sources of funding (public, private) and levels of urbanisation (large city, small or middle city, large city), resulting in a total of 12 explicit strata.

Implicit stratification: no implicit stratification variables used.

Table A E.24. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Kindergarten, public, large city	8		7	1		
Kindergarten, public, small or middle city	10		8	1		1
Kindergarten, public, rural area	10		7			3
Kindergarten, private, large city	36	2	19	6	1	8
Kindergarten, private, small or middle city	32		17	9		6
Kindergarten, private, rural area	8		5	2		1
Childcare centre, public, large city	24		17	3	1	3
Childcare centre, public, small or middle city	14		10	3		1
Childcare centre, public, rural area	16		9	1	1	5
Childcare centre, private, large city	38	4	15	1	1	17
Childcare centre, private, small or middle city	38	3	21	5		9
Childcare centre, private, rural area	24	2	11			11
Total	258	11	146	32	4	65

Table A E.25. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Kindergarten, public, large city	8		7	1		
Kindergarten, public, small or middle city	10		8	1		1
Kindergarten, public, rural area	10		7			3
Kindergarten, private, large city	36	2	20	6	1	7
Kindergarten, private, small or middle city	32		18	9		5
Kindergarten, private, rural area	8		5	2		1
Childcare centre, public, large city	24		17	3	1	3
Childcare centre, public, small or middle city	14		10	3		1
Childcare centre, public, rural area	16		9	1	1	5
Childcare centre, private, large city	38	4	18	1		15
Childcare centre, private, small or middle city	38	3	21	6		8
Childcare centre, private, rural area	24	2	12			10
Total	258	11	152	33	3	59

Table A E.26. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 015	842	160	7	6
Assistants	90	71	15	2	2
Staff for individual children	18	18			
Staff for special tasks	5	5			
Total	1 128	936	175	9	8

Data sources

The centre frame was provided by the Korea Educational Development Institute and the Ministry of Health and Welfare, and the data used relate to 2016.

Norway: ISCED level 02

In Norway, ISCED level 02 consists of kindergartens and family kindergartens. Home-based settings are also available for children of this ISCED level.

Table A E.27. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		5 500	41 162
Exclusions	Home-based settings	270	569
Total exclusions		270	569
Survey population		5 230	40 593
Coverage after exclusions		95.1%	98.6%

Sample design

Sample allocation: proportional to the number of staff, then adjusted to have a minimum of eight centres selected in the smaller strata.

Sample sizes: 32 centres selected for the field trial; 188 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to number of staff; overlap control between ISCED level 02 and services for children under the age of three years by selecting subsamples for each population of interest.

Measure of size: number of staff (ISCED level 02 and services for children under the age of three years (U3)).

Explicit stratification: explicit stratification was organised by overlap categories (ISCED level 02 only, ISCED level 02 and U3), regions (Oslo/Akershus, East Norway, South/West Norway, Middle/North Norway) and sources of funding (public, private), resulting in a total of nine explicit strata.

Implicit stratification: implicit stratification was organised by urbanisation (high level of urbanisation, medium level of urbanisation, low level of urbanisation) and region (Oslo/Akershus, East Norway, South/West Norway, Middle/North Norway).

Table A E.28. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
ISCED level 02 only, all, all	8		6	1		1
ISCED level 02 and U3, Oslo/ Akershus, public	20		8	3	2	7
ISCED level 02 and U3, Oslo/ Akershus, private	22		15	3		4
ISCED level 02 and U3, East Norway, public	22		11	4		7
ISCED level 02 and U3, East Norway, private	22	2	12		2	6
ISCED level 02 and U3, South/ West Norway, public	32	2	18	3	3	6
ISCED level 02 and U3, South/ West Norway, private	30		21	4	1	4
ISCED level 02 and U3, Middle/ North Norway, public	12		7	4		1
ISCED level 02 and U3, Middle/ North Norway, private	20	1	13	3		3
Total	188	5	111	25	8	39

Table A E.29. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
ISCED level 02 only, all, all	8		6	1		1
ISCED level 02 and U3, Oslo/ Akershus, public	20		10	3	2	5
ISCED level 02 and U3, Oslo/ Akershus, private	22		15	3		4
ISCED level 02 and U3, East Norway, public	22		13	4		5
ISCED level 02 and U3, East Norway, private	22	2	13		2	5
ISCED level 02 and U3, South/ West Norway, public	32	2	20	3	3	4
ISCED level 02 and U3, South/ West Norway, private	30		23	4	1	2
ISCED level 02 and U3, Middle/ North Norway, public	12		7	4		1
ISCED level 02 and U3, Middle/ North Norway, private	20	1	12	3		4
Total	188	5	119	25	8	31

Table A E.30. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	449	379	62	7	1
Assistants	582	420	155	7	
Staff for individual children	43	37	6		
Interns	15	10	5		
Total	1 089	846	228	14	1

Data sources

The centre frame was provided by the Norwegian Directorate for Education and Training, and the data used relate to 2015.

Turkey: ISCED level 02

In Turkey, ISCED level 02 consists of pre-schools, kindergarten classrooms and practice classrooms.

Table A E.31. Coverage and exclusions

	Exclusions	ISCED level 02 centres	ISCED level 02 staff
Target population		25 531	74 848
Exclusions	Under the authority of the Ministry of Family and Social Issues	2 134	8 097
Total exclusions		2 134	8 097
Survey population		23 397	66 751
Coverage after exclusions		91.6%	89.2%

Sample design

Sample allocation: proportional to the number of staff, then adjusted to have a minimum of six centres selected in the smaller strata.

Sample sizes: 56 centres selected for the field trial; 354 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to number of staff.

Measure of size: number of ISCED level 02 staff.

Explicit stratification: explicit stratification was organised by sources of funding (public, private), type of centre (pre-school (*anaokulu*), kindergarten classroom (*anasınıfı*), practice classroom (*uygulama sınıfları*)) and regions (Marmara, Central Anatolia, Aegean, Black Sea, Mediterranean, East Anatolia, South Eastern Anatolia), resulting in a total of 28 explicit strata.

Implicit stratification: no implicit stratification variables used.

Table A E.32. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Public, pre-school, Marmara	14		14			
Public, pre-school, Central Anatolia	16	1	15			
Public, pre-school, Aegean	10		10			
Public, pre-school, Black Sea	10		10			
Public, pre-school, Mediterranean	14		14			
Public, pre-school, East Anatolia	8		8			
Public, pre-school, South Eastern Anatolia	12		12			
Public, kindergarten classroom, Marmara	34	1	33			
Public, kindergarten classroom, Central Anatolia	26	1	25			
Public, kindergarten classroom, Aegean	20	1	19			
Public, kindergarten classroom, Black Sea	14	1	13			
Public, kindergarten classroom, Mediterranean	22	1	21			
Public, kindergarten classroom, East Anatolia	14	1	13			
Public, kindergarten classroom, South Eastern Anatolia	20		19			1

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Public, practice classroom, Marmara	6	1	5			
Public, practice classroom, Central Anatolia	6		6			
Public, practice classroom, Aegean	6		6			
Public, practice classroom, Black Sea	6		6			
Public, practice classroom, Mediterranean	6		6			
Public, practice classroom, East Anatolia	6		6			
Public, practice classroom, South Eastern Anatolia	6	1	5			
Private, all types, Marmara	34	1	33			
Private, all types, Central Anatolia	10	2	8			
Private, all types, Aegean	10		10			
Private, all types, Black Sea	6		6			
Private, all types, Mediterranean	6		6			
Private, all types, East Anatolia	6	1	5			
Private, all types, South Eastern Anatolia	6		6			
Total	354	13	340			1

Table A E.33. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Public, pre-school, Marmara	14		14			
Public, pre-school, Central Anatolia	16	1	15			
Public, pre-school, Aegean	10		10			
Public, pre-school, Black Sea	10		10			
Public, pre-school, Mediterranean	14		14			
Public, pre-school, East Anatolia	8		8			
Public, pre-school, South Eastern Anatolia	12		12			
Public, kindergarten classroom, Marmara	34	1	33			
Public, kindergarten classroom, Central Anatolia	26	1	25			
Public, kindergarten classroom, Aegean	20	1	19			
Public, kindergarten classroom, Black Sea	14	1	13			
Public, kindergarten classroom, Mediterranean	22	1	21			
Public, kindergarten classroom, East Anatolia	14	1	13			
Public, kindergarten classroom, South Eastern Anatolia	20		20			
Public, practice classroom, Marmara	6	1	5			

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Public, practice classroom, Central Anatolia	6		6			
Public, practice classroom, Aegean	6		6			
Public, practice classroom, Black Sea	6		6			
Public, practice classroom, Mediterranean	6		6			
Public, practice classroom, East Anatolia	6		6			
Public, practice classroom, South Eastern Anatolia	6	1	5			
Private, all types, Marmara	34	1	33			
Private, all types, Central Anatolia	10	2	7			1
Private, all types, Aegean	10		10			
Private, all types, Black Sea	6		6			
Private, all types, Mediterranean	6		6			
Private, all types, East Anatolia	6	1	5			
Private, all types, South Eastern Anatolia	6		6			
Total	354	13	340			1

Table A E.34. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 621	1 605	14	2	
Total	1 621	1 605	14	2	

Note: Staff responsible for mainly non-pedagogical tasks (including changing and accompanying the child to the toilet) did not fulfil the definition of staff and were therefore not listed.

Data sources

The centre frame was provided by the Ministry of National Education, and the data used relate to 2016.

Denmark: services for children under the age of three years

In Denmark, services for children under the age of three years (U3) consist of nurseries, day care facilities and integrated institutions. Home-based settings are also available for children of this age.

Table A E.35. Coverage and exclusions

	Exclusions	U3 centres	U3 staff
Target population		Unknown	Unknown
• centre-based settings		Unknown	Unknown
• home-based settings		540	Unknown
Exclusions	Private day care	Unknown	Unknown
	Partly private day care	54	Unknown
Total exclusions		Unknown	Unknown
Survey population		3 018	Unknown
Coverage after exclusions		Unknown	99.7%

Note: The coverage rate is based on children, on the assumption that the coverage rate of staff was similar.

Sample design

Sample allocation: upon request, day cares were oversampled to have 100 centres. The smaller strata had a minimum of eight centres selected.

Sample sizes: 32 centres selected for the field trial; 222 centres selected for the main survey.

Method of sample selection: systematic sample with probabilities proportional to centre size; overlap control between ISCED level 02 and services for children under the age of three years by selecting subsamples for each population of interest.

Measure of size: number of children under the age of three years.

Explicit stratification: explicit stratification was organised by overlap categories (U3 only, ISCED level 02 and U3) and type of centre (day care, nursery, integrated institution), resulting in a total of four explicit strata.

Implicit stratification: implicit stratification was organised by management type and source of funding (public, private, self-owned).

Table A E.36. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
U3 only, day care	100	16	25	6	5	48
U3 only, nursery	14	2	2	1		9
U3 only, integrated institution	8		4			4
ISCED level 02 and U3, integrated institution	100	5	32	7	5	51
Total	222	23	63	14	10	112

Table A E.37. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
U3 only, day care	100	16	27	6	5	46
U3 only, nursery	14	2	4	1		7
U3 only, integrated institution	8		4			4
ISCED level 02 and U3, integrated institution	100	5	33	7	6	49
Total	222	23	68	14	11	106

Table A E.38. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	634	467	129	9	29
Assistants	193	112	65	9	7
Staff for individual children	8	3	3		2
Staff for special tasks	1	1			
Interns	14	7	2	2	3
Total	850	590	199	20	41

Data sources

The centre frame was provided by Statistics Denmark, and the data on nurseries and integrated institutions relate to 2014. Data about day cares were collected from the municipalities during the preparation of the sampling frame.

Germany: services for children under the age of three years

In Germany, services for children under the age of three years (U3) consist of day nurseries and mixed-age ECEC centres. Home-based settings are also available for children of this age.

Table A E.39. Coverage and exclusions

	Exclusions	U3 centres	U3 staff
Target population		77 403	unknown
• centre-based settings		45 315	unknown
• home-based settings		32 088	32 088
Exclusions	Not listed home-based settings*	24 381	24 381
Total exclusions		24 381	24 381
Survey population		53 022	unknown
Coverage after exclusions		68.5%	unknown

Note: *A private company provided a list of centres that did not include all home-based settings. The given number is the difference between the number of listed home-based settings and the number of home-based settings according to national statistics.

Sample design

Sample allocation: proportional to the number of staff in services for children under the age of three years on average per stratum.

Sample sizes: 70 centres selected for the field trial; 682 centres selected for the main survey. The sample size was increased to account for expected ineligible centres.

Method of sample selection: systematic random sample with equal probabilities; overlap control between ISCED level 02 and services for children under the age of three years by selecting subsamples for each population of interest.

Measure of size: not applicable.

Explicit stratification: explicit stratification was organised by type of centre (centre-based ECEC settings (*Kindertageseinrichtung*), home-based ECEC settings (*Kindertagespflege*)) and regions (west, east), resulting in a total of four explicit strata.

Implicit stratification: no implicit stratification variables used.

Table A E.40. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Centre-based ECEC settings, west	408	120	142	13	5	128
Centre-based ECEC settings, east	136	28	52	8	2	46
Home-based ECEC settings, west	96	40	31	2	1	22
Home-based ECEC settings, east	42	17	15		1	9
Total	682	205	240	23	9	205

Table A E.41. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres
Centre-based ECEC settings, west	408	120	143	13	4	128
Centre-based ECEC settings, east	136	28	53	8	2	45
Home-based ECEC settings, west	96	40	31	2	1	22
Home-based ECEC settings, east	42	17	15		1	9
Total	682	205	242	23	8	204

Table A E.42. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 387	912	466	1	8
Assistants	414	359	55		
Staff for individual children	40	33	7		
Staff for special tasks	24	20	4		
Interns	81	64	17		
Unknown	27	24	3		
Total	1 973	1 412	552	1	8

Data sources

The centre frame was provided by a private list broker and information technology service provider, and the data used relate to 2017. The list did not include all home-based settings. According to national statistics, the number of *listed* home-based settings and the number of *actual* home-based settings differ.

In the centre frame, there was no information about the level (ISCED level 02 or services for children under the age of three years). Following the centre sampling, centre-based settings were asked whether they belonged to the population of interest they were selected for, and home-based settings were asked whether they have children of the respective age. Centre-based and home-based settings not belonging to the relevant population of interest were counted as ineligible.

Israel: services for children under the age of three years

In Israel, services for children under the age of three years (U3) consist of day care centres and family day cares. Home-based settings are also available for children of this age.

Table A E.43. Coverage and exclusions

	Exclusions	U3 centres	U3 staff
Target population		5 791	unknown
• centre-based settings		2 211	unknown
• home-based settings		3 580	3 580
Total exclusions		0	0
Survey population		5 791	unknown
Coverage after exclusions		100.0%	100.0%

Sample design

Sample allocation: disproportional allocation to follow the country's request to have the same number of centres selected in every region.

Sample sizes: 64 centres selected for the field trial; 254 centres selected for the main survey.

Availability of frame data: because complete frame data were available for centre-based settings, a two-stage sampling design was applied for these centres (see Chapter 9 of this report for a description). For home-based settings, a complete list of co-ordinators was available. Because home-based settings are assigned to co-ordinators, a three-stage design was implemented in which home-based settings were listed for selected co-ordinators.

Method of sample selection: centre-based settings: systematic sample with probabilities proportional to number of children under the age of three years. Home-based settings: three-stage design with co-ordinators as primary sampling units, selected with systematic sampling with probabilities proportional to number of centres. As secondary units, centres were selected with systematic random sampling with equal probabilities.

Measure of size: centre-based settings: number of children under the age of three years. Home-based settings/first level: number of centres for children under the age of three years. Home-based settings/second level: number of staff in services for children under the age of three years.

Explicit stratification: explicit stratification was organised by type of centre (day care centre, family day care) and regions (north, centre, south, Jerusalem), resulting in a total of eight explicit strata.

Implicit stratification: implicit stratification was organised by language (Hebrew, Arab, unknown).

Table A E.44. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres	Excluded centres
Day care centre, north	44	3	38	2	1		
Day care centre, centre	44	2	30	2	1	9	
Day care centre, south	44	6	35	1	1	1	
Day care centre, Jerusalem	44		41	2		1	
Family day care, north	20		20				
Family day care, centre	20		19	1			
Family day care, south	19	1	18				
Family day care, Jerusalem	19	4	13				2
Total	254	16	214	8	3	11	2

Table A E.45. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres	Excluded centres
Day care centre, north	44	3	38	2	1		
Day care centre, centre	44	2	32	2	1	7	
Day care centre, south	44	6	35	1	1	1	
Day care centre, Jerusalem	44		41	1		2	
Family day care, north	20		20				
Family day care, centre	20		19	1			
Family day care, south	19	1	18				
Family day care, Jerusalem	19	4	13				2
Total	254	16	216	7	3	10	2

Table A E.46. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	1 208	1 096	44	43	25
Assistants	8	8			
Staff for individual children	21	19	1		1
Total	1 237	1 123	45	43	26

Data sources

The centre frame was provided by the Ministry of Labour, and the data used relate to 2017. Data about home-based settings overseen by the selected co-ordinators were collected in 2017.

Norway: services for children under the age of three years

In Norway, services for children under the age of three years (U3) consist of kindergartens and family kindergartens. Home-based settings are also available for children of this age.

Table A E.47. Coverage and exclusions

	Exclusions	U3 centres	U3 staff
Target population		5 794	41 996
• centre-based settings		5 176	40 201
• home-based settings		618	1 795
Exclusions	Home-based settings	618	1 795
Total exclusions		618	1 795
Survey population		5 176	40 201
Coverage after exclusions		89.3%	95.7%

Sample design

Sample allocation: proportional to the number of staff, then adjusted to have a minimum of eight centres selected in the smaller strata.

Sample sizes: 30 centres selected for the field trial; 188 centres selected for the main survey.

Method of sample selection: systematic sample proportional to number of staff; overlap control between ISCED level 02 and services for children under the age of three years by selecting subsamples for each population of interest.

Measure of size: number of staff (ISCED level 02 and services for children under the age of three years).

Explicit stratification: explicit stratification was organised by overlap categories (U3 only, ISCED level 02 and U3), regions (Oslo/Akershus, East Norway, South/West Norway, Middle/North Norway) and sources of funding (public, private), resulting in a total of nine explicit strata.

Implicit stratification: implicit stratification was organised by urbanisation (high level of urbanisation, medium level of urbanisation, low level of urbanisation) and region (Oslo/Akershus, East Norway, South/West Norway, Middle/North Norway).

Table A E.48. Participation status of centres: staff population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres	Excluded centres
U3 only, all, all	8	3	3			2	
ISCED level 02 and U3, Oslo/ Akershus, public	20	1	12	5	1	1	
ISCED level 02 and U3, Oslo/ Akershus, private	22		14	4	1	3	
ISCED level 02 and U3, East Norway, public	22	1	13	6		1	1
ISCED level 02 and U3, East Norway, private	22		15	5	1	1	
ISCED level 02 and U3, South/ West Norway, public	32		20	7		5	
ISCED level 02 and U3, South/ West Norway, private	30		19	5	4	2	
ISCED level 02 and U3, Middle/ North Norway, public	12	1	7	3		1	
ISCED level 02 and U3, Middle/ North Norway, private	20	3	14	2		1	
Total	188	9	117	37	7	17	1

Table A E.49. Participation status of centres: leader population

Explicit strata	Total sampled centres	Ineligible centres	Participating sampled centres	Participating first replacements	Participating second replacements	Non-participating centres	Excluded centres
U3 only, all, all	8	3	3			2	
ISCED level 02 and U3, Oslo/ Akershus, public	20	1	12	5	1	1	
ISCED level 02 and U3, Oslo/ Akershus, private	22		15	5	1	1	
ISCED level 02 and U3, East Norway, public	22	1	12	5		3	1
ISCED level 02 and U3, East Norway, private	22		15	5	1	1	
ISCED level 02 and U3, South/ West Norway, public	32		22	8	1	1	
ISCED level 02 and U3, South/ West Norway, private	30		20	5	4	1	
ISCED level 02 and U3, Middle/ North Norway, public	12	1	7	3		1	
ISCED level 02 and U3, Middle/ North Norway, private	20	3	11	2		4	
Total	188	9	117	38	8	15	1

Table A E.50. Participation status of staff

Staff category	Total sampled staff	Participated	Absent	Left centre permanently	Not part of sample
Teachers	469	408	57	4	
Assistants	659	521	127	11	
Staff for individual children	17	16	1		
Interns	9	7	2		
Total	1 154	952	187	15	

Data sources

The centre sampling frame was provided by the Norwegian Directorate for Education and Training, and the data used relate to 2015.

Annex F. Listing and tracking forms

OECD International ECEC Staff Survey MS - [ISCED Level X] Listing Form

Country/Region	
Centre Name	
Centre ID	

❶	❷	❷	❸	❹	❺	❻
Name	**Sequence Number**	**Sequence Number**	**Leader Role**	**Staff Role**	**Year of Birth**	**Gender**

③ **Leader Role:** 1 = Leader of this centre
④ **Staff Role**: 1 = <Only centre leader (no pedagogical work)>; 2 = <Teacher>; 3 = <Assistant>; 4 = <Staff for individual children>; 5 = <Staff for special tasks>; 6 = <Intern>; 7 = <country-specific>; 8 = <country-specific>; 9 = <country-specific>; 10 = <country-specific>; 11 = <country-specific>; 12 = <country-specific>
⑤ **Year of Birth:** YYYY
⑥ **Gender:** 1 = Female; 2 = Male

OECD International ECEC Staff Survey MS - [ISCED Level X] Tracking Form

Country/Region	<_flh_country_>
Centre Name	<_flh_school_name_>
Centre ID	<_flh_school_id_>

❶	❷	❷	❸	❹	❺	❻	❼
Name	**ID**	**ID**	**Year Of Birth**	**Gender**	**Questionnaire Mode**	**Questionnaire Return Status**	**Questionnaire**
<_fld_teacher_name_>	er_ID_>	er_ID2_>	<_fld_yob_>	nder_>	est_mode_>	est_ret_>	uest_>

③ **Year of Birth:** YYYY
④ **Gender**:
1 = Female; 2 = Male; 9 = Not specified
⑤ **Questionnaire** Mode:
1 = Online; 2 = Paper
⑥ **Questionnaire Return Status:**
1 = Returned (paper); 2 = Returned (online); 3 = Not returned; 4 = Left centre permanently; 5 = Person should have been out-of-scope
⑦ **Questionnaire:**
10 = Leader Questionnaire; 20 = Staff Questionnaire; 30 = Centre Questionnaire; 40 = Centre Questionnaire Plus

Annex G. Participation and estimated population sizes

Table A G.1. Participation and estimated size of staff population, ISCED level 02, by participating country

Participating country	Number of participating centres	Responding staff in participating centres	Estimated size of staff population
Chile	228	1 349	58 060
Germany	250	1 401	404 202
Iceland	204	1 378	3 624
Israel	409	1 987	43 478
Japan	216	1 616	307 070
Korea	182	927	91 586
Norway	144	815	39 107
Turkey	340	1 605	65 191
Denmark*	85	544	28 831

Notes:
The estimated population size does not include the exclusions. Annex D provides an overview of the exclusions.
*Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

Table A G.2. Unweighted and weighted participation rates of staff population, ISCED level 02, by participating country

Participating country	Weighting	Centre participation before replacements (%)	Centre participation after replacements (%)	Staff participation in participating centres (%)	Overall participation (%)
Chile	Unweighted	94.0	98.3	98.8	97.1
	Weighted	94.8	98.2	98.7	96.9
Denmark	Unweighted	32.8	45.7	84.3	38.5
	Weighted	32.2	45.4	84.9	38.6
Germany	Unweighted	53.9	69.8	84.8	59.2
	Weighted	54.0	71.1	85.0	60.4
Iceland	Unweighted	86.4	86.4	85.2	73.6
	Weighted	87.0	87.0	84.8	73.8
Israel	Unweighted	95.7	96.7	97.4	94.1
	Weighted	96.0	96.6	96.3	93.1
Japan	Unweighted	87.2	98.6	99.6	98.2
	Weighted	83.0	98.2	99.6	97.9
Korea	Unweighted	59.1	73.7	96.4	71.0
	Weighted	59.7	75.4	96.8	72.9
Norway	Unweighted	60.7	78.7	83.8	65.9
	Weighted	58.0	78.4	83.4	65.4
Turkey	Unweighted	99.7	99.7	99.2	98.9
	Weighted	99.7	99.7	99.6	99.2

Table A G.3. Participation and estimated size of staff population, U3, by participating country

Participating country	Number of participating centres	Responding staff in participating centres	Estimated size of staff population
Germany	272	1 171	268 310
Israel	225	1 113	23 201
Norway	161	938	35 514
Denmark*	87	563	28 303

Notes:
The estimated population size does not include the exclusions. Annex D provides an overview of the exclusions.
*Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

Table A G.4. Unweighted and weighted participation rates of staff population, U3, by participating country

Participating country	Weighting	Centre participation before replacements (%)	Centre participation after replacements (%)	Staff participation in participating centres (%)	Overall participation (%)
Denmark	Unweighted	31.7	43.7	86.1	37.6
	Weighted	30.9	43.6	86.2	37.6
Germany	Unweighted	50.3	57.0	89.7	51.1
	Weighted	49.3	56.2	89.6	50.4
Israel	Unweighted	90.7	95.3	97.4	92.8
	Weighted	86.8	92.2	97.7	90.1
Norway	Unweighted	65.7	90.4	86.9	78.6
	Weighted	67.1	91.1	86.5	78.8

Table A G.5. Participation and estimated size of leader population, ISCED level 02, by participating country

Participating country	Number of participating centre leaders	Estimated size of leader population
Chile	228	9 426
Germany	247	51 942
Iceland	178	236
Israel	416	12 175
Japan	216	35 577
Korea	188	22 722
Norway	152	4 877
Turkey	340	22 380
Denmark*	102	3 034

Notes:
The estimated population size does not include the exclusions. Annex D provides an overview of the exclusions.
*Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

Table A G.6. Unweighted and weighted participation rates of leader population, ISCED level 02, by participating country

Participating country	Weighting	Participation before replacements (%)	Participation after replacements (%)
Chile	Unweighted	94.0	98.3
	Weighted	94.8	98.6
Denmark	Unweighted	40.9	54.8
	Weighted	40.9	55.3
Germany	Unweighted	52.8	69.0
	Weighted	52.8	69.0
Iceland	Unweighted	75.4	75.4
	Weighted	75.4	75.4
Israel	Unweighted	97.4	98.3
	Weighted	97.4	98.3
Japan	Unweighted	87.2	98.6
	Weighted	83.6	98.2
Korea	Unweighted	61.5	76.1
	Weighted	63.3	74.2
Norway	Unweighted	65.0	83.1
	Weighted	64.6	83.8
Turkey	Unweighted	99.7	99.7
	Weighted	99.8	99.8

Table A G.7. Participation and estimated size of leader population, U3, by participating country

Participating country	Number of participating centre leaders	Estimated size of leader population
Germany	273	48 699
Israel	226	5 042
Norway	163	4 916
Denmark*	93	2 852

Notes:
The estimated population size does not include the exclusions. Annex D provides an overview of the exclusions.
*Low response rates in the survey may result in bias in the estimates reported and limit comparability of the data.

Table A G.8. Unweighted and weighted participation rates of leader population, U3, by participating country

Participating country	Weighting	Participation before replacements (%)	Participation after replacements (%)
Denmark	Unweighted	34.2	46.7
	Weighted	35.4	47.5
Germany	Unweighted	50.7	57.2
	Weighted	50.6	57.2
Israel	Unweighted	91.5	95.8
	Weighted	93.3	97.4
Norway	Unweighted	65.7	91.6
	Weighted	66.8	92.6

Annex H. Questionnaires

You can find the TALIS Starting Strong 2018 questionnaires on line at:
http://www.oecd.org/education/school/talisstartingstrongsurvey2018questionnaires.htm

Annex I. Validation of scales and construction of scale scores

Table A I.1. Descriptive statistics for the scaled scores (based on all populations, unweighted)

	Mean	Standard deviation	Lowest score	Highest score
Leader scales				
S1LDEV	10.42	2.17	3.27	15.43
S1LDIVB	13.02	1.99	1.64	16.00
S1LEADP	10.92	1.92	0.64	15.86
S1LEADS	10.99	1.98	3.74	16.98
S1LNIGHB	12.17	2.01	0.77	15.60
S1LPANE	8.70	2.05	1.17	22.09
S1LPART	10.32	2.16	3.91	17.88
Staff scales				
S1ACTIV	12.53	2.10	0.66	17.46
S1COLL	10.76	1.97	4.51	14.99
S1JOBSA	11.42	1.94	1.69	16.08
S1PAREN	12.58	1.92	2.37	15.99
S1PDIV	10.63	1.97	3.71	15.47
S1PDNET	8.95	1.93	5.60	14.29
S1PDPE	9.67	2.01	4.93	15.45
S1PRACT	11.29	1.96	2.98	15.04
S1PRDEV	15.57	1.98	-1.18	19.19
S1PREM	14.56	1.99	1.26	17.44
S1PRINT	13.23	1.94	-2.63	17.43
S1PRLAN	14.38	1.99	-5.77	17.77
S1PROLD	13.65	2.00	-6.91	18.06
S1PROLI	11.69	1.95	0.51	15.67
S1PRONU	12.25	2.01	-11.23	16.63
S1PRSOC	14.55	1.99	-1.58	17.67
S1SOCIAL	13.87	1.96	3.98	17.17
S1STRES	9.29	2.01	4.02	15.52

Note: The descriptive statistics were computed for unweighted scores because they are meant to illustrate the distribution of the scale scores in the sample and not describe the study's target populations. The U3 population in Denmark was excluded from the descriptive statistics for scale S1LDIVB because of untrustworthy scale scores.
Source: OECD, TALIS Starting Strong 2018 database.

Table A I.2. Country-specific invariance levels across countries that participated in both populations (ISCED level 02 and U3): Leader scales

Participating country	S1LDEV	S1LDIVB	S1LEADP	S1LEADS	S1LNIGHB	S1LPANE	S1LPART
Germany	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	-
Israel	Scalar	-	Configural	Configural	Scalar	-	Metric
Norway	Scalar	Scalar	Scalar	Scalar	Scalar	-	-

Note: Missing values became evident when one of the two surveyed populations was dropped from the scale.
See Table 11.6, "Excluded populations from the estimation of scales' international parameters".
Source: OECD, TALIS Starting Strong 2018 database.

Table A I.3. Country-specific invariance levels across countries that participated in both populations (ISCED level 02 and U3): Staff scales

Participating country	S1ACTIV	S1COLL	S1JOBSA	S1PAREN	S1PDIV	S1PDNET	S1PDPE	S1PRACT	S1PRDEV	S1PREM	S1PRINT	S1PRLAN	S1PROLD	S1PROLI	S1PRONU	S1PRSOC	S1SOCIAL	S1STRES
Germany	Metric	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	Scalar	-	Scalar	Metric	Metric	Metric	Scalar
Israel	Metric	Scalar	Metric	Metric	Scalar	Scalar	Metric	Metric	Metric	Metric	Scalar	Metric	Metric	Configural	Metric	Scalar	Configural	-
Norway	Metric	Scalar	Scalar	Scalar	Scalar	Configural	Scalar	Scalar	Scalar	Metric	Scalar	Scalar	Metric	Configural	Metric	Scalar	Configural	Scalar

Note: Missing values became evident when one of the two surveyed populations was dropped from the scale.
See Table 11.6, "Excluded populations from the estimation of scales' international parameters".
Source: OECD, TALIS Starting Strong 2018 database.

Annex J. Table production and verification, analysis

1.1. Recoded variables that are part of the TALIS Starting Strong 2018 public dataset[1]

The first part of the codebook describes the recoding of the variables created for and delivered in the TALIS Starting Strong 2018 public dataset. Variables are ordered (a) by population (first the centre leader, second centre staff) and (b) numerically by the original variable name.

Recodings based on leader data

Variable name	**LAGEGR**		
Description	Leader age groups		
Procedure	Simple recode		
Source	How old are you?	SL1G02	Recoding 1 = Under 50 2 = 50 and above

Variable name	**LISCED**		
Description	Leader education groups		
Procedure	Simple recode		
Source	What is the highest level of formal education you have completed?	SL1G03 1 = Below <ISCED 2011 Level 3> 2 = <ISCED 2011 Level 3> 3 = <ISCED 2011 Level 4> 4 = <ISCED 2011 Level 5> 5 = <ISCED 2011 Level 64> 6 = <ISCED 2011 Level 65> 7 = <ISCED 2011 Level 66> 8 = <ISCED 2011 Level 7> 9 = <ISCED 2011 Level 8>	Recoding 1 = <ISCED 2011 Level 5> and below 2 = <ISCED 2011 Level 6> and above

Variable name	**LEXPGR**		
Description	Leader experience groups		
Procedure	Simple recode		
Source	Years working as an <ECEC centre leader> in total	SL1G05B	Recoding 1 = Under 5 years 2 = 5–9 years 3 = 10 years or more

Variable name	SL1G14CD		
Description	ECEC centres that receive funding from non-governmental organisations, benefactors or both		
Procedure	Multivariable recode		
Missing values	Missing if both variables are missing or if at least one variable is missing at the other has a response of "no" or "don't know". Not administered does not count as missing, only if all variables are not administered.		
Source	In the last 12 months, which of the following sources provided funding for this <ECEC centre>? c) Non-governmental organisations (including religious institutions and employers) d) Benefactors, donations, bequests, sponsorships, parent/guardian fundraising (including subsidies through non-profit ECEC providers)	SL1G14C, SL1G14D 1 = Yes 2 = No 3 = Don't know	Recoding 1 = Yes (if SL1G14C = 1 or if SL1G14D = 1) 2 = No (if SL1G14C = 2 and if SL1G14D = 2) 3 = Don't know (if SL1G14C = 3 and if SL1G14D = 2) or (if SL1G14C = 2 and if SL1G14D = 3) or (if SL1G14C = 3 and if SL1G14D = 3)

Variable name	SUMSTAFGR		
Description	Sum of all staff members, grouped		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre> a) <Leaders> b) <Teachers> c) <Assistants> d) <Staff for individual children> e) <Staff for special tasks> f) <Interns> g) Other <ECEC staff>	SL1G17A, SL1G17B, …, SL1G17G SUMSTAFF (see Annex J.1.2)	Recoding 1 = SUMSTAFF < 15 2 = SUMSTAFF ≥ 15 & SUMSTAFF ≤ 29 3 = SUMSTAFF ≥ 30

Variable name	LTEACHGR		
Description	Proportion of leaders and teachers to total staff, grouped		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre> a) <Leaders> b) <Teachers>	SL1G17A, SL1G17B SUMSTAFF (see Annex J.1.2)	Recoding 1 = ((SL1G17A + SL1G17B) / SUMSTAFF) ≤ 0.50 2 = ((SL1G17A + SL1G17B) / SUMSTAFF) > 0.50 & ((SL1G17A + SL1G17B) / SUMSTAFF) ≤ 0.75 3 = ((SL1G17A + SL1G17B) / SUMSTAFF) > 0.75

Variable name	LASSISTGR		
Description	Proportion of assistants to total staff, grouped		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre> c) <Assistants>	SL1G17C SUMSTAFF (see Annex J.1.2)	Recoding 1 = (SL1G17C / SUMSTAFF) ≤ 0.50 2 = (SL1G17C / SUMSTAFF) > 0.50 & (SL1G17C) / SUMSTAFF) ≤ 0.75 3 = (SL1G17C / SUMSTAFF) > 0.75

Variable name	LSTOTH		
Description	Proportion of "other" staff to total staff, grouped		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre> d) <Staff for individual children> e) <Staff for special tasks> f) <Interns> g) Other <ECEC staff>	SL1G17D, SL1G17E, SL1G17F, SL1G17G SUMSTAFF (see Annex J.1.2)	Recoding 1 = ((SL1G17D + SL1G17E + SL1G17F + SL1G17G) / SUMSTAFF) ≤ 0.50 2 = ((SL1G17D + SL1G17E + SL1G17F + SL1G17G) / SUMSTAFF) > 0.50 & ((SL1G17D + SL1G17E + SL1G17F + SL1G17G) / SUMSTAFF) ≤ 0.75 3 = ((SL1G17D + SL1G17E + SL1G17F + SL1G17G) / SUMSTAFF) > 0.75

Variable name	LSTAFFLEFT		
Description	Proportion of staff who left the centre, grouped		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	<ECEC staff> who permanently left this <ECEC centre> during the last 12 months	SL1G18B SUMSTAFF (see Annex J.1.2)	Recoding 1 = (SL1G18B / SUMSTAFF) ≤ 0.50 2 = (SL1G18B / SUMSTAFF) > 0.50 & (SL1G18B / SUMSTAFF) ≤0.75 3 = (SL1G18B / SUMSTAFF) > 0.75

Variable name	NENRCHLD		
Description	Number of enrolled children		
Procedure	Simple recode		
Source	What is the number of children currently enrolled in <ISCED 2011 Level 0.2> or below in this <ECEC centre>?	SL1G19	Recoding 1 = Under 100 2 = 100 and above

Recodings based on staff data

Variable name	**SAGEGR**		
Description	Staff age groups		
Procedure	Simple recode		
Source	How old are you?	SS1G02	Recoding 1 = Under 20 2 = 20–29 3 = 30–39 4 = 40–49 5 = 50–59 6 = 60 and above

Variable name	**DIVPRAC**		
Description	Index of ECEC centre diversity practices		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	When considering daily interactions with children, to what extent do the following apply to this <ECEC> centre? a) The use of books and pictures featuring people from a variety of ethnic and cultural groups b) Books or toys show people from different ethnic and cultural groups in a variety of professional and social roles c) The children sometimes play with toys and artefacts from cultures other than the ethnic majority d) Some activities emphasise what people from different ethnic and cultural groups have in common	SS1G34A, SS1G34B, SS1G34C, SS1G34D 1 = Not at all 2 = Very little 3 = To some extent 4 = A lot	Recoding DIVPRAC = SS1G34A + SS1G34B + SS1G34C + SS1G34D

Variable name	SUMSTAFFTG		
Description	Sum of staff members working with the <target group>		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables were not administered.		
Source	How many people in each of the following categories were working with the <target group> on that day? a) <Leaders> b) <Teachers> c) <Assistants> d) <Staff for individual children> e) <Staff for special tasks> f) <Interns> g) Other <ECEC staff>	SS1G35A, SS1G35B, …, SS1G35G	Recoding SUMSTAFFTG = SS1G35A + SS1G35B + SS1G35C + SS1G35D + SS1G35E + SS1G35F + SS1G35G + 1 *Please note:* The person filling in the questionnaire was asked not to include herself/himself in the number of staff working with the <target group>. Consequently, "+1" is added to the sum of staff members working with the <target group>.

Variable name	SUMCHILD		
Description	Sum of boys and girls in <target group>		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	How many girls and boys were in this <target group> on that day? a) Girls b) Boys	SS1G37A, SS1G37BB	Recoding SUMCHILD = SS1G37A + SS1G37B

Variable name	HEADCHPAD		
Description	Headcount of children per adult		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	How many people in each of the following categories were working with the <target group> on that day? How many girls and boys were in this <target group> on that day?	SUMSTAFFTG SUMCHILD	Recoding HEADCHPAD = SUMCHILD / SUMSTAFFTG

1.2. Recoded variables used for table production which are not part of the TALIS 2018 public data set[2]

The second part of the codebook describes the recoding of advanced variables generated in the scope of table production. Variables are ordered (a) by population (first leader, second staff) and (b) numerically by the original variable name.

Recodings based on leader data

Variable name	**SL1G04ABC_mulvar_r3**		
Description	Early childhood education, administration or pedagogical leadership included in <education or training>		
Procedure	Multivariable code		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	Did the <education or training> you completed include the following topics? a) Training/education programme or course focused on early childhood b) Administration c) Pedagogical leadership	SL1G04A, SL1G04B, SL1G04C 1 = Yes 2 = No	Recoding 1 = No training in any of the following topics. 0 = Training in at least one of these topics (if SL1G04A or SL1G04B or SL1G04C = 1, then SL1G04ABC_mulvar_r3 = 0, else SL1G04ABC_mulvar_r3 = 1).

Variable name	**SL1G05BC_mulvar_r3**		
Description	Categorised number of years of work experience in an ECEC centre		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	How many years of work experience do you have, regardless of whether you worked full-time or part-time? b) Years working as an <ECEC centre leader> in total c) Years working in other roles in an <ECEC centre> in total (do not include years working as an <ECEC centre leader>)	SL1G05B, SL1G05C	Recoding SL1G05BC_mulvar_r3 = SL1G05B + SL1G05C 1 = Under 5 years 2 = 5--9 years 3 = 10 years or more

Variable name	SL1G14ABCD_mulvar_r1		
Description	ECEC centres that receive both public and private funding		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	In the last 12 months, which of the following sources provided funding for this <ECEC centre>? a) <Government> (including departments, municipal, local, regional, state, national and supranational levels) b) Fees or charges paid <directly or indirectly> to the <ECEC centre> by parents or guardians c) Non-governmental organisations (including religious institutions and employers) d) Benefactors, donations, bequests, sponsorships, parent/guardian fundraising (including subsidies through non-profit ECEC providers)	SL1G14A, SL1G14B, SL1G14C, SL1G14D 1 = Yes 2 = No 3 = Don't know	Recoding 1 = Yes (if SL1G14A=1 and SL1G14BCD_mulvar_r1=1), else 2 = No or don't know *Please note:* Participants who responded with 3 = Don't know to one of the items SL1G14A, SL1G14B, SL1G14C or SL1G14D were excluded from later analysis (filtered out).

Variable name	SL1G14ABCD_mulvar_r2		
Description	ECEC centres that receive funding only from <government>		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	In the last 12 months, which of the following sources provided funding for this <ECEC centre>? a) <Government> (including departments, municipal, local, regional, state, national and supranational levels) b) Fees or charges paid <directly or indirectly> to the <ECEC centre> by parents or guardians c) Non-governmental organisations (including religious institutions and employers) d) Benefactors, donations, bequests, sponsorships, parent/guardian fundraising (including subsidies through non-profit ECEC providers)	SL1G14A, SL1G14B, SL1G14C, SL1G14D 1 = Yes 2 = No 3 = Don't know	Recoding 1 = Yes (if SL1G14A=1 and SL1G14BCD_mulvar_r1=2), else 2 = No or don't know *Please note:* Participants who responded with 3 = Don't know to one of the items SL1G14A, SL1G14B, SL1G14C or SL1G14D were excluded from later analysis (filtered out).

Variable name	SL1G14BCD_mulvar_r1		
Description	ECEC centres that receive funding from at least one <private> source		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	In the last 12 months, which of the following sources provided funding for this <ECEC centre>? b) Fees or charges paid <directly or indirectly> to the <ECEC centre> by parents or guardians c) Non-governmental organisations (including religious institutions and employers) d) Benefactors, donations, bequests, sponsorships, parent/guardian fundraising (including subsidies through non-profit ECEC providers)	SL1G14B, SL1G14C, SL1G14D 1 = Yes 2 = No 3 = Don't know	Recoding 1 = Yes (if SL1G14B OR SL1G14C OR SL1G14D = 1), else 2 = No or don't know *Please note:* Participants who responded with 3 = Don't know to one of the items SL1G14B, SL1G14C or SL1G14D were excluded from later analysis (filtered out).

Variable name	SUMSTAFF		
Description	Sum of all staff members		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. Not administered does not count as missing, only if all variables are not administered.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre> a) <Leaders> b) <Teachers> c) <Assistants> d) <Staff for individual children> e) <Staff for special tasks> f) <Interns> g) Other <ECEC staff>	SL1G17A, SL1G17B, …, SL1G17G	Recoding SUMSTAFF = SL1G17A + SL1G17B + SL1G17C + SL1G17D + SL1G17E + SL1G17F + SL1G17G

Variable name	SL1G17AG_mulvar_T_r1		
Description	Share of teachers among all staff members		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (headcount) currently working in this <ECEC centre>. b) <Teachers>	SUMSTAFF, SL1G17B	Recoding SL1G17AG_mulvar_T_r1 = SL1G17B / SUMSTAFF

Variable name	SL1G17AG_mulvar_A_r1		
Description	Share of assistants among all staff members		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (headcount) currently working in this <ECEC centre>. c) <Assistants>	SUMSTAFF, SL1G17C	Recoding SL1G17AG_mulvar_A_r1 = SL1G17C / SUMSTAFF

Variable name	SL1G17BC_mulvar_r2		
Description	Headcount of assistants per teacher		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (head count) currently working in this <ECEC centre>. b) <Teachers> c) <Assistants>	SL1G17C, SL1G17B	Recoding SL1G17BC_mulvar_r2 = SL1G17C / SL1G17B *Please note*: If SL1G17B = 0, SL1G17BC_mulvar_r2 is set to equal SL1G17C.

Variable name	SL1G1718A_mulvar_r1		
Description	Percentage of staff who began working at the ECEC centre in the past year		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable missing.		
Source	For each type of position listed below, please indicate the number of <ECEC staff> (headcount) currently working in this <ECEC centre>. <ECEC staff> who began work at this <ECEC centre> during the last 12 months.	SUMSTAFF, SL1G18A	Recoding SL1G1718A_mulvar_r1 = SL1G18A / SUMSTAFF × 100

Variable name	SL1G19_02_r1 (ISCED level 02), SL1G19_U3_r1 (U3)		
Description	National quarters of centre size for ISCED level 02, U3		
Procedure	Complex recode		
Source	What is the number of children currently enrolled in <ISCED 2011 level 0.2> or below in this <ECEC centre>?	SL1G19	Recoding 0 = Other (middle national quarters of centre size) 1 = Lowest national quarter of centre size 2 = Top national quarter of centre size *Please note:* Further information on the recoding of the national quarters can be found in Chapter 12 of this report.

Variable name	SL1G35DEF_mulvar_r1 (ISCED level 02)		
Description	Index for centre engagement with the community		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	To what extent do the following statements apply to this <ECEC centre>? d) The <ECEC centre> works with the local neighbourhood e) The <ECEC centre> collaborates with the local <ISCED 2011 Level 1> schools f) The <ECEC centre> has excursions to outdoor areas, such as parks, forests, beaches or other nature areas	SL1G35D, SL1G35E, SL1G35F 1 = Not at all 2 = To some extent 3 = Quite a bit 4 = A lot	Recoding SL1G35DEF_mulvar_r1= SL1G35D + SL1G35E + SL1G35F

Variable name	S1LPART_02_r1 (ISCED level 02), S1LPART_U3_r1 (U3)		
Description	National quarters of the scale S1LPART (opportunities for parental participation)		
Procedure	Complex recode		
Source	Scale (opportunities for parental participation)	SL1LPART	Recoding 0= Other (middle national quarters of SL1LPART) 1 = Lowest national quarter of SL1LPART size 2 = Top national quarter of SL1LPART *Please note:* Further information on the recoding of the national quarters can be found in Chapter 12 of this report.

Recodings based on staff data

Variable name	**SS1G13AJ_mulvar_r1**		
Description	No participation in professional development activities during the last 12 months		
Procedure	Multivariable recode		
Missing values	Missing only if all variables are missing.		
Source	During the last 12 months, did you participate in any of the following professional development activities? a) Courses/seminars attended in person b) Online courses/seminars c) Conferences where <ECEC staff> and/or researchers present their research or discuss educational issues d) Qualification programme (e.g. a degree programme) e) Observation visits to other <ECEC centres> f) Peer and/or self-observation and coaching as part of a formal arrangement g) On-site coaching by an external person h) Participation in a network of professionals working with children i) Induction or mentoring activities j) Other	SS1G13A, SS1G13B, …, SS1G13J 1 = Yes 2 = No	Recoding 1 = No participation in professional development activities over the last 12 months 0 = Participation in professional development activities over the last 12 months (if SS1G13A or SS1G13B, …, or SS1G13J = 1)

Variable name	**SS1G13ABC_mulvar_r1**		
Description	Participation in courses, online courses or conferences		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	During the last 12 months, did you participate in any of the following professional development activities? a) Courses/seminars attended in person b) Online courses/seminars c) Conferences where <ECEC staff> and/or researchers present their research or discuss educational issues	SS1G13A, SS1G13B, SS1G13C 1 = Yes 2 = No	Recoding 1 = Participation in courses, online courses or conferences (if SS1G13A or SS1G13B or SS1G13C=1), else 0 = No participation in courses, online courses or conferences

Variable name	SS1G13FG_mulvar_r1		
Description	Participation in peer- or self-observation or coaching		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	During the last 12 months, did you participate in any of the following professional development activities? f) Peer and/or self-observation and coaching as part of a formal arrangement g) On-site coaching by an external person	SS1G13F, SS1G13G 1 = Yes 2 = No	Recoding 1 = Participation in peer- or self-observation or coaching (if SS1G13F or SS1G13G = 1), else 0 = No participation in peer- or self-observation or coaching

Variable name	SS1G13EHIJ_mulvar_r2		
Description	Participation in visits, networks of professionals, induction, mentoring or other activities		
Procedure	Multivariable recode		
Missing values	Missing only if all variables are missing.		
Source	During the last 12 months, did you participate in any of the following professional development activities? e) Observation visits to other <ECEC centres> h) Participation in a network of professionals working with children i) Induction or mentoring activities j) Other	SS1G13E, SS1G13H, SS1G13I, SS1G13J 1 = Yes 2 = No	Recoding 1 = Participation in visits, networks of professionals, induction, mentoring or other activities (if SS1G13E or SS1G13H or SS1G13I or SS1G13G = 1), else 0 = No participation in visits, networks of professionals, induction, mentoring or other activities

Variable name	SS1G15AH_mulvar_r1		
Description	No support for professional development activities during the last 12 months		
Procedure	Multivariable recode		
Missing values	Missing only if all variables are missing.		
Source	For the professional development in which you participated during the last 12 months, did you receive any of the following? a) Release from working with children for activities during regular working hours b) Non-monetary support for activities outside working hours (e.g. reduced contact time with children, days off, study leave) c) Reimbursement or payments of costs d) Materials needed for the activities e) Monetary supplements for activities outside working hours f) Non-monetary rewards (e.g. <classroom/playgroup/group> resources/materials, book vouchers, software/apps) g) Non-monetary professional benefits (e.g. fulfilling professional development requirements, improving my promotion opportunities) h) Increased salary	SS1G15A, SS1G15B, …, SS1G15H 1 = Yes 2 = No	Recoding 1 = No support received for professional development activities over the last 12 months 0 = Support received for professional development activities over the last 12 months (if SS1G15A or SS1G15B, …, or SS1G15H = 1)

Variable name	SS1G15AB_mulvar_r1		
Description	Time resources for professional development		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. "Not administered" does not count as missing, only if all variables are not administered.		
Source	For the professional development in which you participated during the last 12 months, did you receive any of the following? a) Release from working with children for activities during regular working hours b) Non-monetary support for activities outside working hours (e.g. reduced contact time with children, days off, study leave)	SS1G15A, SS1G15B 1 = Yes 2 = No	Recoding 1 = Staff receive time to attend professional development (if SS1G15A or SS1G15B = 1), else 0 = Staff receive no time to attend professional development

Variable name	SS1G15CD_mulvar_r1		
Description	Cost resources for professional development		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. "Not administered" does not count as missing, only if all variables are not administered.		
Source	For the professional development in which you participated during the last 12 months, did you receive any of the following? c) Reimbursement or payments of costs d) Materials needed for the activities	SS1G15C, SS1G15D 1 = Yes 2 = No	Recoding 1 = Staff receives payment of costs or material for professional development (if SS1G15C or SS1G15D = 1), else 0 = Staff receives no payment of costs or material for professional development

Variable name	SS1G15FG_mulvar_r1		
Description	Non-monetary incentives for professional development		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. "Not administered" does not count as missing, only if all variables are not administered.		
Source	For the professional development in which you participated during the last 12 months, did you receive any of the following? f) Non-monetary rewards (e.g. <classroom/playgroup/group> resources/materials, book vouchers, software/apps) g) Non-monetary professional benefits (e.g. fulfilling professional development requirements, improving my promotion opportunities)	SS1G15F, SS1G15G 1 = Yes 2 = No	Recoding 1 = Staff receives non-monetary rewards or non-monetary benefits for professional development (if SS1G15F or SS1G15G = 1), else 0 = Staff receives no non-monetary rewards or non-monetary benefits for professional development

Variable name	SS1G15EH_mulvar_r1		
Description	Monetary incentives for professional development		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing. "Not administered" does not count as missing, only if all variables are not administered.		
Source	For the professional development in which you participated during the last 12 months, did you receive any of the following? e) Monetary supplements for activities outside working hours h) Increased salary	SS1G15E, SS1G15H 1 = Yes 2 = No	Recoding 1 = Staff receives monetary supplements or increased salary for professional development (if SS1G15E or SS1G15H = 1), else 0 = Staff receives no monetary supplements or increased salary for professional development

Variable name	SS1G35A36_mulvar_r1		
Description	Total number of leaders per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35A is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? a) <Leaders> Which of the following categories best represents your role when you worked with the <target group> on that day? 1 = <Leader>	SS1G35A, SS1G36	Recoding SS1G35A + 1 Or, SS1G35A *Please note:* If the person filling in the questionnaire is a leader according to SS1G36, "+1" was added to SS1G35A36_mulvar_r1.

Variable name	SS1G35B36_mulvar_r1		
Description	Total number of teachers per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35B is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? b) <Teachers> Which of the following categories best represents your role when you worked with the <target group> on that day? 2 = <Teacher>	SS1G35B, SS1G36	Recoding SS1G35B + 1 Or, SS1G35B *Please note:* If the person filling in the questionnaire is a teacher according to SS1G36 , "+1" was added to SS1G35B36_mulvar_r1.

Variable name	SS1G35C36_mulvar_r1		
Description	Total number of assistants per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35C is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? c) <Assistants> Which of the following categories best represents your role when you worked with the <target group> on that day? 3 = <Assistant>	SS1G35C, SS1G36	Recoding SS1G35C + 1 Or, SS1G35C *Please note:* If the person filling in the questionnaire is an assistant according to SS1G36, "+1" was added to SS1G35C36_mulvar_r1.

Variable name	SS1G35D36_mulvar_r1		
Description	Total number of staff for individual children per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35D is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? d) <Staff for individual children> Which of the following categories best represents your role when you worked with the <target group> on that day? 4 = <Staff for individual children>	SS1G35D, SS1G36	Recoding SS1G35D + 1 Or, SS1G35D *Please note:* If the person filling in the questionnaire was a staff member for individual children according to SS1G36, "+1" was added to SS1G35D36_mulvar_r1.

Variable name	SS1G35E36_mulvar_r1		
Description	Total number of staff for special tasks per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35E is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? e) <Staff for special tasks> Which of the following categories best represents your role when you worked with the <target group> on that day? 5 = <Staff for special tasks>	SS1G35E, SS1G36	Recoding SS1G35E + 1 or SS1G35E *Please note:* If the person filling in the questionnaire was a staff member for special tasks according to SS1G36, "+1" was added to SS1G35E36_mulvar_r1.

Variable name	**SS1G35F36_mulvar_r1**		
Description	Total number of interns per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35F is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? f) <Interns> Which of the following categories best represents your role when you worked with the <target group> on that day? 6 = <Intern>	SS1G35F, SS1G36	Recoding SS1G35F + 1 or SS1G35F *Please note:* If the person filling in the questionnaire was an intern according to SS1G36, "+1" was added to SS1G35F36_mulvar_r1.

Variable name	**SS1G35G36_mulvar_r1**		
Description	Total number of other staff per <target group>		
Procedure	Multivariable recode		
Missing values	Missing if SS1G35G is missing		
Source	How many people in each of the following categories were working with the <target group> on that day? g) Other <ECEC staff> Which of the following categories best represents your role when you worked with the <target group> on that day? 7 = Other	SS1G35G, SS1G36	Recoding SS1G35G + 1 or SS1G35G *Please note:* If the person filling in the questionnaire marked the category "Other staff" in item SS1G36, "+1" was added to SS1G35G36_mulvar_r1.

Variable name	**SS1G3537_mulvar_02_r3 (ISCED level 02), SS1G3537_mulvar_U3_r3 (U3)**		
Description	National quarters of the recoded variable HEADCHPAD (headcount of children per adult)		
Procedure	Complex recode		
Source	How many people in each of the following categories were working with the <target group> on that day? How many girls and boys were in this <target group> on that day?	HEADCHPAD	Recoding 0= Other (Middle national quarters of HEADCHPAD) 1 = Lowest national quarter of HEADCHPAD 2 = Top national quarter of HEADCHPAD *Please note:* Further information on the recoding of the national quarters can be found in Chapter 12 of this report.

Variable name	SS1G37A_r1		
Description	Percentage of girls in the <target group>		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	SS1G37A How many girls were in this <target group> on that day? SS1G37B How many boys were in this <target group> on that day?	SS1G37A, SS1G37B	Recoding SS1G37A_r1 = SS1G37A / SUMCHILD × 100

Variable name	SS1G37B_r1		
Description	Percentage of boys in the <target group>		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	SS1G37A How many girls were in this <target group> on that day? SS1G37B How many boys were in this <target group> on that day?	SS1G37A, SS1G37B	Recoding SS1G37B_r1 = SS1G37B / SUMCHILD × 100

Variable name	SS1G37AB_mulvar_r2		
Description	Unbalanced gender group (<target group> consists of 67% or more children who have the same gender)		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	SS1G37A How many girls were in this <target group> on that day? SS1G37B How many boys were in this <target group> on that day?	SS1G37A_r1, SS1G37B_r1	Recoding 1 = Unbalanced gender group (if SS1G37A / SUMCHILD × 100 > 67 or SS1G37B / SUMCHILD × 100 > 67), else 0 = Balanced gender group

Variable name	SS1G39_ISCED_02_mulvar_r3 (ISCED 02)		
Description	Outside expected age range if more than 33% of the children are 2 years of age or younger		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	Please indicate the number of children of the following ages who were in this <target group> on that day. a) Children aged less than 1 year b) Children aged 1 year c) Children aged 2 years d) Children aged 3 year e) Children aged 4 years f) Children aged 5 years g) Children aged 6 years and above	SS1G39A, SS1G39B, , …, SS1G39G	Recoding 1 = Outside of expected age range (if (SS1G39A+SS1G39B+SS1G39C) / (SS1G39A+SS1G39B+…+SS1G39G) × 100 > 33 %,), else 0 = Inside expected age range

Variable name	SS1G39_Under_3_mulvar_r3 (U3)		
Description	Outside expected age range if more than 33% of the children are 3 years or older		
Procedure	Multivariable recode		
Missing values	Missing if at least one variable is missing.		
Source	Please indicate the number of children of the following ages who were in this <target group> on that day. a) Children aged less than 1 year b) Children aged 1 year c) Children aged 2 years d) Children aged 3 year e) Children aged 4 years f) Children aged 5 years g) Children aged 6 years and above	SS1G39A, SS1G39B, , …, SS1G39G	Recoding 1 = Outside of expected age range (if (SS1G39D+SS1G39E+SS1G39F+SS1G39G) / (SS1G39A+SS1G39B+…+SS1G39G) × 100 > 33 %,), else 0 = Inside expected age range

Notes

1. A decision was made with respect to listing items to interpret an omitted answer to the response category “other” as “no” or “zero” in case all other items of the list had a valid entry.

2. Some recodings in this paragraph might be counterintuitive to the reader, such as the recoded variable SS1G15AH_mulvar_r1: “No support for professional development activities during the last 12 months”, in which “No support” is coded to 1 and “Support” is coded to 0. The reasons for this are technical and relate to the analysis software, the IEA IDB Analyzer, used to produce the tables.

TALIS Starting Strong 2018 Technical Report

The OECD Starting Strong Teaching and Learning International Survey (TALIS Starting Strong) is the first international survey that focusses on the early childhood education and care workforce to learn about staff and leaders' characteristics, their practices with children and their views on the sector. This technical report details the steps, procedures, methodologies, standards and rules that TALIS Starting Strong 2018 used to collect high-quality data. The primary purpose of the report is to support readers of the international and subsequent thematic reports, as well as users of the public international database when interpreting results, contextualising information and using the data.

For more information, please contact: startingstrongsurvey@oecd.org

Further information about TALIS Starting Strong project can be found at: http://www.oecd.org/education/school/oecd-starting-strong-teaching-and-learning-international-survey.htm.