Organisation for Economic Co-operation and Development

EDU/WKP(2020)8

Unclassified **English text only**

23 April 2020

DIRECTORATE FOR EDUCATION AND SKILLS

Teaching and Learning International Survey (TALIS) 2018 Analysis Plan

OECD Education Working Paper No. 220

By Heather Price (Marian University) and Ralph Carstens (International Association for the Evaluation of Educational Achievement [IEA])

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Contact:
Heather Price (hprice@alumni.nd.edu)
Ralph Carstens (ralph.carstens@iea-hamburg.de)

JT03461144

OECD EDUCATION WORKING PAPERS SERIES OECD

Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed herein are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome, and may be sent to the Directorate for Education and Skills, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

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.

Comment on the series is welcome, and should be sent to edu.contact@oecd.org.

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

Acknowledgements

Work on the TALIS 2018 analysis plan was an international and collaborative effort. The authors would like to express their sincere thanks and gratitude to the members of the Questionnaire Expert Group (QEG) who provided the substance of Sections 2 and 3 as per their respective areas of expertise: Sigrid Blömeke, Hilary Hollingsworth, David Kaplan, Daniel Muijs, Trude Nilsen, Ronny Scherer, and Fons van de Vijver. Key members of the international research consortium contributed the statistical, analytical and methodological deliberations in Section 4 and conducted reviews of all other sections: Alena Becker, Minge Chen, Jean Dumais, Sabine Meinck and Agnes Stancel-Piatak.

John Ainley ensured consistency and follow-through from the conceptual framework. Steffen Knoll reviewed the plan at key stages. The members of the Technical Advisory Group (TAG) provided valuable comments during the development: Timothy L. Kennel, Bart Meuleman and Christian Monseur. We gratefully acknowledge all contributions of the late Fons van de Vijver, who was Chair of the TAG and an advisor to TALIS since the first cycle in 2008.

The authors are grateful for all inputs, discussions and critical scaffolding provided the TALIS analysts at the OECD Secretariat to ensure consistency with the eventual reporting of TALIS data, in particular Pablo Fraser, Gabor Fulöp, Noémie Le Donné and Karine Tremblay. We also acknowledge the many reflections provided by the members of the TALIS Governing Board at key stages.

Emily Groves (OECD Secretariat) provided the proofreading and managed the layout process, with support from Rachel Linden (also OECD Secretariat).

Abstract

This analysis plan proposes the specifications for analysis for various studies using the third cycle of the Teaching and Learning International Survey (TALIS 2018) data. It serves as a link between the *TALIS 2018 Conceptual Framework* and the reporting plan. After a brief introduction in Section 1, Sections 2 and 3 suggest analyses for the eleven focal themes collected in the 2018 dataset. The analyses are grounded in literature discussed in detail in the conceptual framework and questions of interest based on the priorities set forth by the Board of Participating Countries (BPC) and later in subsequent TALIS Governing Board (TGB) meetings. Section 2 highlights, by theme, new items, items measured across TALIS cycles, scale constructs (when applicable), and within-theme multivariate analyses. Section 3 provides recommendations for research questions to test between thematic indicators and includes a reference look-up table that identifies the most appropriate aggregation unit for analysis (teacher-, school-, or system-level) based on the item construction and the policy questions of interest. Section 4 outlines the reporting standards for the TALIS 2018 data with subsections that discuss: defining the participants, estimation statistics, displaying statistical information, specifications for scales and specifications for statistical models. The target audience for this analysis plan are internal analysts working on the project, such as those at the OECD and, by extension, analysts in national centres.

Résume

Le plan d'analyse propose les spécifications pour les différentes études utilisant les données du 3e cycle de l'Enquête internationale sur l'enseignement et l'apprentissage (TALIS 2018). Il sert de lien entre le cadre conceptuel de TALIS 2018 et le plan de communications. Après une brève introduction dans la section 1, les sections 2 et 3 suggèrent des analyses pour les 11 thèmes principaux pour lesquels des données ont été recueillies dans TALIS 2018. Les analyses reposent sur les études qui ont été discutées en détail dans le cadre conceptuel et des questions d'intérêt à partir des priorités établies par le Conseil des pays participants (CPP) et lors des réunions des Comités directeurs de TALIS (CDT) qui ont suivi. La section 2 met en évidence, par thèmes, de nouveaux items, des items mesurés dans les cycles de TALIS, et des indices construits (si applicable), et des analyses multivariées par thème. La section 3 fournit des recommandations pour des questions de recherche pour tester les liens entre les indicateurs thématiques et comprend une référence à un tableau qui identifie les unités d'analyse les plus pertinentes (au niveau des enseignants, des établissements scolaires ou des systèmes) s'appuyant sur la construction des items et les questions d'intérêt sur le plan des politiques. La section 4 décrit les normes de présentation des données de TALIS 2018 comprenant des sous-sections qui discutent la définition des participants, les estimations statistiques, les manières de présenter l'information statistiques, les spécifications des échelles et des modèles statistiques. Ce plan d'analyse a comme public cible les analystes internes qui travaillent sur le projet, comme ceux à l'OCDE et les analystes dans les centres nationaux.

Table of contents

TABLES

Teaching and Learning International Survey (TALIS) 2018 Analysis Plan

Introduction

The third cycle of the Teaching and Learning International Survey (TALIS 2018) aims to continue the tradition of providing timely, relevant, accurate, comparable, and interpretable data regarding the conditions of teaching and learning environments to participating OECD members, non-member and partner economies and sub-national entities (TALIS participants).

This document proposes the specifications for analysis for various studies using the TALIS 2018 data, as proposed in the TALIS 2018 reporting plan (OECD, 2017[1]), as well as for analyses by participating countries and economies. Each theme in the survey is presented with corresponding literature-informed research questions and proposed analyses to address research questions based on the priorities set forth by the Board of Participating Countries (BPC) in a priority rating exercise at the 18th meeting of the BPC in 2015 (OECD, 2015[2]) and henceforth in subsequent TALIS Governing Board (TGB)[1] meetings. System-level descriptions of the themes are also presented. Reporting standards for calculating averages, missing data, and other data validation specifications conclude this document. Some proposed analyses may need to be adapted to fit scales and indices established during the main survey (MS) analyses.

The guidelines presented in this document aim at enhancing the quality of the statistical outputs and inferences that can be derived from the TALIS 2018 databases. The expression "statistical outputs" covers a broad field, ranging from basic charts and other displays to inferences on population model parameters. The idea of "quality" is a multi-faceted notion that comprises – in no specific order – accuracy, precision, comparability, relevance, interpretability, and utility.

This *TALIS 2018 Analysis Plan* was developed by the TALIS 2018 Consortium at the International Association for the Evaluation of Educational Achievement (IEA) and the TALIS 2018 Questionnaire Expert Group (QEG), in consultation with the OECD Secretariat and the TALIS 2018 Technical Advisory Group (TAG). This draft of the analysis plan was presented at the 6th TGB meeting in February 2018. Feedback from this meeting and other sources was incorporated and this analysis plan was revised according to the data validation and scaling work conducted with the main survey data.

It is important to highlight that this analysis plan is intended to be used in co-ordination with the *TALIS 2018 Conceptual Framework*, reporting plan, and *TALIS 2018 Technical Report* (Ainley and Carstens, 2018[3]; OECD, 2019[4]; OECD, 2017[1]). These documents will be referenced throughout this document to emphasise the links between them.

[1] Referred to as the TALIS Board of Participating Countries (BPC) when TALIS 2018 commenced until the end of 2015.

1. Overview of the analytic framework

A central goal of TALIS is to monitor and compare the conditions of teaching and learning in the participating education systems. In addition to descriptions of useful and relevant information about teachers, teaching conditions, and learning environments, TALIS aims to help to identify the conditions that are associated with variability within and across education systems.

TALIS 2018 covers issues that have endured over the ten-year period encompassing the three cycles (2008, 2013 and 2018) and has been adjusted to include newly emergent issues as well. Therefore, TALIS 2018 combines aspects from 2013 and 2008 with new aspects developed for 2018. The 2018 cycle is an important refinement rather than a major redevelopment of the survey. Specifically, TALIS 2018 provides indicators on teachers' instructional practices, school leadership, teachers' professional practices, teacher education and initial preparation, teacher feedback and development, teacher human resource measures and stakeholder relations, school climate, teacher self-efficacy, job satisfaction, innovation, and diversity and equity.

The overall objective of the TALIS is to provide robust international indicators and policy-relevant analysis on teachers and teaching to help countries and economies review and develop policies that promote conditions for effective teaching and learning. The principles guiding the survey, as detailed in the conceptual framework, are that it has policy relevance, added value, validity, reliability, comparability, rigour, interpretability, efficiency, cost-effectiveness and is indicator-oriented.

The intended use of the data is first to give teachers a voice so that educational stakeholders, including policy makers, at the system- and school-level can develop improvements directed at the teaching and learning conditions in the education system. School administrators and teachers are expected to use the data to learn from the variability of conditions and policies across the globe to improve their schools. Academics and scholars are expected to use the data to further develop, test, and advance theories of teaching and learning. The public are expected to benefit from such data as education systems are improved with data-informed decisions.

As is the case with large, nationally representative survey data such as TALIS, the resulting research will provide descriptions and associations about issues asked in the survey. This research can give shape to discussions about educational policies, programmes, and theories (Berliner, 2002[5]; Farley-Ripple et al., 2018[6]; Slavin, 2002[7]; Teddlie and Reynolds, 2000[8]). These policies, programmes, and theories will need their own set of studies to test the fidelity, effectiveness, impact, and consequences of implementation (Berliner, 2002[5]; Farley-Ripple et al., 2018[6]; Pant, 2016[9]; Teddlie and Reynolds, 2000[8]).

For example, imagine the survey research finds that a "books in the home" measure has an impact on children's literacy. But an implementation study that only drops off books at homes with no further action misses the importance of the interpretation of the survey item. Understanding that "books in the home" is an item used to signal parents' literacy and socio-economic state helps explain the factors external to school that may impact child literacy. Understanding these external conditions can be used to shape programmes seeking to improve literacy. For implementation studies, longitudinal, experimental, and multi-site designs and analyses are often used (Creemers and Kyriakides, 2008[10]; Teddlie and Reynolds, 2000[8]).

Connecting descriptive and correlational research to policy and implementation is best achieved when there are co-design partnerships between academic researchers and school educators (Coburn, Penuel and Geil, 2013[11]; Creemers and Kyriakides, 2008[10]; Farley-Ripple et al., 2018[6]; Honig and Coburn, 2008[12]). These partnerships have longer lasting impact and value to education than top-down mandates from local or national policy makers (Coburn, Penuel and Geil, 2013[11]; Farley-Ripple et al., 2018[6]; Honig and Coburn, 2008[12]). It is often only when the partnership dialogues take place that the political nuances and ramifications of policies are discussed (Coburn, Penuel and Geil, 2013[11]; Creemers and Kyriakides, 2008[10]; Farley-Ripple et al., 2018[6]). Considering these political conditions are critical to legitimacy in implementation and to mediating the effects of policy (Farley-Ripple et al., 2018[6]; Honig and Coburn, 2008[12]). School leaders are central bridges for implementation (Leithwood et al., 2004[13]; Seashore Louis et al., 2010[14]; Wallace Foundation Staff, 2011[15]).

Given this research orientation, TALIS is designed to collect nationally representative data from the perspective of teachers and give teachers a voice in the dialogue of educational policy. These data are supplemented by a school survey that provides an additional voice, that of the school leader. These data focus on the International Standard Classification of Education (ISCED) level 2 teachers; however, international options did collect data at the ISCED 1 and ISCED 3 levels as well. ISCED 2 analyses are the focus of this document.

TALIS data are intended as a complement to other OECD data, such as that of the Programme for International Student Assessment (PISA) and Education at a Glance (EAG). However, it is not possible to link the teacher-level, classroom-level, and student-level data. At the school level, data can be linked between TALIS and PISA. Those specifications are not discussed here, but can be found in the TALIS 2013 and 2018 technical reports (OECD, 2019[4]; OECD, 2014[16]).

Given the scope of these teacher data, and as the following specifications will reiterate, careful interpretations of the results are warranted. All analyses must underscore that these data are a cross-sectional snapshot in time and were not designed to collect data in a longitudinal design that can prove any cause-and-effect. Relationships between items and concepts can be tested, but there is no scientific method to confirm a causal process.

The teachers surveyed in 2018 are not the same teachers that were surveyed in 2013 or 2008. However, the sampling method that was used in all three cycles of TALIS does allow for comparisons of system-level statistics across the time periods. For multi-cycle participating countries and economies, analyses will be proposed to test whether or not there have been significant changes in certain areas of teaching and learning.[2]

The target audience for this analysis plan are internal analysts working on the project, such as those at the OECD and, by extension, analysts in national centres. This document serves as a link between the conceptual framework and the reporting plan (Ainley and Carstens, 2018[3]; OECD, 2017[1]). This document can be used as a reference for analysts of the TALIS 2018 data because Sections 2 and 3 summarise the development of the TALIS 2018 survey questions by theme and outline how the questions link to the TALIS Governing Board priorities, prior TALIS results, and the literature and theories described in the conceptual framework. Section 4 details technical specifics for consideration, given the deployment of the TALIS 2018 survey, such as how to calculate 2018 averages so that they can be comparable to the 2008 or 2013 data since the scope of participating countries and economies has widened.

[2] Specifications to calculate comparable averages for participating countries that expanded their data collection between 2013 and 2018, such as Flanders in 2013 to Belgium in 2018, are detailed in Section 4.

2. Within-theme analyses

This section organises the analyses by theme. Each thematic subsection frames the analytic plan using the priorities set forth in the 18th meeting of the BPC in co-ordination with the evolving suggestions from the TGB during the development phase. Each thematic subsection highlights new items and items of high priority to gauge change over time using system-wide statistics. When applicable, subsections also discuss the proposed scale constructs for new items. Following the current educational policy issues flagged by the TGB as of high importance to them, and in accordance with existing research described in the conceptual framework and TALIS 2013 results (Ainley and Carstens, 2018[3]; OECD, 2014[17]), multivariate analyses are proposed where appropriate.

2.1. Teacher background and initial teacher education

Theme priorities

The TALIS Board of Participating Countries (BPC) – now the TGB – expressed a particular desire to learn about:[3]

- 2.1.1) attracting motivated candidates into teaching

Regarded as lesser priority were:

- 2.1.2) education and qualifications of teachers
- 2.1.3) initial teacher education and pathways into the profession

Least priority was given to:

- 2.1.4) sociological composition of teachers.

System-wide statistics

Given the low priority of "Education and qualifications of teachers", "Initial teacher education and pathways into the profession", and "Sociological composition of teachers", system-wide descriptive statistics (mean and variability) can be focused on the main variables of gender (TQ01), age (TQ02), and highest level of formal education (TQ03), while putting less emphasis on the type of teacher education programme (TQ04), the year teacher education was completed (TQ05), and elements included in teacher education (TQ06).

Given the high priority of "Attracting motivated candidates into teaching" more attention should be paid to: motivation to become a teacher (TQ07) and teaching as first choice (TQ08). System-wide descriptive statistics (mean and variability) with respect to TQ07 can be reported by its two subscales and both TQ07 and TQ08 could be reported by core subgroups: gender, education level by system-specific median split, for example; type of teacher education programme, by options 1/2 versus 3-6, and teacher generation by, for instance, a system-specific median split of the year teacher education was completed, for example.

[3] Henceforth, all references will be made as "TGB" since this is the governance structure in place at the time of writing.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	Teacher questionnaire item(s) (TQ)	Principal questionnaire item(s) (PQ)	Type of response
Motivation to teach	2.1.1	TQ07a,b,c,d,e,f,g	N/A	4-point importance scale
Teaching as first choice career	2.1.1	TQ08	N/A	Yes/no
Qualification pathway	2.1.2	TQ04	N/A	7-option exclusive choice
Qualification elements and preparedness	2.1.2	TQ06d,e,f,g,h,i,j,k,l	N/A	Matrix: Yes/no & 4-point preparedness scale
Teacher generation	2.1.3	TQ05	N/A	Numeric year

There are many items regarding motivation to teach and qualification elements. All these measures can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently nominated per country/economy and the item cited as the most common across all countries/economies.

Estimates of change

Several measures regarding teacher background and initial teacher education are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures.

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Gender	2.1.4	TQ01	✓	✓	✓	PQ01
Age	2.1.4	TQ02	✓	✓	✓	PQ02
Highest educational attainment	2.1.2	TQ03	*	✓	✓	PQ03
Qualification elements and preparedness	2.1.2	TQ06a,b,c	✓	✓		N/A
Work experience	2.1.4	TQ11	✓	✓	+	PQ04

* The question in 2018 uses new (2011) ISCED thresholds that differ from TALIS 2008 and 2013
+ The question in 2008 used optional ranges of years as categories instead of numeric write-in answer.

Recoding adjustments will be needed to assess changes on highest education and work experience measures because of coding differences between cycles.

In order to measure changes across TALIS surveys, the TALIS 2013 statistics would need to be mirrored. To do this, the coverage of teachers needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Few opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses. However, system-wide proportional representation of teachers and principals by gender could point to gender disparities in the education sector.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Personal utility motivation to teach	N/A	TQ07a	TQ07b	TQ07c	TQ07d		
Social utility motivation to teach	N/A	TQ07e	TQ07f	TQ07g			
Perceived value of teachers	N/A	TQ54c	TQ54d	TQ54e			

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

Given the high priority of "Attracting motivated candidates into teaching", analyses can be conducted to better understand the process under which teachers self-select into the profession. The question about whether teaching was the first career choice of in-service teachers may be indicative of high levels of motivation. Using this approach, regression parameters for the relation between motivation (TQ07) to become teacher and teaching as first choice (TQ08) may be reported.

Regression parameters can be estimated for subgroups to relate motivation and teaching as a first choice to all the variables mentioned: gender (TQ01), education level (TQ03), and teacher generation (TQ05) by a system-specific median split of the year teacher education was completed, for example.

It is possible that teacher generations differ substantially with respect to almost every variable, given the large changes teacher education had to undergo over the course of the past three to four decades, and this applies to almost all countries and economies. Therefore, it may be worthwhile to check all analyses to see whether generational differences (TQ05) need to be taken into account because, otherwise, they erase other important relations. This is a general recommendation that also applies to all other work in TALIS that goes beyond mere descriptive statistics.

2.2. Human resources

Theme priorities

The TGB expressed the desire to:

- 2.2.1) add new material on linking teacher perception of the value of the profession with more specific stakeholders, teachers' views regarding the mechanisms leading to the professionalism of teaching and for improving the status of teaching (including horizontal and vertical career trajectories and teachers' involvement in educational reforms), as well as new material on teachers' perceptions of linking pay to performance.
- 2.2.2) add new material on teachers' views on: stakeholder relations, relations between the unions and the government, on the extent to which they feel governments relate to teachers, and on their desired focus of educational reforms.

Human resources and stakeholder issues were not presented as a separate section in the TALIS 2013 surveys. Rather, the various aspects of each were addressed at different places in the questionnaires. TALIS 2018 is adding to the TALIS 2013 items in this area by addressing the following main aspects of the theme:

- 2.2.3) attracting good students into teaching; recognition, reward and evaluation of teachers; and teachers' working time.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the job satisfaction theme.

New to 2018

For each of the following measures, new to TALIS 2018 system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Attracting good students into teaching	2.2.1	TQ7, TQ8	N/A	TQ7 4-point scale, TQ8 yes/no
Extent to which teachers are valued in school, community and society	2.2.1 2.2.2	TQ54	N/A	4-point scale
Desired focus of educational reform	2.2.2	TQ55	N/A	3-point scale

There are many items regarding teachers' value and desired focus of educational reform. These measures can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding human resources are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures.

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Teachers' working time	2.2.3	TQ16,TQ17, TQ18	✓	✓	✓	N/A
Principal working time	2.2.3	N/A	✓	✓		PQ21

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be replicated. To do this, the sample frame coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Triangulation is not possible with the human resource items.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Personal utility motivation to teach	N/A	TQ07a	TQ07b	TQ07c	TQ07d		
Social utility motivation to teach	N/A	TQ07e	TQ07f	TQ07g			
Perceived value of teachers	N/A	TQ54c	TQ54d	TQ54e			

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

Attracting high-quality applicants to the profession is an issue in many countries and economies. According to a number of researchers – see, for example, Sahlberg (2011[18]) – this factor distinguishes some high-performing education systems from those systems that do not perform as well. Both intrinsic and extrinsic value (e.g. rewards) and both personal and social utility appear to influence a person's motivation for choosing teaching as a career, as do social factors, such as the esteem in which the profession is held. However, the balance of these factors shows some cross-country variation (Watt and Richardson, 2008[19]; Watt et al., 2012[20]). The variables here will allow comparisons to be made across countries and economies, and to be related to working hours as well as teachers' perceived value by school members, communities and society. Teacher workload and time spent on different activities differ between, but also within, countries and economies and is likely related to principal workload.

Priorities for educational reform may be related to perceived value in society, and lends itself to interesting possible between-country comparisons.

2.3. Instructional practices

Theme priorities

The TGB expressed the desire to:

- 2.3.1) repeat and improve indicators on overall instructional teaching practices; and
- 2.3.2) add new material on instructional practices that captures core practices that are related to student outcomes, including classroom management, clarity of instruction, and cognitive activation.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the teacher instructional practices theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Classroom management	2.3.1	TQ42i,j,k,l	N/A	4-point frequency scale
Clarity of instruction	2.3.1	TQ42a,b,c,d	N/A	4-point frequency scale
Cognitive activation	2.3.2	TQ42e,f,g,h	N/A	4-point frequency scale

There are many items regarding classroom management, clarity of instruction, and cognitive activation that reflect teachers' behaviour in the classroom. All these measures can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies. Another approach is to report classroom management, clarity of instruction and cognitive activation as scales, and the two items pertaining to project work and use of

information communications technology (ICT) as stand-alone items. All of the above items refer to teacher instructional practices, more specifically to teachers' instructional practices in the target class.

Estimates of change

Some measures regarding teachers' instructional practices are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures.

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item
Target class subject focus	2.3.1	TQ37				N/A
Target class size	2.3.1	TQ38				N/A
Target class time distribution	2.3.1	TQ39a,b,c				N/A
Engagement in collaborative activities	2.3.1	TQ33a,b,c,d,e,f,g,h			partial	N/A
Summarise lesson content	2.3.1	TQ42a				N/A
Small group work	2.3.2	TQ42g				N/A
Quieten the class for lesson	2.3.1	TQ42l				N/A
Connect lesson to life	2.3.2	TQ42m				N/A
Mastery of content	2.3.1	TQ42n				N/A
Project work	2.3.2	TQ42o				N/A
Use of ICT	2.3.2	TQ42p				N/A
Assessment and feedback	2.3.1	TQ43a,b,c,d,e				N/A

To compare changes across TALIS surveys, the TALIS 2013 statistics would need to be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

In addition, some opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses. The school leader is asked the following about teachers' views in the school on: teacher reliance on each other. This is expected to be correlated with TQ33a-h (collaborative activities). With this, statistical analyses could be performed with the principal questionnaire results to identify the extent to which there is dissonance between the school leaders' perception overall of their teachers and the actual variation among teachers in the school.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5
Teaching practices, overall	✓	Clarity of instruction scale	Cognitive activation scale	Classroom management scale		
Clarity of instruction	✓	TQ42a	TQ42b	TQ42c	TQ42d	
Cognitive activation	✓	TQ42e	TQ42f	TQ42g	TQ42h	
Classroom management	✓	TQ42i	TQ42	TQ42k	TQ42l	
Instructional assessment and feedback	✓	TQ43a	TQ43b	TQ43c	TQ43d	TQ43e

In addition, the configurations for the collaboration scales are outlined in Section 2.4. The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

The dimensions and indicators concerned with teachers' instructional practices may be reported by making two profiles of teachers' instruction:

- A profile of instructional teaching practices with respect to the dimensions of instructional quality (Decristan et al., 2015[21]; Fauth et al., 2014[22]; Klieme, Pauli and Reusser, 2009[23]):
  - classroom management as indicated by a positive disciplinary climate (as in TALIS 2008 and 2013)
  - clarity of instruction
  - cognitive activation
  - feedback to students
  - assessment strategies
  - lesson time spent on teaching and learning
- A profile of instructional practices with respect to 21st century instruction:
  - fostering students' cross-curricular skills (e.g. TQ42s, t)
  - accounting for equity and diversity in classrooms (e.g. TQ35, TQ45).

In response to the priority given to indicators of overall instructional teaching practices (2.3.1), TALIS 2018 included three sets of questions concerned with factors associated with teacher instructional practices. If stakeholders are interested in understanding how individual aspects relate to the overall measure of instructional teaching practices, then analyses would be suggested to correlate classroom management, clarity of instruction and cognitive activation to the overall scale of teachers' instructional practices (TQ42).

Moreover, the three aspects of teaching practices (classroom management, clarity of instruction and cognitive activation) should correlate to instructional assessment and feedback practices (TQ43a, b, c, d, e).

The strength and magnitude of the coefficients from these analyses can point data users to areas on which to focus deeper exploration. Deeper exploration could include analyses that allow for multi-collinear associations to identify the co-dependence (spuriousness) between similar scales. The results from such analyses could then inform stakeholders of the scales that are most central to the association with the overall measure. Associations that are weak or insignificant could, likewise, inform policy makers where efforts would not be well spent.

If data users are interested in understanding whether or not these research findings apply in a broader global context, then the hypotheses could be tested among TALIS participating countries and economies, such as:

	2.	3.	4.	5.
1. Overall teaching practices scale	+	+	+	
2. Classroom management scale	+	+	+	+
3. Clarity of instruction scale	+	+	+	
4. Cognitive activation scale	+	+	+	+
5. Instructional assessment and feedback	+	+	+	+

The statistical method with which to analyse these scale measures will depend on the level of invariance that is reached. These model specifications are described in detail in Section 4.

2.4. Teachers' professional practices

Theme priorities

The TGB expressed the desire to:

- 2.4.1) repeat and improve indicators on teacher collaboration and teachers' participation in school decision-making processes
- 2.4.2) develop new material on connection between teachers' readiness for and openness to diversity and collaboration among staff, as well as their views on the conditions and resources needed to foster deeper forms of collaboration in the school, between schools, and between the school and external stakeholders.

System-wide averages

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the teachers' professional practices theme.

New to 2018

For the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Collaborative professional learning	2.4.2	TQ26f	N/A	Yes/no
Collaboration during innovation	2.4.2	TQ32d	N/A	4-point agreement scale
Principal collaboration involvement and support	2.4.2	N/A	PQ22a,e,j	4-point frequency scale
Academic mobility purpose*	2.4.2	TQ56	N/A	Yes/no
Academic mobility total time abroad*	2.4.2	TQ58	N/A	3- point duration scale

*County national option (CNO) question

The single items listed above would be adequately represented by within-country/economy proportions as measures of central tendencies.

Estimates of change

Several measures regarding teachers' professional practices are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be done with the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Time distribution total hours		TQ16	✓	✓	✓	N/A
Time distribution teaching hours		TQ17	✓	✓	✓	N/A
Time distribution non-teaching hours		TQ18	✓	✓	✓	N/A
Participation in school management	2.4.1	TQ18e	✓	✓		N/A
Participation in school decisions	2.4.1	TQ48a	✓	✓		PQ26a,e
Distribution of responsibilities/ leadership	2.4.1	N/A	✓	✓		PQ20a,b,c,d,e,f, g,h,i,j,k
Collaborative school culture	2.4.1	TQ48e	✓	✓		PQ26f
Engagement in collaborative activities, overall	2.4.1 2.4.2	TQ33a,b,c,d,e,f,g,h	✓	✓	partial	N/A
Exchange and co-ordination for teaching	2.4.1 2.4.2	TQ33d,e,f,g	✓	✓	partial	N/A
Professional collaboration	2.4.1 2.4.2	TQ33a,b,c,h	✓	✓	partial	N/A
Academic mobility activities*	2.4.2	TQ57	✓	✓		N/A

*County national option (CNO) question

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Several opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses.

- The school leader is asked the following about teachers' views in the school on: participation in school decisions, distribution of responsibilities/leadership, collaborative school culture. With these, statistical analyses could be performed with the principal questionnaire results to identify the extent to which there is dissonance between the school leaders' perception of teacher involvement in decision making in the school and the culture of collaboration in the school and the views of teachers related to decision making and collaboration in the school.
- The school leader is also asked about the following: school leader decision making, distribution of responsibilities/leadership, principal collaboration involvement and support. These items allow for analyses to be supplemented by the school leader's perceptions if there is a preference among policy makers to gain a more comprehensive understanding of the relationship between the school leaders' views and the views of their teachers.

Scale configurations

Scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Engagement in collaborative activities, overall (TCOOP)	✓	TCEXCH scale	TCCOLL scale				
Exchange and co-ordination for teaching (TCEXCH)	✓	TQ33d	TQ33e	TQ33f	TQ33g		
Professional collaboration (TCCOLL)	✓	TQ33a	TQ33b	TQ33c	TQ33h		

The specifics of the statistical approaches to confirm scales as well as the appropriate statistics to report are outlined in Section 4.

Multivariate analyses recommendations

As noted in the conceptual framework, collaboration is a professional practice of high interest because it appears to play a role in various elements of teachers' work, including instructional practice, professional learning and job satisfaction (Ainley and Carstens, 2018[3]; Desimone, 2009[24]; Goddard, Goddard and Tschannen-Moran, 2007[25]; Timperley et al., 2007[26]). Collaboration appears to be associated with aspects of school culture, although the direction of causation is not clear.

Researchers have cautioned against simplistic interpretations of collaboration, including assumptions that increased collaboration will lead to improvements in student outcomes (Little, 1990[27]). Some researchers have distinguished between collaboration and collegiality. Kelchtermans (2006[28]), for example, suggests that collaboration describes co-operative actions, whereas collegiality refers to the quality of relationships among staff in schools. As part of the understanding of collaboration, it may be interesting to examine questions about the form of collaborative activities and about collegial staff relationships (TQ33a, b, c, d, e, f, g, h; TQ49e) to determine any associations these have with teachers' views of collaboration within their school (TQ48e).

The conceptual framework highlights that, while TALIS 2013 asked teachers about the frequency of their involvement in collaborative activities, it did not specifically focus on the ways that collaborative activity might shape teachers' professional work and their attitudes towards that work. Some researchers have proposed that a balance between autonomy and collaboration seems to provide a promising basis for examining the goals of teacher professional development and school improvement – see, for example, Kelchtermans (2006[28]). To understand collaboration, it might be interesting to examine questions related to teacher activities and time use and teacher autonomy (TQ18; TQ40; TQ48a).

To understand some relationships between collaboration and professional practices, as discussed in priorities 2.4.1 and 2.4.2, the scales measuring engagement in collaborative activities, exchange and co-ordination for teaching and professional collaboration are suggested as being central to the analyses. In addition, the analyses would incorporate the single items concerned with whether a professional learning activity with the greatest impact on teaching provided opportunities for collaborative learning, as well as incorporating aspects of working conditions (non-teaching time) and teacher background (academic mobility time abroad). These are listed as:

	2.	3.	4.	5.	Academic mobility time abroad (TQ58)	Non-teaching time (TQ18)
Engagement in collaborative activities scale	+	+	+	+	+	-
Exchange and co-ordination for teaching scale		+	+	+	+	-
Professional collaboration scale	+		+	+	+	-
Collaborative professional learning item (TQ26f)	+	+		+	+	-
Collaborative school practices (TQ 48, PQ26a,e,f)						

Note: The engagement in collaborative activities scale is made up of the items that contribute to both the exchange and co-ordination for teaching scale and the professional collaboration scale. Thus, analyses can be based on either the overall scale or the two subscales, but not the combination.

Analyses would focus on identifying factors that correlate with exchange and co-ordination for teaching and professional collaboration as well as the associations between these two constructs. Among these factors, the analyses would examine whether the amount of time used for "team work and dialogue with colleagues" and professional experience abroad, as

well as other aspects of background, explain differences in current teaching (e.g. subjects taught) and collaborative school practices including participation in school decisions.

2.5. Teacher feedback and development

Theme priorities

The TALIS Governing Board expressed the desire to:

- 2.5.1) repeat and improve indicators on sources and types of feedback provided to teachers and the perceived impact of this feedback
- 2.5.2) repeat and improve indicators on types of professional development activities, participation rates, intensity of participation, needs/demand, and frequency of mentoring and induction programmes
- 2.5.3) develop new material on sustained, collaborative and school-embedded (or contextualised) professional learning, on the connection between professional learning and innovation and on teachers' views on effective forms of professional learning; and new material on connections between teacher feedback and teacher development.

System-wide averages

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the teacher feedback and development theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Feedback types/sources	2.5.1	TQ29a,b,c,d,e,f	PQ24a,b,c,d,e,f	4-option non-exclusive choice
Feedback impact (filter)	2.5.1	TQ30	N/A	Yes/no
Feedback impact	2.5.1	TQ31a,b,c,d,e,f	N/A	Yes/no
Induction types/formats	2.5.2	TQ20a,b,c,d,e,f,g,h,i,j	PQ33a,b,c,d,e,f,g,h,i,j	Yes/no
Involvement in mentoring	2.5.2	TQ21a,b	(PQ34 – trend)	Yes/no
Professional development topics	2.5.2	TQ23a,b,c,d,e,f,g,h,i,j, k,l,m,n,o	N/A	Yes/no
Professional development incentives/support	2.5.2	TQ24c,d,e,f,g,h	N/A	Yes/no
Professional development needs	2.5.2	TQ27l,m,n	PQ08a,b,c,d,e,f,g,h,i,j,k	4-point extent scale
Professional development importance	2.5.2	TQ55g	N/A	3-point importance scale
Professional development impact (filter)	2.5.3	TQ25	N/A	Yes/no
Professional development impact	2.5.3	TQ26a,b,c,d,g,h,i,l	N/A	Yes/no

There are many items regarding feedback impact, induction types and formats, and professional development topics, incentives, supports, and impact. All these measures tally the presence or absence of the specific areas and topics and thus can be represented as proportion of teachers with affirmative ("yes") answers. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding teacher feedback and development are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be done with the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Teacher formal appraisal – agency/frequency	2.5.1		✓	✓	partial	PQ23a,b,c,d,e
Teacher formal appraisal – frequency of actions	2.5.1		✓	✓	partial	PQ25a,b,c,d,e,f,g,h
Teacher feedback to other teachers	2.5.1	TQ33b	✓	✓	*	
Formal induction target group	2.5.2		✓	✓	partial	PQ32
Mentoring availability for teachers	2.5.2		✓	✓	✓	PQ34
Mentoring subject field matching	2.5.2		✓	✓	✓	PQ35
Mentoring importance	2.5.2		✓	✓		PQ36a,b,c,d,e,f
Professional development activities participation hours	2.5.2	TQ18g	✓	✓	*	N/A
Participation in induction	2.5.2	TQ19a,b	*	*		PQ31a,b
Professional development types/formats	2.5.2	TQ22a,b,c,d,e,f,g,h,i,j	✓	partial	partial	PQ07a,b,c,d,e,f,g,h,i,j
Professional development incentives/support	2.5.2	TQ24a,b	✓	partial		
Professional development needs	2.5.2	TQ27a,b,c,d,e,f,g,h,i,j,k	✓	✓	partial	(PQ08 – new)
Professional development barriers	2.5.2	TQ28a,b,c,d,e,f,g	✓	✓	*	PQ09a,b,c,d,e,f,g
Professional development impact	2.5.3	TQ26e,f,j,k	✓	✓		N/A
Collaborative professional learning	2.5.3	TQ33h	✓	✓	✓	N/A

*Slight changes in wording in 2018 may prohibit comparisons to prior survey years.

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Several opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses.

- The school leader is asked the following about teachers' views in the school on: feedback types/sources, participation in induction, induction types/formats, involvement in mentoring. With these, statistical analyses could be performed with the principal questionnaire results to identify the extent to which there is dissonance

between the school leaders' perception of teacher feedback and development in the school and the views of teachers related to feedback and development in the school.

- The school leader is also asked about the following: professional development types/formats, professional development needs, professional development barriers. These items allow for analyses to be supplemented by the school leader's perceptions if there is a preference among policy makers to gain a more comprehensive understanding of the relationship between the school leaders' views and the views of their teachers.

Scale configurations

Scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6	Item7
Professional development needs [TPDPEDS]*	N/A	TQ27a	TQ27b	TQ27c	TQ27d	TQ27f		
Professional development for diversity [TPDDIV]	✓	TQ27h	TQ27i	TQ27j				
Effective professional development	N/A	TQ26a	TQ26b	TQ26c	TQ26d			
Professional development barriers	N/A	TQ28a	TQ28b	TQ28c	TQ28d	TQ28e	TQ28f	TQ28g

* Since the 2013 26L is not asked on the 2018 survey, this is no longer comparable to the 2013 scale.

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

To address priorities 2.5.1 and 2.5.2, TALIS 2018 expanded questions to understand the sources and types of feedback and professional development. It was anticipated that there would be interest in understanding which types and sources of feedback and professional development have the strongest perceived impact on teachers work. For this purpose, analyses are proposed to examine the relationships between types and sources of professional development and the impacts of professional development and feedback (TQ26, TQ31, respectively). The magnitude of the coefficients resulting from these analyses may point to areas in which to concentrate efforts to improve student learning (Yoon et al., 2007[29]).

As discussed in the conceptual framework, TALIS 2018 (Ainley and Carstens, 2018[3])provides opportunities to obtain information about the impact of teacher professional development and feedback at both teacher and institutional levels. Given priority 2.5.3, the links between feedback and teacher professional development and their roles in teachers' ongoing professional learning are of particular interest (Ainley and Carstens, 2018[3]; Isoré, 2009[30]; OECD, 2005[31]). Comparing international results from TALIS to those in the literature that are typically based on studies from individual countries (Desimone, 2009[24]; Hattie, 2009[32]; Ingvarson, Meiers and Beavis, 2005[33]; Timperley et al., 2007[26]) can add a wider perspective to the discussions about these topics.

As part of the understanding of the impact of professional development, it may also prove useful to investigate which barriers to (TQ28), or incentives for (TQ24), professional development were associated with types or formats (TQ22) and foci (TQ23) of professional development, as well as the expressed professional development needs (TQ27) among teachers.

The pilot phase of TALIS 2018 revealed that teachers in some countries and economies reported an absence of induction activities at the system or local school level. A lack of, or

limited, induction provision may, in conjunction with a consideration of data about related constructs, reveal new understandings about the impact of induction on teaching quality.

2.6. Teacher self-efficacy

Theme priorities

The TGB expressed the desire to:

- 2.6.1) repeat and improve indicators on overall teacher self-efficacy and its subscales (teaching and classroom management in general)
- 2.6.2) add new material on teachers' readiness for diversity.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the teacher self-efficacy theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Teacher self-efficacy in using technology for educational purposes	2.6.1	TQ34m	N/A	4-point extent scale
Teacher self-efficacy in multicultural environments	2.6.2	TQ45a, b, c, d, e	N/A	4-point extent scale

Although specified under the theme Diversity and equity, teacher self-efficacy in multicultural environments taps self-beliefs in specific instructional practices and, thus, supplements the general teacher self-efficacy measure – therefore, the former can be considered part of teacher self-efficacy. These measures can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding self-efficacy are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Teacher self-efficacy in classroom management	2.6.1	TQ34d, f, h, i	✓	✓	✓	N/A
Teacher self-efficacy in instruction	2.6.1	TQ34c, j, k, l	✓	✓	✓	N/A
Teacher self-efficacy in student engagement	2.6.1	TQ34a, b, e, g	✓	✓	✓	N/A
Overall teacher self-efficacy scale	2.6.1	compound scale; average of three scales: Teacher self-efficacy in classroom management, Instruction, and Student engagement	✓	✓	✓	N/A

To compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Given that the assessment of teacher self-efficacy refers only to individuals at the teacher level, thus representing a teacher-specific measure, triangulation with principals' or other school-level measures is not possible.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Teacher self-efficacy, overall	✓	Teacher self-efficacy in classroom management scale	Teacher self-efficacy in instruction scale	Teacher self-efficacy in student engagement scale			
Teacher self-efficacy in classroom management	✓	TQ34d	TQ34f	TQ34h	TQ34i		
Teacher self-efficacy in instruction	✓	TQ34c	TQ34j	TQ34k	TQ34l		
Teacher self-efficacy in student engagement	✓	TQ34a	TQ34b	TQ34e	TQ34g		

The configurations for the teacher self-efficacy in multicultural classrooms scales are outlined in Section 2.11. The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

Given priority 2.6.1 and the results from TALIS 2013 that demonstrated positive associations among the teacher self-efficacy scales, it is suggested that these associations be replicated. Examining the relationships between the three general self-efficacy scales (i.e. self-efficacy in classroom management, instruction, and student engagement) provides information about the coherence of the teacher self-efficacy construct and the extent to which self-beliefs in different instructional practices go together. From a measurement point of view, information about the associations is needed to establish valid associations to other constructs, such as job satisfaction.

Given priority 2.6.2, it is suggested that the relationships between teacher self-efficacy in instruction scale and teacher self-efficacy in multicultural environments scale – a measure

that refers to more specific instructional practices – be examined. The strengths of associations will provide policy makers with information about whether teachers generally lack self-efficacy in instruction (independent of the type of instruction) or whether teachers may lack self-efficacy in teaching culturally diverse classrooms, yet may compensate for this by their self-efficacy in general instructional practices. Moreover, the relationships among the three subscales of general teacher self-efficacy and self-efficacy in multicultural environments may uncover the extent to which aspects of self-efficacy in general instructional practices contribute to self-efficacy in teaching culturally diverse classrooms. For instance, teachers may lack self-efficacy in both classroom management and teaching culturally diverse classrooms.

Priority 2.6.2 further links the theme of Teacher self-efficacy with that of Diversity and equity.

	2.	3.	4.	5.
Overall teacher self-efficacy scale				+
Teacher self-efficacy in classroom management scale		+	+	+
Teacher self-efficacy in instruction scale			+	+
Teacher self-efficacy in student engagement scale				+
Teacher self-efficacy in multicultural environments scale				

The statistical method with which to analyse these scale measures will depend on the level of invariance that is reached. These model specifications are described in detail in Section 4.

2.7. Job satisfaction

Theme priorities

The TGB expressed the desire to:

- 2.7.1) repeat and improve indicators on overall job satisfaction and teacher perception of the value of the teaching profession
- 2.7.2) add new material on teachers' views on the factors that would increase their job satisfaction and perception of the value of the profession
- 2.7.3) improve material on the risk of job attrition
- 2.7.4) improve questions on teacher well-being and stress.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the job satisfaction theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Motivation to teach	2.7.1	TQ07a,b,c,d,e,f,g		4-point importance scale
Satisfaction with classroom autonomy	2.7.1	TQ40a,b,c,d,e		4-point agreement scale
Teacher reliance on each other	2.7.2	TQ49e	PQ26k	4-point agreement scale
Satisfaction with salary and benefits	2.7.2	TQ54a,b	PQ45a,b	4-point agreement scale
Feeling valued as a profession by various stakeholders	2.7.2	TQ54c,d,e		4-point agreement scale
Risk of job attrition; number of years continuing in profession	2.7.3	TQ50	PQ42	continuous
Workplace well-being	2.7.4	TQ51a,b,c,d		4-point extent scale
Workload stressors	2.7.4	TQ52a,b,c,d,e,f,g,h,i,j,k	PQ43a,b,c,d,e,f,g,h,i	4-point extent scale

There are many items regarding motivation to teach, feeling valued by various stakeholders, and workload stressors. All these measures (except TQ50/PQ42) can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding job satisfaction are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Value of teaching	2.7.1	TQ53h	✓	✓		PQ44h
Job satisfaction, overall scale	2.7.1	compound scale; average of two scales: "Satisfaction with this school" and "Satisfaction with the profession"	✓	✓		compound scale; average of two scales: "Satisfaction with this school" and "Satisfaction with the profession"
Satisfaction with this school scale	2.7.1	TQ53c,e,g,j	✓	✓		PQ44c,e,g,j
Satisfaction with the profession scale	2.7.1	TQ53a,b,d,f	✓	✓		PQ44a,b,d,f

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Several opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses.

- The school leader is asked the following about teachers' views in the school on: teacher reliance on each other. With this, statistical analyses could be performed with the principal questionnaire results to identify the extent to which there is

dissonance between the overall school leaders' perception of their teachers and the actual variation among teachers in the school.

- The school leader is also asked about the following: value of teaching, satisfaction with this school, satisfaction with the profession, satisfaction with salary and benefits, years continuing in the profession, workload stressors. These items allow for analyses to be supplemented by the school leader's perceptions if there is a preference among stakeholders to gain a more comprehensive understanding of the school leaders' views in relation to their teachers' views.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Job satisfaction, overall	✓	Satisfaction with this school scale	Satisfaction with the profession scale				
Satisfaction with this school -- teacher	✓	TQ53c	TQ53e	TQ53g	TQ53j		
Satisfaction with this school – principal	✓	PQ44e	PQ44g	PQ44i	PQ44j		
Satisfaction with the profession -- teacher	✓	TQ53a	TQ53b	TQ53d	TQ53f		
Satisfaction with the profession – principal		PQ44a	PQ44b	PQ44d	PQ44f*		
Satisfaction with classroom autonomy	N/A	TQ40a	TQ40b	TQ40c	TQ40d	TQ40e	
Workplace well-being	N/A	TQ51a	TQ51b	TQ51c	TQ51d		
Student behaviour stress – teacher	N/A	TQ52f	TQ52g	TQ52h			
Workload stress – Teacher	N/A	TQ52a	TQ52b	TQ52c	TQ52d	TQ52e	
Workload stress – Principal	N/A	PQ43a	PQ43b	PQ43c			
Personal utility motivation to teach	N/A	TQ07a	TQ07b	TQ07c	TQ07d		
Social utility motivation to teach	N/A	TQ07e	TQ07f	TQ07g			
Perceived value of teachers	N/A	TQ54c	TQ54d	TQ54e			

*PQ44f was not asked of the principal in TALIS 2013. For a 2018 comparison to principal satisfaction with profession to 2013, do not include PQ44f in the scale construction. To triangulate satisfaction with the profession in 2018 between teachers and principals, include item PQ44f.

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

Per the priorities of 2.7.2 and 2.7.4, TALIS 2018 included three sets of questions to understand the factors associated with workplace well-being and feeling valued as a teacher. If data users are interested in understanding how individual aspects relate to the overall measures of well-being and feeling valued, then analyses would be suggested to correlate well-being and feeling valued with: workplace stressors (TQ52a,b,c,d,e,f,g,h,i,j,k), teachers' ability to rely on each other (TQ49e), and stakeholders views of teachers (TQ54c,d,e). The magnitude of the coefficients from these analyses can point data users to areas on which to concentrate deeper exploration. Deeper exploration could include analyses that allow for multi-collinear associations to identify the co-dependence between similar items. The results from such analyses could then inform

stakeholders about the items that are most central to the association with the overall measures.

Given priority 2.7.1 and the results from TALIS 2013 that demonstrated that job satisfaction and feeling valued as a teacher related to teacher and classroom characteristics (OECD, 2013[34]), it is suggested that, similarly, the new 2018 indicators of classroom autonomy, satisfaction with salary and benefits, well-being, and risk of job attrition be analysed. The conceptual framework describes existing research that shows positive associations between satisfaction measures and teaching work attributes (e.g. classroom autonomy, teaching duties) and negative associations with non-teaching aspects of work (e.g. stress, salary) (Ainley and Carstens, 2018[3]; Butt et al., 2005[35]; Crossman and Harris, 2006[36]; Dinham and Scott, 1998[37]; Stearns et al., 2015[38]). Initial motivations for teaching are also expected to relate to differences in satisfaction with the profession (Karabenick and Urdan, 2010[39]; Richardson and Watt, 2010[40]; Watt and Richardson, 2008[19]). If data users are interested in understanding whether or not these research findings apply in a broader global context, then the hypotheses could be tested among TALIS participating countries and economies, for example:

	Satisfaction with classroom autonomy scale	Satisfaction with salary and benefits items: TQ54a,b	Workplace well-being scale	Motivation to teach: social or personal subscale
Overall job satisfaction scale	+	+	+	
Satisfaction with the profession scale	+	+	+	+
Satisfaction with this school scale	+	+	+	
Feeling valued (TQ53h)	+	+		

The statistical method with which to analyse these scale measures will depend on the level of invariance that is reached. These model specifications are described in detail in Section 4.

2.8. School leadership

Theme priorities

The TGB expressed the desire to:

- 2.8.1) repeat and improve indicators on school leadership style, role, style
- 2.8.2) develop new material on distributed and teacher leadership and on teachers' perception of school leadership.

TALIS 2018 conceptualises school leadership in terms of four key dimensions:

1. who principals are: qualifications, recruitment and development of principals
2. what principals do: role, function and work of the principal, instructional leadership, terms and conditions, workload, hours, autonomy, function, and actions
3. distributed leadership and teacher leadership and teachers' perceptions of these types of leadership
4. system leadership and leadership in networks of schools.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the job satisfaction theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
System leadership	2.8.2	N/A	PQ22j	4-point scale
Distributed leadership	2.8.1	TQ48b,c,h	N/A	

Estimates of change

Several measures regarding job satisfaction are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Principal demographic characteristics	2.8.1	N/A	✓	✓	✓	PQ1,2
Principal qualifications and experience	2.8.1	N/A	✓	✓	✓	PQ3,4,5,6
School management team	2.8.2	N/A	✓	✓	✓	PQ18,19
Responsibility for staffing issues scale	2.8.2	N/A	✓	✓	✓	PQ20a,b
Responsibility for budget scale	2.8.2	N/A	✓	✓	✓	PQ20c,d,e
Responsibility for educational policies scale	2.8.2	N/A	✓	✓	✓	PQ20f,g,h,i,j,k
Principal time use	2.8.2	N/A	✓	✓	✓	PQ21
Principal activities	2.8.2	N/A	✓	✓		PQ22
Instructional leadership scale	2.8.2	N/A	✓	✓		PQ22a,b,d,e,f,g,i
Distributed leadership (teacher and principal)	2.8.2	TQ48a	✓	✓		PQ26a,b,c

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Triangulation is possible with the teacher and principal questionnaire survey responses regarding the distributed leadership items about the decision-making opportunities for parents or guardians, students, and staff in general and, in particular, with leading new initiatives.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
School autonomy for staffing	✓	PQ20a	PQ20b				
School autonomy for budget	✓	PQ20c	PQ20d	PQ20e			
School autonomy for educational policies	N/A	PQ20f	PQ20g	PQ20h			
School autonomy for curriculum	✓	PQ20i	PQ20j	PQ20k			
School autonomy for instructional policies	✓	PQ20f	PQ20g	PQ20j	PQ20k		

Instructional leadership	✓	PQ22d	PQ22e	PQ22f			
Lack of personnel [PLACKPER]	✓	PQ29a	PQ29b	PQ29c			
Lack of resources [PLACKMAT]	✓	PQ29d	PQ29e	PQ29f	PQ29g		
Distributed leadership – principal	✓	PQ26a	PQ26b	PQ26c			
Distributed leadership – teacher	✓	TQ48a	TQ48b	TQ48c			
Participation among stakeholders [PSCSTAKE] – principal	✓	PQ26a	PQ26b	PQ26c	PQ26d	PQ26f	
Participation among stakeholders [TSCSTAKE] – teacher	✓	TQ48a	TQ48b	TQ48c	TQ48d	TQ48e	

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

The information that TALIS 2018 collects on principals' characteristics, including age, gender, and previous experience and training (PQ1 to PQ6), will enable development and analysis of principal types across countries and economies, as well as exploration of how these types relate to aspects of leadership, teaching and learning environments. Principal characteristics such as experience or training can be related to levels of distributed and instructional leadership, for example. A key question concerns the associations between and among principal qualifications and leadership development. This question was an important one for TALIS 2013 and remains so for TALIS 2018, especially given evidence of the differential distribution of highly qualified principals across schools (Boyd et al., 2011[41]).

The extent of autonomy that principals exercise in their role (the three scales responsibility for staffing issues, responsibility for budget and responsibility for educational policies) were found to differ significantly between countries and economies in TALIS 2013 and it is likely that this is still the case. These scales could be usefully related to principal activities and time use, which are likely to depend, at least partially, on autonomy (Pashiardis, Brauckmann and Muijs, 2011[42]).

The statistical method with which to analyse these scale measures will depend on the level of invariance that is reached. These model specifications are described in detail in Section 4.

2.9. School climate

Theme priorities

The TGB expressed the desire to:

- 2.9.1) repeat the study's indicators of student-teacher relationships and classroom disciplinary climate
- 2.9.2) add new material designed to capture teachers' views on the school climate conditions that foster effective teaching and learning (e.g. teacher leadership structure)
- 2.9.3) integrate aspects of equity, diversity and innovation into the existing instruments.

System-wide averages

These priorities and preferences are linked to the final teacher survey instrument and can be used to describe the system-wide averages of ideas related to the school climate theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Teacher-teacher trust	2.9.2	TQ49e	PQ26k	4-point agreeance
Common teaching beliefs	2.9.2	TQ48f	PQ26g	4-point agreeance
Resource shortages	2.9.2	N/A	PQ29j,k,l,m,n,o	4-point extent
Cyber/electronic bullying	2.9.2	N/A	PQ30f,g	5-point frequency
Academic press	2.9.2	N/A	PQ27a,b,c,f	4-point extent
Parent-community involvement	2.9.2	N/A	PQ27d,e,g	4-point extent
Climate of shared rules	2.9.2 2.9.3	TQ48g	PQ26h	4-point agreeance
Climate of teacher initiative	2.9.2 2.9.3	TQ48h	PQ26i	4-point agreeance

There are many items regarding resource shortages. If the preference is to provide data users with a more concise report of averages, an example of a condensed approach of reporting averages could be to highlight the item cited most frequently per country/economy, the item cited as the most common across all countries and economies, and/or the item cited the least frequently per country/economy.

Estimates of change

Several measures of school climate are repeated from prior TALIS cycles. Comparisons of the averages and variation between survey years can be executed with the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Classroom disciplinary climate	2.9.1	TQ41a,b,c,d	✓	✓	✓	N/A
Student-teacher relations scale	2.9.1	TQ49a,b,c,d	✓	✓		PQ26j,k
Participation among stakeholders scale	2.9.2	TQ48a,b,c,d,e	✓	✓		PQ26a,b,c,d, f
Lack of resources [PLACKPMAT]	2.9.2	N/A	✓	✓	✓	PQ29d,e,f,g
Lack of personnel [PLACKPER]	2.9.2	N/A	✓	✓		PQ29a,b,c
Student delinquency scale	2.9.2	N/A	✓	✓	✓	PQ30a,b,c,d

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Several opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses. The principal was asked about teachers in the school: student-teacher relations, participation among stakeholders, teacher-teacher trust, common teaching beliefs, climate of shared rules, and climate of teacher initiative. Therefore, the analyses listed above can be triangulated with the principal questionnaire results to identify the extent to which there is dissonance between the overall school leaders' perception of their teachers and the actual variation among teachers in the school.

Scale configurations

Scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6	Item7
Academic press	N/A	PQ27a	PQ27b	PQ27c	PQ27f			
Parent-community involvement	N/A	PQ27d	PQ27e	PQ27g				
Classroom disciplinary climate	✓	TQ41a	TQ41b	TQ41c	TQ41d			
Student-teacher relations scale	✓	TQ49a	TQ49b	TQ49c	TQ49d			
Participation among stakeholders scale – teacher	✓	TQ48a	TQ48b	TQ48c	TQ48d	TQ48e		
Participation among stakeholders scale – principal	N/A	PQ26a	PQ26b	PQ26c	PQ26d	PQ26f		
Material resource shortages	N/A	PQ29d	PQ29e	PQ29f	PQ29g	PQ29i	PQ29j	PQ29m
Special needs personnel shortages	N/A	PQ29b	PQ29k	PQ29l				
Lack of personnel	✓	PQ29a	PQ29b	PQ29c				
Lack of resources	✓	PQ29d	PQ29e	PQ29f	PQ29g			
Student delinquency scale	✓	PQ30a	PQ30b	PQ30c	PQ30d			

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

Per the priorities of 2.9.2, TALIS 2018 added questions to more deeply understand the dimensions of school climate. If stakeholders are interested in understanding the academic and community aspects of school climate, then analyses would be suggested to correlate these measures with resource shortages (PQ29a,b,c,d,e,f,g,h,i,j,k,l,m,n,o) and common teaching beliefs (PQ26g), and the school-level mean score of classroom disciplinary climate items (TQ41a,b,c,d).

The addition of the measures of the teachers' perception of shared rules and perception about taking initiative allows for triangulation analyses at the school-level to compare the teachers' perceptions to that of the school leader. Moreover, understanding the associations of teachers' level of trust in relation to perceptions about common teaching beliefs, shared rules, taking initiative, disciplinary climate, student-teacher and stakeholder relations may serve well if stakeholders seek to identify where trust between teachers intersects with various dimensions of school climate.

Lastly, the theme of school climate asks questions of teachers' perceptions of their target classroom disciplinary climate. This provides an opportunity for data users to test the strength of the (expected) inverse relationships between teachers' perceptions of classroom disciplinary climate in relation to their perceptions of the overall school climate of teacher trust, common teaching beliefs, shared rules, taking initiative, disciplinary climate, student-teacher and stakeholder relations.

2.10. Innovation

Theme priorities

The TGB expressed the desire to:

2.10.1) add new material and refine existing material to measure the degree to which innovation is implemented in learning environments (classrooms)

2.10.2) add new material to measure conditions for innovation in schools and in education systems manifested by, for example, teachers' and principals' views on the school conditions needed to foster innovation.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the innovation theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	Theme priority	TQ item(s)	PQ item(s)	Type of response
Teacher self-efficacy in using technology	2.10.1	TQ34m	N/A	4-point extent scale
Inclusion of cross-curricular skills and technology in teacher education	2.10.2	TQ6g,h (part A)	N/A	Dichotomous inclusion scale
Teacher preparedness to teach cross-curricular skills and use technology	2.10.2	TQ6g,h (part B)	N/A	4-point extent scale
Inclusion of cross-curricular skills in professional development	2.10.2	TQ23k	N/A	Dichotomous inclusion scale
Teachers' perceived needs for professional development in teaching cross-curricular skills	2.10.2	TQ27k	N/A	4-point extent scale
Team innovativeness	2.10.2	TQ32a,b,c,d	N/A	4-point agreement scale
Organisational innovativeness	2.10.2	N/A	PQ28a,b,c,d	4-point agreement scale

All these measures of aspects of innovation can be represented from central tendencies to more detailed distributions. If the preference is to provide data users with a more concise report of statistics, an example of a general central tendency statistic could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding innovation are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures:

Measure	Theme priority	TQ item(s)	2018	2013	2008	PQ item(s)
Instructional practices to use technology for educational purposes	2.10.1	TQ42p	✓	✓		N/A
Inclusion of teaching with technology in professional development	2.10.2	TQ23e	✓	✓		N/A
Teachers' perceived needs for professional development in using technology	2.10.2	TQ27e	✓	✓	✓	N/A

To compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Only limited opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses. The school leader is also asked about the following: organisational innovativeness. These items allow analyses of teachers' perceptions of team innovativeness within the school to be supplemented by the school leaders' perceptions if there is a preference to pursue a more comprehensive understanding of the relationship of the school leaders' views in relation to their teachers.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Team innovativeness	N/A	TQ32a	TQ32b	TQ32c	TQ32d		
Organisational innovativeness	N/A	PQ28a	PQ28b	PQ28c	PQ28d		

The specifics of the statistical approaches to confirm scales, as well as the appropriate statistics to report, are outlined in Section 4.

Multivariate analyses recommendations

To address priority 2.10.1, TALIS 2018 includes a measure of instructional practices that includes technology (TQ42p). This measure supplements the items and scales measuring further, more general instructional practices and provides information about the enacted use of digital technology for teaching purposes – the latter is of relevance, given the trend toward digitalisation of teaching and learning processes in education systems around the world. If data users are interested in the extent to which teacher education and professional development in this area might be related to the enacted use of technology for educational purposes, the single-item measure (TQ42p) can be related to the inclusion items (TQ6g, TQ23e). Furthermore, teachers' perceptions of their preparedness to use technology in classrooms (TQ6g) and their self-efficacy (TQ34m) may further explain variation in the extent to which teachers use technology during instruction (TQ42p). Information about the strengths of associations could point to the need for interventions that specifically target strengthening teachers' mastery experiences and, thus, self-efficacy in using technology. Moreover, the results of these analyses can be combined with teachers' expressed needs for professional development in this area (TQ27e).

To address priority 2.10.2, TALIS 2018 includes several measures, two of which represent indicators of the schools' climates of innovation. Given that these two indicators are assessed by teachers' and principals' perceptions, it is recommended that their relations at the school level be examined. This necessitates the aggregation of teacher perceptions of team innovativeness – a step that provides relevant information on the extent to which teachers agree on the perceived level of team innovativeness within a school. More importantly though, the relation between the team innovativeness scale (TQ32) and organisational innovativeness scale (PQ28) provides policy makers with information about how facilitating conditions at the school level (PQ28) might impact enacted innovation in teacher teams (TQ32). This conditions-enactment link may point to the need for openness to change and innovation in schools as organisations. The associations between the innovativeness scales and instructional practices shed light on this link even further.

	2.	3.	4.	5.	6.
Team innovativeness scale (TQ32)	+	+		+	+
Organisational innovativeness scale (PQ28)		+		+	+
Instructional practices to use technology (TQ42p)			+	+	+
Inclusion of cross-curricular skills and technology in teacher education (TQ6g, h) and professional development (TQ23e, k)				+	+
Teachers' perceived needs for professional development in teaching cross-curricular skills and using technology (TQ27e, k)					+
Instructional practices to use technology for educational purposes (TQ42p)					

The statistical method with which to analyse these scale measures will depend on the level of invariance that is reached. These model specifications are described in detail in Section 4.

2.11. Equity and diversity

Equity and diversity is a new theme in TALIS 2018. Items are both in the teacher and principal questionnaires and refer to both attitudes/preferences and practices.

Theme priorities

Equity and diversity were originally considered to be encapsulated in the substance of each of the nine themes, but consultation with TALIS participants and policy stakeholders led to this stand-alone theme emerging as a theme of high contemporary importance.

System-wide statistics

These priorities and preferences are linked to the final teacher survey instrument and can be used to compute and report system-wide statistics that describe ideas related to the job satisfaction theme.

New to 2018

For each of the following measures, new to TALIS 2018, system-level statistics can be computed. Statistically significant differences among countries and economies could also be highlighted for individual items, keeping in mind that the results from TALIS are not to be construed as league tables.

Measure	TQ item(s)	PQ item(s)	Type of response
Target class student immigrants	TQ35g,h	PQ17d,e	5-point proportion
Teaching in multicultural classroom (filter)	TQ44	PQ37	Yes/no
Self-efficacy in multicultural classrooms	TQ45a,b,c,d,e	N/A	4-point extent scale
Diversity practices (filter)	TQ46	N/A	Yes/no
Diversity practices	TQ47a,b,c,d	PQ38a,b,c,d	Yes/no
Diversity policies	N/A	PQ39a,b,c,d	Yes/no
Diversity beliefs	N/A	PQ40a,b,c,d	4-point quantity scale
Equity beliefs	N/A	PQ41a,b,c,d	4-point quantity scale

All these measures of aspects of equity and diversity can be represented, from central tendencies to more detailed distributions. If the preference is to provide data users with a concise report of statistics, an approach could be to highlight the item cited most frequently per country/economy and the item cited as the most common across all countries and economies.

Estimates of change

Several measures regarding equity and diversity are repeated from prior TALIS cycles. Comparisons of the statistics and change between survey years could be performed using the following measures:

Measure	TQ item(s)	2018	2013	2008	PQ item(s)
Target class student characteristics	TQ35a,b,c,d,e,f	✓	✓	partial	PQ17a,b,c
Target class special needs focus	TQ36	✓	✓		N/A

In order to compare changes across TALIS surveys, the TALIS 2013 statistics would be mirrored. To do this, the coverage needs to be comparable within- and between-systems, the details of which are described in Section 4.

Triangulation

Several opportunities exist to triangulate the teacher survey responses with the principal questionnaire survey responses.

- Both the school leader and the teachers were asked about diversity practices, which allows for analyses to be supplemented by the school leader's perceptions in relation to their teachers' perceptions.
- Comparing teachers' target class diversity levels to that of the whole school is also possible using data from the school survey data in comparison to the teacher data.

Scale configurations

2018 scale	2013 scale	Item1	Item2	Item3	Item4	Item5	Item6
Self-efficacy in multicultural classrooms	N/A	TQ45a	TQ45b	TQ45c	TQ45d	TQ45e	
Diversity practices (teacher)	N/A	TQ47a	TQ47b	TQ47c	TQ47d		
Diversity practices (school)	N/A	PQ38a	PQ38b	PQ38c	PQ38d		
Diversity policies	N/A	PQ39a	PQ39b	PQ39c	PQ39d		
Diversity beliefs	N/A	PQ40a	PQ40b	PQ40c	PQ40d		
Equity beliefs	N/A	PQ41a	PQ41b	PQ41c	PQ41d		

The specifics of the statistical approaches to confirm scales as well as the appropriate statistics to report are outlined in Section 4.

Multivariate analyses recommendations

As all scales are new, reporting of univariate statistics will be important. However, there is an important issue that has troubled the development and administration of the equity and diversity questionnaire in some countries and economies where equity and diversity is a sensitive topic. As a consequence, any country-by-country presentation of results can trigger adverse reactions as the normativity of the topic is large and countries and economies may not want to be located in the tail of the distribution. This sensitivity is probably stronger for attitudinal than for practice questions. In addition, psychometric adequacy (internal consistencies and invariance) has to be documented clearly. This sensitivity has to be factored into all data reports.

The questionnaires offer much opportunity for within-theme analysis. First, correlations between the scale scores are relevant both for principals and teachers. Second, teacher and principal data address related topics. As a consequence, correlations can be computed at the school level. These correlations can be computed in each participating country/economy separately as well as combined (note, that associations between variables

will not face the same sensitivity as reports of univariate statistics, notably averages). There is very little literature on the associations of teachers and their principals vis-à-vis equity and diversity. If there is a specific and well-communicated policy with regard to equity and diversity, positive correlations can be expected. The absence of such correlations could suggest the lack of a clear policy or lack of endorsement of such a policy. Third (possibly most controversial), are associations at the country/economy level where institutions could try to understand country/economy variations in equity and diversity; again, countries and economies may prefer not to be compared.

It can be concluded that the theme of equity and diversity provides considerable scope for triangulating teacher and principal data. Their correspondence would provide information about the shared perception of preferences and practices among both parties. In these analyses, the school is the unit of analysis and within-country/economy correlations provide information about this correspondence in each country/economy. By aggregating data at the country/economy level, a very different and equally interesting perspective is obtained, namely how these perspectives are correlated at the country/economy level.

3. Between-theme analyses

This section revisits the conceptual links between themes based on the research discussed in the conceptual framework (Ainley and Carstens, 2018[3]). It provides recommendations for analytic procedures to test relationships between thematic indicators using correlational, cross-tabulation, and regression-based approaches, as informed by existing research, TALIS 2013 results (OECD, 2014[17]), and current educational policy issues of high importance to the TGB. This section also includes a reference look-up table that identifies the most appropriate aggregation unit for analysis (teacher-, school-, or system-level) based on the item construction and the policy questions of interest.

3.1. Teacher background and initial teacher education

In addition to the within-theme analysis of factors related to various measures of teacher background and initial teacher education, existing research points to the fact that teachers have great influence on instructional practices (Hattie, 2009[32]; Kyriakides, Christoforou and Charalambous, 2013[43]). Relations vary in strength though. Furthermore, teacher background and initial teacher education represent antecedents of outcomes, such as teacher self-efficacy and job satisfaction (Holzberger, Philipp and Kunter, 2014[44]; OECD, 2013[34]; Tschannen-Moran and Woolfolk Hoy, 2007[45]), and perhaps also risk of job attrition. The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

Ability to describe and compare the composition of the teaching force across countries and economies relies on having at hand information about teacher background in terms of age, gender, employment status, and job experience. At the system-level, differences in teacher background by initial teacher education could be described. Within-system variability of these characteristics is likely large and may resemble more descriptive clusters than linear relationships to other variables. Therefore, it may be a good first step to examine whether such descriptive clusters exist (e.g. with respect to teacher education level and major).

Cross-theme recommendations

Given the TALIS 2013 findings and the research discussed in the conceptual framework (Ainley and Carstens, 2018[3]), it may be of interest to data users to analyse the relationship of selected teacher background variables (e.g. gender [TQ01], years of teaching experience [TQ11]) and of selected initial teacher education characteristics (e.g. teacher qualifications [TQ06], educational level [TQ03], teacher qualification pathway [TQ04], teacher subject STEM [TQ15], teacher subject Humanities [TQ15] and teachers' sense of preparedness [TQ06 part B]) to instructional practices (e.g. classroom management [TQ42], student project work [TQ42o], ICT use in instruction [TQ42p]) but also to outcomes such as teachers' job satisfaction (TQ53) and the risk of job attrition (TQ53h), teachers' self-efficacy (TQ34), possibly also on the subscale level, and with respect to specific challenges such as multicultural classrooms (TQ45). As pointed out, teacher qualifications and initial education are generally known to influence these outcomes. Given the research discussed in the conceptual framework, an analysis of teacher background and initial teacher education may also be of policy interest in relation to professional development needs (TQ27) and professional development topics (TQ23).

Recent research (Blömeke et al., 2018[46]) shows that, for some countries and economies, job motivation (TQ07) significantly predicts job satisfaction (TQ53; two subscales), teacher self-efficacy (TQ34; three subscales) and risk of job attrition (TQ50). Such regression analyses are also recommended for TALIS 2018. Furthermore, they may be expanded to teaching as a first choice career (TQ8) as predictor, on the one hand, and instructional practices (TQ42; three subscales) as a dependent variable, on the other hand.

3.2. Human resources

While teacher human resource issues and stakeholder relations are generally found to have only an indirect association with student outcomes, they can significantly affect recruiting teachers to the profession and retaining them once there. They can also significantly affect teachers' job satisfaction and teachers' teaching and learning conditions. As such, they form a part of the school context that creates conditions for educational effectiveness, as shown in various educational effectiveness models – see, for example, Reynolds et al. (2014[47])

Teacher- or school-characteristic averages

Due to the diversity of the teacher workforce, there are likely to be differences in human resources by such demographics as teachers' age or certification type. Cross-tabulations by school type may also reveal differences in human resources if there are system-level differences in societal perceptions of public versus private schools.

Cross-theme recommendations

Cross-theme relationships: all aspects of human resources and stakeholder relations are important to understand. For example, relations between teachers and between teachers and school leaders may associate with available time not spent on teaching and lesson preparation.

Both teachers' views on how they are valued and workload may be related to job satisfaction, self-efficacy and to teacher demographic characteristics, and may influence teachers' likelihood of staying in the profession. The extent to which teachers feel valued may also be related to innovation potential.

Educational reform priorities may be influenced by teacher characteristics such as experience, but may also be related to professional practices. There is likely to be a correlation with views on diversity and equity.

3.3. Instructional practices

In addition to the within-theme analysis of factors related to various measures of instructional practices, research discussed in the conceptual framework describes strong associations of teachers' instructional practices with teacher education and initial preparation, including professional development (Baumert et al., 2010[48]; Klieme, Pauli and Reusser, 2009[23]), school climate (Kyriakides and Creemers, 2008[49]; Scherer and Nilsen, 2016[50]) and teacher self-efficacy (Holzberger, Philipp and Kunter, 2014[44]; OECD, 2013[34]; Tschannen-Moran and Woolfolk Hoy, 2007[45]). The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

Given the TALIS 2013 findings and the research discussed in the conceptual framework, it may be of interest to data users to analyse the relationship of selected instructional practices (e.g. classroom management [TQ42], project work [TQ42o], ICT use in instruction [TQ42p]) and teacher characteristics (e.g. gender [TQ01], years of teaching experience [TQ11], teacher subject STEM [TQ15], teacher subject Humanities [TQ15], preparedness for instructional practices [TQ6 part B]).

Cross-theme recommendations

Given priority 2.3.1 and the research discussed in the conceptual framework, an analysis of instructional practices (TQ42 and TQ43a,b,c,d,e) in relation to items related to professional development needs (TQ27), professional development topics (TQ23), teacher qualifications (TQ6), educational level (TQ3), and self-efficacy (TQ34) may be investigated out of policy interest. This is because teacher qualifications and characteristics are generally known to influence their instructional practices, e.g. Klieme, Pauli and Reusser (2009[23]).

In addition, school climate has been shown to interact with teachers' instructional practices, e.g. Kyriakides and Creemers (2008[49]). Hence, instructional practices (TQ42 and TQ43a,b,c,d,e) may be related to disciplinary climate (TQ41), student-teacher relations (TQ49), academic focus (PQ27), and school safety (PQ30) at the school level. At the teacher level, classroom management scale should further correlate with target class disciplinary climate (TQ41).

3.4. Teachers' professional practices

In addition to the within-theme analysis of factors related to the range of professional activities that comprise teachers' professional practices, research discussed in the conceptual framework describes associations of teachers' professional practices with teacher professional development, instructional practices, school leadership, school culture, innovation, motivation, job satisfaction and self-efficacy.

Key cross-theme questions related to the Teachers' professional practices theme identified in the conceptual framework are:

- What connections exist between collaboration and development? Do teachers perceive collaboration to be a feature of effective professional development? Does collaboration stimulate further teacher professional development?
- In what ways does collaboration stimulate and support innovation in teaching practice?

Teacher- or school-characteristic averages

At the teacher level, differences in engagement in collaborative activities (collaborative activities scale), collaborative professional learning (TQ26f), and collaboration during innovation (TQ32d), may vary by teacher demographics, such as age, education level, and experience levels teachers.

At the school level, differences in perceptions of a collaborative school culture (TQ48e), teacher participation in school decisions (TQ48a), and non-teaching working hours, could be associated with school type (public/private) and principal leadership style.

Cross-theme recommendations

Firestone and Pennell (1993[51]) argued that autonomy is central to teachers' motivation. They reported that a reduction in teacher autonomy can be associated with teachers becoming dissatisfied with teaching. Therefore, an analysis of the relationship between collaboration and questions related to teacher motivation and job satisfaction may be of interest.

Also of interest are relationships between the forms of collaborative activity in which teachers engage (TQ33) and how participation in these activities is associated with various elements of teachers' work, including instructional practices, professional learning, decision making, and job satisfaction.

As noted in the conceptual framework, there is interest in gaining a better understanding of the emphases placed on teacher leadership in schools and the extent to which synergies exist between teachers' and principals' views of decision-making processes in schools. Together with comparisons between teacher and principal views in this area, it would be interesting to explore school-level relationships with questions related to school leadership, school climate and innovation.

Specific questions related to teacher academic mobility are offered as a national option in TALIS 2018. Interest in academic mobility is high in some countries and economies because of its perceived benefits relating to teacher learning and teaching quality, as well as professional characteristics such as motivation, job satisfaction and self-efficacy. Examination of relationships between academic mobility and questions related to instructional practices, professional development, innovation, motivation, job satisfaction and self-efficacy would be interesting to explore.

3.5. Teacher feedback and development

As noted in the conceptual framework, the associations between teaching quality and student learning outcomes are well documented in the research literature – see, for example, (Darling-Hammond, 2000[52]; Hattie, 2009[32]; Jensen et al., 2016[53]; Rowe, 2003[54]; Wenglinsky, 2002[55]) – and there is interest in identifying ways that teachers' professional development and feedback can contribute to teacher learning and improved instruction. TALIS 2018 provides opportunity to investigate these relationships at both teacher and institutional levels.

Key cross-theme questions related to the teacher feedback and development theme identified in the conceptual framework are:

- What forms of feedback, and what forms of professional development, do teachers perceive as having an impact on their teaching and other aspects of their professional practices (e.g. job satisfaction and motivation, self-efficacy, instructional practices and school climate)?
- In what ways does professional development stimulate and support innovation in teaching and learning?

Teacher- or school-characteristic averages

Teachers' perceptions of their professional development needs, and the impact of their professional development, may be influenced by teacher demographics such as age, educational level, and experience, as well as the characteristics of their classes (e.g. subject domain and student composition).

Teachers' views of professional development incentives and barriers, as well as their involvement in mentoring, may be associated with school resource needs. Their perceptions of the impact of feedback may be influenced by principal type, school decision-making processes and characteristics of the school student body.

Cross-theme recommendations

It can be inferred from some research literature that professional development experiences can motivate, inform, and support the development of teachers' instructional practices (Ingvarson, Meiers and Beavis, 2005[33]). Therefore, the relationship of teacher professional development with instructional practices would be a fruitful topic to explore. The literature also argues that contextual factors at classroom, school and system levels can either facilitate or hinder teachers' ability to engage in professional experimentation and develop their instructional practices (Clarke and Hollingsworth, 2002[56]). Relationships of teacher professional development with principal leadership, school climate and school resources would also be of interest.

Given priority 2.5.3, and the research discussed in the conceptual framework, an analysis on the relationship between collaborative professional learning (TQ22, TQ33) and innovation would be strongly recommended.

The ways that different types of feedback affect teaching and other aspects of teachers' working lives (e.g. self-efficacy and job satisfaction at the teacher level, and school climate at the institutional level) are of interest.

3.6. Teacher self-efficacy

In addition to the within-theme analysis of factors related to various measures of teacher self-efficacy, research discussed in the conceptual framework describes strong associations of teacher self-efficacy with instructional practices (Holzberger, Philipp and Kunter, 2014[44]), school climate (Hsu, Hou and Fan, 2011[57]), job satisfaction (Skaalvik and Skaalvik, 2010[58]), teacher education, feedback, and training (Klassen and Chiu, 2010[59]). Generally, these associations tap the relations between teacher self-efficacy, the classroom environment, the school environment, teacher background and satisfaction. The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

At the teacher level, differences in self-efficacy in general instructional practices scale (TQ34) and in teaching culturally diverse classrooms scale (TQ45) could be described by associating these measures with teachers' gender (TQ01), age (TQ02), years of experience in the teaching profession (TQ11), focus of teacher education (TQ06 part A), and their perceived preparedness (TQ06 part B). At the school level, the differences in teacher self-efficacy could be described by associating these measures with school type (public or private) (PQ12), the provision of induction and mentoring (PQ31, PQ34), and the schools' policies and practices to deal with diversity and equity (PQ39).

Cross-theme recommendations

Given the priorities 2.6.1 and 2.6.2, the research discussed in the conceptual framework, and the previous TALIS findings, an analysis of teacher self-efficacy in relation to items related to: class size (TQ38) and composition (TQ35), job satisfaction (e.g. TQ54), teacher-student relations (TQ49), teacher collaboration (TQ33), instructional practices

(TQ41, TQ42), self-efficacy in multicultural classrooms (TQ45), diversity practices (TQ47) and general school climate (TQ48). These variables and themes represent possible sources of teacher self-efficacy (Tschannen-Moran and Hoy, 2001[60]).

3.7. Job satisfaction

In addition to the within-theme analysis of factors related to various measures of job satisfaction, research discussed in the conceptual framework describes strong associations of job satisfaction with school climate (Collie, Shapka and Perry, 2012[61]; Miller, Brownell and Smith, 1999[62]; Stearns et al., 2015[38]; Weiss, 1999[63]), self-efficacy (Caprara et al., 2003[64]; Klassen et al., 2009[65]; OECD, 2013[34]; Tschannen-Moran and Woolfolk Hoy, 2007[45]) and school conditions (National Academy of Education, 2008[66]; OECD, 2013[34]). The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

Given the TALIS 2013 findings and the research discussed in the conceptual framework, it may be of interest to stakeholders to analyse the relationship of classroom student characteristics (TQ35) with the average mean differences of feeling valued (TQ53h) and intention to continue in the profession (TQ50) and whether or not there is a significant and substantial association with the scale measures of overall job satisfaction scale, satisfaction with the profession scale, satisfaction with this school scale and workplace well-being scale. At the school level, it may be of interest whether or not there is a significant and substantial association of workplace well-being, satisfaction overall, or satisfaction with the profession and school related to the covariates of school type (public or private) (PQ12), school turnover (PQ14), and/or school resource obstacles (PQ29). Examining the mean differences for feeling valued and risk of job attrition with school type, school turnover, and/or school resource obstacles would also be informative.

Cross-theme recommendations

Given the priorities 2.7.2, 2.7.3, and 2.7.4 and the research discussed in the conceptual framework, an analysis of job satisfaction, workplace well-being, and risk of job attrition in relation to items related to professional development needs (TQ27), engagement in collaborative activities (TQ33) and self-efficacy in multicultural classrooms (TQ45) may be investigated out of policy interest.

Results from the TALIS 2013 main survey demonstrated the relationships job satisfaction has with teacher self-efficacy, opportunities to participate in school decision making, collaborative work and disciplinary climate (OECD, 2013[34]). If it is of interest to build on these findings to accommodate priorities 2.7.3 and 2.7.4, similar analyses could be performed to additionally understand the relationship of those same aspects to the new measures of workplace well-being and risk of job attrition.

3.8. Principal leadership

School leadership is a key factor in effective schools (Chapman et al., 2015[67]) and has been found to be related to a range of school and classroom characteristics. Initial studies, positing a direct relationship between leadership and pupil outcomes often showed very little relationship, leading some researchers to state that the influence of leadership had been overstated (Scheerens and Bosker, 1997[68]). However, studies employing indirect effects models were more successful in establishing a leadership impact (Hallinger,

2008[69]), allowing Huber and Muijs (2010[70]) to state that "The impact [of leadership on pupil outcomes] is indirect, and modest rather than strong. Context is an important factor here, however, in that the influence of leadership at the school level is clearly stronger where school autonomy is greater." (p. 70[42]).

As Hallinger's (2008[69]) study confirmed, there is also growing evidence of the reciprocal and contextual nature of effective leadership, whereby student body context itself will influence leadership, which, in turn, needs to be congruent with context. This means that leadership is posited to affect school climate and teacher behaviours, which will, in turn, affect pupil outcomes and, therefore, indicate a range of cross-theme analyses in the TALIS dataset.

Teacher- or school-characteristic averages

Perceptions of leadership style may vary by teacher demographics, such as age and novice or seasoned teachers. Comparing perceptions of leadership styles by whether or not the teacher and principal are the same gender may reveal differences.

Cross-theme recommendations

That there are cross-theme relationships is true for all four aspects of school leadership.

Literature described in the conceptual framework describe potential links to teacher factors, such as teaching practices and job satisfaction but also to factors such as job satisfaction and views on diversity. Principal autonomy shows potential links to job satisfaction.

Instructional leadership and distributed leadership are the two central constructs in terms of leadership styles in TALIS 2018. Both have been identified as being related to pupil outcomes and to teacher behaviours (Harris, Muijs and Crawford, 2003[71]; Pashiardis, Brauckmann and Muijs, 2011[42]), so relationships with the teacher questionnaire items on classroom teaching would be interesting to explore here. Distributed leadership has also been posited to be related to teacher job satisfaction and classroom climate (Hulpia, Devos and Rosseel, 2009[72]).Leadership has also been shown to be related to the extent to which schools are open to innovation (Leithwood and Riehl, 2003[73]). Instructional leadership and distributed leadership may also be related to teacher autonomy.

3.9. School climate

In addition to the within-theme analysis of factors related to various measures of school climate, research discussed in the conceptual framework describes strong associations of school climate with stress levels (Collie, Shapka and Perry, 2012[61]), teacher retention (Miller, Brownell and Smith, 1999[62]; Weiss, 1999[63]), job satisfaction (Collie, Shapka and Perry, 2012[61]; Miller, Brownell and Smith, 1999[62]; Stearns et al., 2015[38]; Weiss, 1999[63]), school conditions (National Academy of Education, 2008[66]; OECD, 2013[34]), and commitment to teaching (Fulton, Yoon and Lee, 2005[74]; Hoy and Woolfolk, 1993[75]; Weiss, 1999[63]). The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

Given the TALIS 2013 findings and the research discussed in the conceptual framework, it may be of interest to data users to analyse the relationship of classroom student characteristics (TQ35) with the classroom disciplinary climate scale, student-teacher relations scale and participation among stakeholders scale. It may be informative to stakeholders to see the relationship of teachers' years of experience (TQ11b), teaching

qualifications (TQ04), and feelings of preparedness (TQ06) in relation to the classroom disciplinary climate scale and student-teacher relations scale.

In the school-level aggregate, the differences of climate of common beliefs (TQ48f), shared rules (TQ48g), teacher initiative (TQ48h), teacher-teacher trust (TQ49e) and stakeholder relations scale could be examined in relation to school type (public or private) (PQ12). Among principals, these same comparisons by school type (public or private) (PQ12) could also be examined in relation to the overall academic climate scale and parent-community involvement (PQ27).

Cross-theme recommendations

Given the TALIS 2013 findings and the research discussed in the conceptual framework, it may be of interest to data users to analyse the relationship of classroom disciplinary climate scale with measures of stress (TQ52), teacher intention to continue in the profession (TQ50), satisfaction with this school scale, self-efficacy scale and self-efficacy in multicultural classrooms (TQ45). TALIS 2013 findings also discussed a link of school leadership with school climate. With these 2018 data, ideas of satisfaction with classroom autonomy scale can now be linked to the classroom disciplinary climate scale and the student-teacher relations scale.

The new teachers' perception of shared rules and perception about taking initiative may serve well to inform policy regarding how aspects of school climate relate to the teachers' satisfaction with their school scale, intention to stay in the teaching profession (TQ50) and feeling valued as a teacher (TQ53h).

3.10. Innovation

In addition to the within-theme analysis of factors related to various measures of innovation, research discussed in the conceptual framework describes strong associations of innovation with school climate (Andrews, 2007[76]; Bakker and Demerouti, 2007[77]), self-efficacy (DeYoung, Peterson and Higgins, 2002[78]; Hurt, Joseph and Cook, 1977[79]; Tschannen-Moran and Hoy, 2001[60]), instructional practices (Anderson and West, 1998[80]; Dumont, Istance and Benavides, 2010[81]; Yi, Fiedler and Park, 2006[82]), and school conditions (Patterson et al., 2005[83]). The following section suggests these between-theme analyses based on prior research and the TGB priorities.

Teacher- or school-characteristic averages by theme

At the school-level, the differences in team and organisational innovativeness could be described by associating these measures with school type (public or private) (PQ12), school turnover (PQ14), and/or school resource obstacles (PQ29).

Cross-theme recommendations

Given the priorities 2.10.1 and 2.10.2 and the research discussed in the conceptual framework, an analysis of innovation in relation to items related to: professional development needs (TQ27), teacher collaboration (TQ33), instructional practices (TQ42), self-efficacy in multicultural classrooms (TQ45) and diversity practices (TQ47) may be investigated out of policy interest. These relationships shed light on how innovation is linked to school conditions, teacher characteristics and needs for professional development – these are relevant factors to examine possible challenges associated with the enactment of innovation in schools and education systems. Furthermore, the associations between principals' perceptions of innovation within schools as organisations and their perceptions

of the school climate (PQ26, PQ27) can be examined to delineate the relevance of school climate factors for innovation at the school level.

3.11. Equity and diversity

There are two types of between-theme analyses that can be conducted. The first refers to associations of the scale scores of the equity and diversity domain, with separate items about this topic elsewhere in the questionnaires. The theme of equity and diversity was originally intended to be "cross-cutting", indicating that it could be combined with virtually all other domains assessed in TALIS. For example, within the domain of teacher education, items could be asked about the perceived preparedness; within the domain of school climate, items could be asked about the perceived diversity policy in the school. These ambitions were later scaled back somewhat; yet, diversity is still part of the following domains (as single or twin items) in the teacher questionnaire, in addition to being a domain on its own: teacher preparation section, professional development in the last 12 months, need for professional development (PD) and influence of feedback. Associations with the scores of the scales in the diversity domain could be computed (again, this could be done within schools, across schools, and across countries and economies). In the principal questionnaire, there is one item about diversity in the scale about hindrances to provide quality education. Again, correlations with the scales of the diversity domain would be relevant.

The second type of between-theme analysis refers, essentially, to all other domains in the questionnaire. The background of these exploratory analyses is that, in the literature, no study of a similar scale has been reported. As a consequence, there is no clear guidance from the literature that other domains could be relevant for equity and diversity. For example, what are the characteristics of schools with high/low scores in the equity and diversity domain? Is diversity policy related to other school policies? Is it correlated with climate variables? Most of these correlations can be computed within schools (some can only be computed at the school level as they combine teacher and principal data).

In the conceptual framework it is mentioned that the theme of equity and diversity straddles teacher and school levels. As a consequence, it will be interesting to examine associations between equity/diversity and various TALIS constructs that are collected at the teacher level and also at the school level and possibly even at the country/economy level (although the latter may be controversial). For example, it will be interesting to establish the link between equity/diversity and job satisfaction at the teacher level, but it will also be interesting to examine this relationship at the school level.

From an analysis perspective, it is important to realise that data at the teacher level are restricted to ethnic diversity whereas data at the principal level involve both cultural and gender/socio-economic diversity. Opportunities to triangulate teacher and principal data are, therefore, restricted to cultural diversity.

3.12. Cross-theme summary

The following table provides a streamlined look-up reference that identifies the most appropriate aggregation unit for analysis (teacher-, school-, or system-level) based on the item construction and the policy questions of interest. This approach is meant to underscore which question items in the TALIS teacher survey ask teachers about their own perceptions versus the perceptions of others in their school. Depending on this perspective, some analyses are best suited at the school level because the "individual as agent" perception datum does not exist. For example, if the dependent (explanatory) variable assesses a teachers' perception of "teachers at this school", the analysis can only address questions at

the school level or aggregate up to the systems level. An analysis could not disaggregate this school-level variable down to the individual level. An analysis can only accurately represent levels at or above the level that is measured in the dependent variable.

The items about teachers' relationships with students (TQ49) are an example of school-level items. With this item, a cross-theme analysis of teachers' satisfaction "with this school" scale with "teachers' relationships with students" would be best suited at the school level. Similarly, analyses that aim to address system-level policy questions where the variation is not likely due to school variation but rather system-wide regulations and laws would be best suited for a system-level analysis. An example of such a system-level question would be analysing how teacher certification associates with teachers' job satisfaction "with the profession" scale. The premise underlying such an inquiry would conceptually stipulate aggregating the measures to a system-level estimate in order to understand the differential impact of teacher certification on teachers' professional satisfaction.

The use of the bold-faced and box-style formatting in the table signals that there is research literature in the conceptual framework that provides immediate justification for the level specification. The identified unit of analysis in Table 3.1 is not intended to limit analyses but, rather, to assist in pointing to those levels that may be of highest policy interest, including those areas that are most malleable with policy. Most cells in the table could have multiple levels listed, but analyses of other levels may provide few levers for policy makers to engage in. For example, it may be the case that teachers' number of years in the profession correlates with induction programming but there is little to act on policy-wise with this information if it is the case that there are now standards in place to induct new teachers and it's the veteran teachers who didn't receive induction.

Table 3.1. Reference table of between-theme unit of analysis

		Explanatory variables										
	Policy priority (dependent variables)	1	2	3	4	5	6	7	8	9	10	11
1	Teacher education and initial preparation		S	**T**	T	S	**T**	**T**	S	S	T	T
2	Human resource issues and stakeholder relations	T			**T**	**S**	**T**	**T**	**S**	**S**	**T**	**T**
3	Teachers' instructional practices	**T**			**T**	T	**T**	T	S	**S**	T	T
4	Teachers' professional practices	T	C	T		T	T	T	S	S	T	S
5	Teacher feedback and development	T	S	T	T		T	T	S	S	S	
6	Self-efficacy	**C**		**T**	T	**S**		**T**	S	**S**	C	T
7	Job satisfaction and motivation	C	S	T	T	S	**T**		**S**	**S**	C	S
8	School leadership		S	**S**	**S**	S		**S**		**S**	**S**	S
9	School climate	S	S	S	S	S	**T**	**S**	**S**		S	S
10	Innovation	C	C	T		T	T	T		**S**		S
11	Equity and diversity	C	C	S	C	S	T	T	S	S	C	

Legend:
T = teacher-level analysis
S = school-level analysis
C = country (system)-level analysis
Boxed and bold-faced indicates research-based associations highlighted in the conceptual framework (Ainley and Carstens, 2018[3]).

4. Reporting standards

This final section outlines the reporting standards for the TALIS 2018 data. To facilitate ease of country/economy use of these data, reporting standards that align with TALIS 2013 will be recommended whenever feasible. However, changes are recommended when there was an analytic change that warrants a different approach, or, if the TGB recommends a change to improve upon the TALIS 2013 reporting standards. For example, the change of the data collection from all paper to mostly electronic collection could elicit reporting standard changes.[4] Similarly, access and ease of online data analysis tools may provide an opportunity to introduce a new technological delivery tool to participating countries and economies.

This guide aims at enhancing the quality of statistical outputs derived from TALIS 2018. The expression "statistical outputs" covers a broad field ranging from basic charts and other displays to inferences on population model parameters. The idea of "quality" is a multi-faceted notion that comprises – in no specific order – accuracy, precision, comparability, relevance, interpretability and utility.

The subsections discussed here include: defining the participants, estimation statistics, displaying statistical information, specifications for scales, and specifications for statistical models.

The main sources used to guide the decisions stated herein are *Statistics Canada Quality Guidelines* (2009[84]), Bethlehem's *Applied Survey Methods* (2009[85]), the OECD's *Handbook for Internationally Comparative Education Statistics* (OECD, 2017[86]), UNSD's *Designing Household Survey Samples: Practical Guidelines* (UNSD, 2005[87]), UNECE's series *Making Data Meaningful*, especially *Part 2: A Guide to Presenting Statistics* (2009[88]) and Wainer's paper on "How to display data badly" (Wainer, 1984[89]). Plenty of authors have published on the topic over the years, these are but a small selection that provide a good summary of the guiding principles in this type of social science survey research.

4.1. Population definitions and implications for analysis

The "school"

In most countries and economies, a school is what one expects it to be: a building where students attend classes taught by teachers who are grouped under one administration led by a principal. Thus, the building, the principal and the teachers become the three components of the "school". To ease teacher sampling, a list of such schools is prepared by the national project managers (NPMs) and the consortium draws a probabilistic sample of schools. When selecting an entry on the list, one actually selects a building, a principal and a set of teachers. All estimates derived from that sample speak to the population(s) covered by the list.

In some countries and economies (e.g. Mexico), the school buildings are shared by more than one set of principal-teachers who take turns teaching different populations of students;

[4] To address the threat of this common-method variance bias, a diagnostic test could be performed with data from countries that took the survey in both paper and electronic versions. In this diagnostic, a series of regression models could test if there is a significant difference in average scores among respondents who took the survey in paper versus electronic delivery by creating a flagged binary variable of survey type. These models would include country dummy variables as controls.

these are usually referred to as "shifts" or "school shifts". Care is taken at sampling to select shifts rather than buildings to allow for cheaper and simpler collection and easier comparisons with the single-shift schools found in most countries and economies.

In a few countries and economies (e.g. the Netherlands), entries on the sampling frames are administrative entities recognised by the respective authorities but need not coincide with what TALIS has in mind when thinking about schools. Variations on frame entries may be clusters of school buildings under the management of a single principal and teachers may be working in more than one of the buildings. In other countries and economies, more than one administration may share the set of buildings – depending, for example, on grade levels – and teachers may work in a single building under a single administration, in a single building under multiple administrations, in more than one building under a single administration, or in more than one building under more than one administration. Whenever possible, the network of frame entries and schools has been simplified prior to sampling so that easy references could be made; however, this was not always possible. In 2013, in Flanders (Belgium), for example, the question on the number of schools where a teacher worked made no sense.

Regardless of the nature of the frame entries, a single school or a cluster, national estimates are design-unbiased and correctly relate to the population covered by the frame entries. However, some school-level statistics ought to be read with care and users should consult the relevant documentation (eventually published as part of a Technical Report) and relevant footnotes before making comparative statements in order to accurately contextualise school-level interpretations.

For reasons of practicality and comparability, sampling, weighting, reporting and analysis is limited to schools and not extended to the wider organisation they may belong to or be part of.

The "target" teacher

Since the first cycle of TALIS, the aim has been to report on teachers, as defined in OECD documentation. The definition often quoted in the documents produced by the Consortium was the following:

> *"The formal definition of a classroom teacher is a person whose professional activity involves the planning, organising and conducting of group activities whereby students' knowledge, skills and attitudes develop as stipulated by educational programmes." (OECD, 2004[90])*

In 2008, schools entirely devoted to adult education or to children with special needs were considered out-of-scope. Consequently, teachers who worked with either population were excluded from school-level teacher rosters if encountered in regular schools during data collection.

From 2013, while schools entirely devoted to adult education or to children with special needs remained out-of-scope, teachers of children with special needs were listed – and eligible for teacher sampling – if encountered in regular schools, and flagged as such on the data files to ease comparisons with 2008.

Consequently, the data from TALIS 2013 can be subset to match the sample definition in 2008. The sample definitions in 2013 and 2018 are equivalent and do not require any sub-setting.

Users should remind themselves that TALIS never imposed a minimum number of hours, a minimum number of students or a specific set of subject matter areas to consider a teacher eligible to sampling. However, practicality and economy may have dictated to exclude the

smaller schools or those located in areas hard or perilous to reach. Those national exclusions are documented in the technical report (OECD, 2019[4]).

The "participating" teacher

What constitutes a participating school was established early on, has not changed over the TALIS cycles and is documented in the *TALIS 2018 Technical Standards* – see Annex B of the technical report (OECD, 2019[4]) – and at Paragraphs 61 and 66 of the *TALIS 2018 Sampling Manual* (internal document). At least half of the teachers selected in a school must have participated for their school to be considered a participating school. This rule was set regardless of the principal's response to the principal questionnaire.

What constitutes a participating teacher has evolved little over time. No rules were set in 2008: a single response to the teacher questionnaire qualified the respondent as a "participating teacher". This is how a "participating teacher" was, again, defined in 2013 and 2018, in the latter case following a TGB level discussion on the selection and minimum number of responses to key questions data fields.

There should not be any issue in comparing teacher-based estimates over time – assuming the wording of questions and answer choices haven't changed. Tests conducted using 2013 data on several stringent minima revealed that differences in estimates were hardly ever significant (Gouzi and Dumais, 2017[91]).

4.2. Producing sample statistics and estimating population and model parameters

There are many software packages available to data users and analysts that can compute averages and proportions, create cross-tables and a flurry of multivariate statistics. However, the proper handling of TALIS data, like data from any complex design survey, requires software that is able to handle survey weights and replication techniques.

The survey weight must be used in order to obtain design-unbiased estimates of population or model parameters and the survey design must be accounted for when estimating sampling errors for those estimates. TALIS data files, the national versions and the international comprehensive international file, all contain the appropriate survey estimation and replication weights. Specialised software is available that can produce design-unbiased estimates of population values, with their correct design-based sampling errors, for example, the IEA interface for SPSS called IEA IDB Analyzer, SAS, Stata and WesVar. Users need to verify which procedure options need to be "switched on" to turn to survey estimation mode.

Univariate statistics (e.g. means or averages, proportions, medians) are likely to be estimated for each participating country and economy and displayed as a table, one line per country/economy, one column for each statistic. Since the number of such tables is likely to be large, as in 2008 and 2013, the list of countries and economies on the left hand-side column should be sorted alphabetically to avoid the impression of and misinterpretation as "league tables" and to simplify following one or many countries across tables and charts. See section 4.3.

Sub-national entities or other adjudicated sub-regions not corresponding to a whole country or economy participating in TALIS should be reported in separate table segments and not contribute to averages. The OECD will need to consider inclusion in main tables and figures or appendices and online tables only and, in particular, if sub-entities and entities are displayed in the same table, that both do not contribute to averages.

Lines, cells, or countries and economies where item non-response is substantial should be clearly marked. Where it is high (i.e. at or above 50%) estimates should not be reported at

all. In instances of 100% non-response, a practice in PISA is to identify this cell of data with an "(m)" if the country/economy data were removed for technical reasons and "(w)" if the country/economy data have been withdrawn or not collected by request of the country/economy (OECD, 2016[92]). This convention should be adopted for TALIS 2018 as well.

There should be a minimum number of responses in a cell to display it; otherwise the OECD's standard "(c)" notation should be displayed in the table to identify there are too few observations to provide reliable estimates (OECD, 2017[86]). For consistency with other OECD programmes, most significantly PISA, TALIS 2018 and later should suppress estimates if they are based on fewer than 30 teachers or 5 schools with valid data, respectively. Proportions under 0.05 (i.e. 5%) could also be clearly marked, e.g. greyed out (or the cell shaded), as was the standard in TALIS 2013 reporting. In addition, it would be informative to use "(a)", as previously done in PISA and TALIS, to identify if an item category does not apply to the country/economy because it was not administered or was an optional part of a questionnaire (i.e. a structural zero instead of a sample zero).

While some indication of the quality of the estimates should be readily visible in the tables, it may not be necessary to print all sampling errors in the body of the document. The tables could gain in legibility if only the less precise cells were highlighted, for example, using the OECD's standard "(r)" notation (OECD, 2017[86]).

If averages are computed (e.g. average taken over TALIS participating countries and economies, over OECD members), they should be displayed as the last lines of a table. In keeping with TALIS 2013 and the standards set forth by the OECD in the *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications*, international averages would be computed using a one-country-one-datum contribution to the average calculation (2017, p. 86[86]). So-called weighted averages of country/economy values should not be supported, as the sizes of the participating countries and economies can be extremely different (from Brazil and the Russian Federation to Malta and Singapore) and, thus, skew the estimate so that many countries and economies would not effectively contribute to a weighted average. If TALIS 2018 estimates were averaged over participating countries and economies, weighting each contribution proportionally to their respective population, Brazil, the Russian Federation and the United States would account about equally for nearly 46% of the average, nine countries/economies would account for the next 45%, leaving 24 countries/economies with the last 10%.[5]

If tests of independence or of distribution are conducted on the data tables, the test statistics should be corrected for design effect using the adjustment suggested by Rao and Scott (1987[93]). This adjustment to the χ^2-2-statistic is readily available in SAS, Stata and WesVar; the adjustment is also available in SPSS Complex Samples, but impossible to use with public TALIS data files because the details of the various national sampling plans are not disclosed to users (first and second order probabilities of inclusion at each stage, stratification, etc.). The IEA IDB Analyzer could not be used for this purpose either as it supports parametric analyses only. These same considerations apply to comparing statistics when triangulating teachers' and principals' responses.

When computing estimates of change between TALIS cycles, analysts and data users should keep in mind the following: over the course of five years, changes come from many sources, some caused by our own decisions as survey designers. When estimates of change

[5] Data extracted from TALIS 2018 sampling frames and, if necessary, TALIS 2013 estimates; for some countries, no data for the size of the ISCED level 2 population were available at the time of preparing this note.

are required, analysts should verify that the variables of interest haven't changed; then they may infer a change in the population itself. Analysts should also account for possible changes in population coverage (national exclusions) and if required, filter the datasets so that coverage be as similar as they can make it, or otherwise annotate and document the differences in population scoping and coverage. This would require an additional "2013 comparative average" to be calculated among countries and economies with changing population coverage from the 2013 coverage (e.g. Belgium, United Arab Emirates). TALIS does not take longitudinal measures; therefore, all changes are changes in snapshots, not individual paths.

Table 4-1 lists TALIS Participants across the cycles and options and whether analysis between cycles is possible (or where restrictions apply), assuming good, fair or poor response rates in 2018. Insufficient data should not be reported according to TALIS 2018 technical standards. The table does not include indications on whether countries and economies elected the same options in multiple cycles since no related reporting was planned from the outset given expected low numbers, e.g. between the ISCED 3 option in 2013 and 2018. The cross-cycle participation indicates Participants who took part in TALIS 2008 and 2018 only (i.e. skipped 2013, such as Austria). However, questionnaires changed in substantial ways between the first and third cycle and analyses were known to be limited to a smaller set of measures.

Table 4.1. List of TALIS participants (core and international options) between TALIS survey cycles

Country/economy	2018 participation (core and options)				Cross-cycle participation for ISCED 2				Comment
	2018 ISCED 2	2018 ISCED 1	2018 ISCED 3	2018 TALIS-PISA	2008 > 2013 (ISCED 2)	2008 > 2013 > 2018 (ISCED	2008 > 2018 only (ISCED	2013 > 2018 (ISCED 2)	
Argentina (Ciudad Autónoma de Buenos Aires)	+	+		+					
Australia	+	+		+	+	+		+	
Austria	+						+		
Belgium	+								
Flemish Community of Belgium		+			+	+		+	Adjudicated separately as a part of Belgium in 2018 for ISCED 2 (only)
French Community of Belgium	+								Adjudicated separately as part of Belgium in May 2019
Brazil	+		+		+	+		+	
Bulgaria	+				+	+		+	
Canada (Alberta)	+		+					+	
Chile	+							+	
China (Shanghai)	+							+	
Chinese Taipei	+	+	+						
Colombia	+			+					
Croatia	+		+					+	
Cyprus[1]	+							+	
Czech Republic	+			+				+	
Denmark	+	+	+	+	+	+		+	
Estonia	+				+	+		+	

Country/economy	2018 participation (core and options)				Cross-cycle participation for ISCED 2				Comment
Finland	+							+	
France	+	+						+	
Georgia	+			+				+	
Hungary	+						+		
Iceland	+				+	+		+	
Ireland									
Israel	+							+	
Italy	+				+	+		+	
Japan	+	+						+	
Kazakhstan	+								
Korea	+	+			+	+		+	
Latvia	+							+	
Lithuania	+						+		
Malaysia					+				
Malta	+			+			+		
Mexico	+				+	+		+	
Netherlands	+	+			- *	- *		+	* Insufficient response rate in in 2008; no cross-cycle estimates should be reported
New Zealand	+							+	
Norway	+				+	+		+	
Poland					+				
Portugal	+		+		+	+		+	
Romania	+							+	
Russian Federation	+							+	
Saudi Arabia	+								
Serbia									
Singapore	+							+	
Slovak Republic	+				+	+		+	
Slovenia	+		+				+		
South Africa	+								
Spain	+	+			+	+		+	
Sweden	+	+	+					+	
Thailand									
Turkey	+	+	+	+			+		
United Arab Emirates	+	+	+						Only Abu Dhabi participated in TALIS 2013
United Arab Emirates (Abu Dhabi)								- *	* No separate adjudication
United Kingdom (England)	+	+						+	
United States	+							- *	* Insufficient response rate in 2013; no cross-cycle estimates should be reported
Viet Nam	+	+	+	+					
Total	48	15	11	9	18	15	6	34	

1. Note by Turkey
The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
Note by all the European Union Member States of the OECD and the European Union
The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

TALIS does not control the national conditions under which teachers live and work; therefore, causal inferences are not be appropriate. Causality is especially tempting when interpreting the slope parameters of a regression model. A common interpretation given to the slope b is a rate of change, as in this statement: "if a country rose from X-value X_1 to $X_1+\delta$, its Y-value would be expected to change from Y_1 to $Y_1+b\delta$". A more appropriate interpretation would be a statement like "countries whose X-value is close to X_1 are expected to show a Y-value around Y_1, while countries whose X-value is close to $X_1+\delta$ are expected to show a Y-value around $Y_1+b\delta$". It is important to remember that X and Y are measured at the same point in time and that TALIS has not tracked countries and economies nor individual teachers over time.

When models are tested (e.g. independence, regression), the usual practice is to report the so-called p-value. Given the large sample sizes (teachers or schools) drawn in each participating country/economy, even after adjusting for the clustering and inefficiencies of the stratification, the p-values are bound to be small, making about any model specification statistically significant. Then, the utility of the model has to be gauged using some other device. Reporting some measure of goodness-of-fit along the model parameters and the overall p-value (e.g. χ^2, R^2) in order to evaluate the relevance of the contribution is also recommended. Similarly, slope coefficients (effect sizes) near zero can be misleading to readers and policy makers. TALIS 2013 adopted a writing rule to report on standardised coefficients with "weak", "moderate", or "strong" associations where the absolute values were <0.2, 0.2-0.3, and ≥0.3 standard deviations, respectively.

4.3. Displaying statistical information: charts and graphs

There are many ways to display statistical information and even basic desktop software packages propose a number of graph and chart options. If, as the saying goes, "a picture is worth a thousand words", then the picture should be selected with care. The following guidelines come from best practices.

Firstly, to avoid giving an artificial impression of "growth", all charts, tables and graphs that have countries and economies on X axis should generally display countries and economies in a fixed order throughout the report; alphabetical order is arbitrary enough to make references easier when one tries to follow a few countries and economies throughout the report or several tables and charts. This may be contrary to advice given elsewhere – for example, in UNECE (2009[88]) – but may be preferable, given the potentially large number of tables and charts where the list of participating countries and economies will appear. A fixed order would also reduce the misinterpretation of the results as cross-comparative league ratings. When some other unordered categorical variable is used, such as the abscissa axis, then its modalities should appear in the order in which they appear in the questionnaire. If a "refused / don't know" category is also reported on the graph, then it should be to the right of the response categories as it hardly constitutes an intermediary stance between agreement and disagreement, satisfaction and dissatisfaction.

In monochromatic bar charts, the darkest shade should be associated with the highest value of the X variable and the lightest shade with smallest value of the X variable. This is especially intuitive and easy to read when displaying, say, a 4-point Likert scale as stacked bars for each country/economy.

If an average value is to be displayed (e.g. the average of the TALIS participating countries and economies or the average of the participating OECD members), then it should appear as the rightmost X-value. Alternatively, to facilitate the quick comparison of a country/economy value to the average, the average could be displayed (possibly with its confidence interval) as a straight line spanning the width of the graph.

If appropriate, some marker could be used to point out which country/economy estimates (for example) are significantly different from the average. Likewise, there could be an indicator of low item response (say, less than 50% item response). For scaled measures, averages could only be displayed if scalar invariance was reached (see section 4.4 below).

Pie charts should be avoided if more than eight slices have to be displayed; the first slice should start at noon, going clockwise.

The use of "3D effects" is strongly discouraged: *Y*-scales are hard to read and it becomes difficult to compare bar heights or the surface of pie wedges that are a bit distant, especially with dozens of countries and economies to display.

Clearly, "line charts" or connecting points on a graph only make sense if *X* is ordinal (at the very least) or an actual numerical variable (discrete or continuous).

When displaying two continuous variables using scatter plots, one might prefer to use "bubble plots" where each point of the scatter plot is replaced by a disk whose area is proportional to the sampling variance associated to the point displayed. This avoids confusing confidence interval overlays on traditional scatterplots.

When displaying scale scores for countries and economies, an average should not be displayed if the scale is not shown to be scalar invariant (see section 4.4 below).

4.4. Specifications for scale configuration

The scales are defined to reflect the underlying latent construct that cannot be directly observed (for instance, teachers' beliefs or self-efficacy). In order to make decisions on the use of scales (or latent constructs) for further analysis, it is crucial to consider the specific scale properties and their quality, that is, their reliability and validity in a cross-cultural context. The following description provides an overview of the tests applied to scales in TALIS 2018 and the implication of the respective results. Technical details on the analysis are reported in the *TALIS 2018 Technical Report* (OECD, 2019[4]).

The process of scale construction and validation in the TALIS 2018 main survey (MS) is applied as follows:[6]

1. **Item identification and modification based on the field trial (FT) data analysis:** From a theoretical standpoint, the experts have identified the items believed to inform each latent construct. Based on a set of validity and reliability checks of the FT data, the scales were modified or, in some cases, removed from the MS.
2. **Reliability and validity analysis** in the main survey (MS): the dimensionality, reliability and validity of the scale is tested.

   If the scale fails these tests, adjustments are made to the scale as indicated empirically and in agreement with the experts and the reliability and validity analysis will be repeated.
3. **Measurement invariance testing and model evaluation** in the main study (MS): the cross-country and cross-ISCED-levels comparability of scales is tested.

   This includes comparing countries and economies from all ISCED levels participating in TALIS 2018. If a scale fails these tests, in some cases further

[6] It should be noted that, following the TAG suggestion on the creation of scale scores for cross-cycle analysis using pooled data from respective cycles, the consortium decided to abstain from the creation of such scales in the recent cycle due to time constraints.

adjustments will be made in agreement with the TALIS Consortium, followed by a new analysis round (#2 and #3)

4. **Construction of the factors score for each scale:** The scale scores are constructed as indicated by the results from prior analysis

Item identification and modification based on the FT data analysis

The process starts by the theoretical identification of items for a latent construct. In TALIS 2018, the identification of items is based on 1) lessons learned from prior TALIS cycles (in particular with respect to the repeated survey constructs), 2) research theories from the respective fields and 3) expert knowledge on item and scale construction. The latent constructs were evaluated using the FT data and modified for the MS data analysis. In some cases, the decision was taken to remove the latent construct from the final analysis and to analyse the respective topic using single items. Also, to reduce the length of the questionnaire, some of the items were removed from the questionnaire after the FT. This decision was taken considering the FT results of the scale validation.

The FT analysis of the scales was conducted using linear modelling,[7] which statistically requires the response options to be interval (or continuous). The same analysis was further conducted using ordinal response modelling that corresponded to the ordinal level of response categories (i.e. categorical) in TALIS background questionnaires. The results were very similar, with some model improvements (in terms of model fit and measurement invariance across countries and economies) using ordinal response modelling, while the opposite also occurred, i.e. some of the scales based on ordinal response modelling performed worse than when based on linear modelling. Practical challenges related to a possible change of the approaches from linear modelling in previous cycles to ordinal response modelling, leading the TAG to recommend using the linear measurement model for TALIS 2018 as the differences in the results are, overall, minor. Thus, in the MS, all analysis and scaling procedures are based on linear modelling.

Reliability and validity analysis

The constructs used to compute scale scores in the MS should undergo extensive checks with regard to scale validity and reliability, including item statistics to check the distribution of missing data, responses per category and skewness, as well as Cronbach's alpha and item-total correlation, to analyse the scale internal consistency. Moreover, confirmatory factor analysis (CFA) are conducted to evaluate how well the theoretically defined latent construct mirrors the information in the actual empirical data. CFA allows for inferring on the conceptual scale from the associated empirical items, by establishing associations between them. The association between each item and the scale is described in a regression line (OECD, 2014[17]). Using well-established model evaluation criteria, the model fit indices, and the theoretically defined model (latent variable and its indicators/items) will be evaluated with respect to its match to the empirical data.

The scale evaluation leads to a selection of scales that meet the majority of the following criteria:

1. The items of the scales are normally distributed and are not skewed.
2. The proportion of respondents contributing to each category is at least 5%.

[7] Models assuming metric level of items, i.e. based on variables with interval or continuous response options.

3. The reliability coefficient (omega for unidimensional scales, stratified Cronbach's alpha for multidimensional scales) is at least 0.7.
4. The item total correlation is more than 0.4 (Ladhari, 2010[94]).
5. The scale is unidimensional (EFA).
6. The scale has passed the cut-of criterion of model evaluation of CFI≥0.90, TLI≥0.90, RMSEA≤0.08, SRMR≤0.06.[8,9]

Scales that do not pass the quality checks are reviewed for scale dimensionality by exploratory factor analysis (EFA). If the analysis reveals multi-dimensionality, one of the following modifications is applied: 1) reducing the number of items in order to build a uni-dimensional construct as indicated by the structure matrix provided by the oblimin rotation in the principal axis factoring (PAF) analysis, 2) splitting the scale into two constructs as indicated by the EFA results, or 3) keeping the scale as uni-dimensional in further analysis if there are strong content-related or other arguments. The latter is relevant, for instance, in the case of lack of theoretical arguments to support multi-dimensionality.

It is important to stress that the statistical criteria are reviewed in the light of theoretical considerations. Some scales that do not pass the above-mentioned statistical criteria will be kept based on content-related arguments. The details on each scale will be reported in the technical report.

Measurement invariance testing and model evaluation

The TALIS 2018 scales are used to conduct analysis in a cross-national perspective. When comparing scales across groups (such as nations, education systems, cultures, ISCED levels), it is crucial to make sure that the scales have an equivalent meaning in each of the groups of comparison. Thus, in addition to the analysis of the overall scale reliability and validity, the TALIS 2018 scales will be evaluated with respect to their equivalence across ISCED levels and participating education systems (country/economy level).

To analyse the cross-ISCED-levels and cross-country comparability, all scales will be evaluated using item statistics, Cronbach's alpha, item total correlation and CFA at country/economy level. Checks will be carefully performed to consider the results per country/economy based on the country/economy-level sample sizes of each item. The criteria applied here will be the same as for the overall analysis mentioned above. If a scale passes this validity and reliability check, it is assumed the overall model can be applied to create scales at the country/economy level.

However, further validity checks are necessary to assume that the scales can be constructed using constant parameters across all participating education systems (i.e. the parameters are held to be equal across groups) in order to conduct statistical comparisons. The statistical method to evaluate the statistical comparability of scales across different groups/-subpopulations, such as ISCED levels and education systems, is called *measurement invariance testing*. This method assumes that the comparability of a latent

[8] Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error Approximation (RMSEA) and the Standardised Root Mean Square Residual (SRMR).

[9] The cut-off criterion of the SRMR in TALIS 2013 MS was less strict (SRMR≤.1). To enhance the alignment with the established cut-off criteria for model evaluation proposed in the literature (Hu and Bentler, 1999[106]; OECD, 2013[107]; Schermelleh-Engel Karin, Moosbrugger and Müller, 2003[109]; Steiger, 1990[108]; Yu, 2002[110]) a stricter cut-off of SRMR will be applied in TALIS 2018 MS. This is justified by the fact that the scale construction in TALIS 2018 benefits from prior TALIS cycles as well as from the TALIS 2018 FT. These prior experiences and analysis have led to improvements of scales in the recent cycle and justify applying slightly stricter criterion of SRMR than in TALIS 2013 in order to improve the robustness of the results.

construct among groups (e.g. education systems) can be reached, to a varying extent. Latent (not directly observable) constructs are constructed based on associations between the construct and several indicators/items in a joint model, which is normally called the measurement model. The measurement model consists of several parameters (i.e. means, regression coefficients, variances) of which comparability across groups can be tested. The optimal level of comparability is given, if all parameters of a model (except the residual) are fixed to be the same across groups (e.g. education systems). This level of invariance is called the *scalar level of measurement invariance*.[10] In this case, cross-group comparisons of scale means are justified and the results can be assumed to be robust and reliable.

It is not easy to establish constructs that reach the scalar level of measurement invariance in large-scale assessments with a large number of education systems participating. Sound theoretical assumptions support the idea that measurement constructs of latent traits (such as self-beliefs) are not necessarily fully comparable across cultures, but they might be comparable to a certain degree. Thus, the degree to which the stricter model (i.e. a model that assumes equity of parameters across groups) suits the data comparing to the less strict model (i.e. a model that assumes some degree of flexibility of parameters between groups) will need to be evaluated. The underlying idea is that if the measurement models are not comparable across education systems' ISCED levels with respect to their levels (means), they might at least be comparable with respect to their specific meaning. The meaning of the latent scale is defined by the content of the questions participants were asked that are used to create the construct. If the strength of the associations (i.e. the magnitude of the regression parameters) is the same across groups, then the meaning of the latent construct can be assumed to be comparable. This level of measurement invariance is called the *metric level of measurement invariance*.[11] If a scale has established the metric level of invariance, correlational analyses (such as correlation or regression analysis) are assumed to be free of the cross-cultural bias.

If a scale cannot be assumed to be comparable across education systems with respect to its meaning or content, it is still possible that there is, at least, comparability at the conceptual level. Using fit indices for model evaluation, the analysis can reveal whether a latent construct can be specified by a particular configuration of items in all analysed education systems in the same way. This level of measurement invariance is called the *configural level of measurement invariance*.[12] It indicates that the configuration of items is the same in all groups of comparison. If a scale only reaches the configural level of measurement invariance, no statistical methods can be applied to compare the scale scores across

[10] *Scalar level of measurement invariance* is given 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 for all countries are equivalent (Cheung and Rensvold, 2002[111]; Cheung and Rensvold, 1999[112]; Davidov, 2008[113]; OECD, 2013[107]; Steenkamp and Baumgartner, 1998[114]). If the intercepts for all countries are equivalent, the expected value of the items is the same across countries when the value of the construct is zero, meaning that, given a certain value of the observed item, we can claim an equivalent value/degree of the construct across different countries. This allows for unbiased comparisons of scale means across education systems.

[11] *Metric level of measurement invariance* is given 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 allows for claiming that one unit change of the construct will lead to the same amount of change in the items that constitute the construct across different groups (i.e. education systems) (Cheung and Rensvold, 2002[111]; Cheung and Rensvold, 1999[112]; Davidov, 2008[113]; OECD, 2013[107]; Steenkamp and Baumgartner, 1998[114]).

[12] *Configural level of measurement invariance* is given when the construct is measured by the same items. It implies that structure of the construct indicated by the configuration of items is equivalent across countries (Cheung and Rensvold, 2002[111]; Cheung and Rensvold, 1999[112]; Davidov, 2008[113]; OECD, 2013[107]; Steenkamp and Baumgartner, 1998[114]).

education systems without violating the basic assumption of the comparability of the measurement construct. The comparability is only given at a conceptual level, where results from different countries and economies can be discussed referring to the country/economy level without referring to "differences" or "similarities across education systems". The modelling method chosen for TALIS 2018 (as for TALIS 2013) is the multiple group confirmatory factor analysis (MGCFA). To evaluate the cross-ISCED-levels and cross-country comparability the following criteria will be used:

- To evaluate the scalar level of invariance (which enables for comparisons of the means of the scales across different education systems), worsening fit indices changes (Δ) are acceptable within the following criteria: ΔCFI<0.010 or ΔTLI<0.015 and ΔRMSEA<0.015 or ΔSRMR<0.01/WRMR<0.04 (Chen, 2007[95]).
- To evaluate the metric level of invariance (which enables comparisons of associations between education systems), worsening fit indices changes (Δ) are acceptable within the following criteria: ΔCFI<0.010 or ΔTLI<0.015 and ΔRMSEA<0.015 or ΔSRMR/WRMR<0.03 (Chen, 2007[95]).[13]
- To evaluate the configural level of invariance (no statistical comparisons across education systems are permitted), the following model evaluation criteria will be used: CFI≥0.90 or TLI≥0.90 and RMSEA≤0.08 or SRMR≤0.06/WRMR<0.90.

Construction of the factor scores for data release and recommendations for analysis

An important consideration is that the results from the scale reliability and validity analysis as well as the cross-country comparability described above have major implications, not only for 1) the construction of factor scores but also for 2) the use of the factor scores attached to the data set for further analysis.

1. Factor scores indicate each individual's scoring on the scale or latent construct. To construct the factor scores the information on the measurement invariance of each scale is used to specify a suitable model. Since it is not obvious to users of the TALIS 2018 database which scale fulfils which level of measurement invariance, the level of invariance will be included in the variable label of each scale constructed.
2. The 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:
    a. Cross-country analysis of scales with scalar level of invariance:
        - *Recommendation for analysis:* For these scales, the comparison of the mean score of the scale is meaningful across groups. Scale mean can be compared across countries and economies (e.g. "Staff-beliefs are significantly higher in country A compared to countries B and C")
        - *Limitation:* If the scale reaches the highest level of invariance, the scalar invariance, the comparability of the measurement construct between education systems (and over-time) is mirrored in the way the index is created.

[13] According to Chen (2007[95]), simulation study SRMR used for measurement invariance testing applied to big data sets can be less strict than in the TALIS 2013 MS, where the cut-off criterion of absolute changes SRMR ≤ .005.

b. Cross-country analysis of scales with metric level of invariance:

- *Recommendation for analysis:* For such a scale, the strength of the association between the scale and items are comparable across countries and economies, and statistical methods such as correlation and regression are applicable. Comparisons of associations between countries and economies are justified (e.g. "The association of staff beliefs and level of education are stronger in country A and B than in country C"). With the metric level of measurement invariance, scales can be used for analysis based on linear regression (please refer to the next paragraph).
- *Limitation:* If the scale reaches the metric level of invariance, the index of the scale is created with equivalent factor loadings but with non-equivalent means, respectively.

c. Cross-country analysis of scales with configural level of invariance:

- *Recommendation for analysis:* Statistical procedures should be limited to analysis of populations within countries and economies.[14] At the cross-national level, only qualitative (descriptive) comparisons are statistically justified (e.g. "Associations of staff beliefs and staff education is positive in country A and B, whereas there is no significant association in country C"), which should be presented together with the limitations concerning the interpretation of the results, in particular concerning the differences between countries and economies of the meaning of the construct. Alternatively, analysis at item level could be carefully considered if there are theoretical assumptions that specific items are less impacted by the cross-cultural bias.
- *Limitation:* If the scale only reaches configural invariance, the scale index is constructed in such a way that the factor loadings and means are allowed to vary across education systems. Further analysis aiming at cross-country comparisons can only be conducted at the conceptual level, meaning that no statistical methods of comparison (such as t-test) are applicable in such cases.

Violating these assumptions may lead to biased results. For instance, if scale scores created based on the assumption of metric invariance are used to compare country/economy means of a scale, the differences between countries and economies will likely be biased. Thinking, for example, of staff self-beliefs, if the construct did not reach the scalar level of measurement invariance, this would mean that country/economy differences in the average staff-beliefs can also be due to cross-cultural differences in the way in which response patterns (agreement and disagreement) are reflected in the latent construct. In case the construct does not reach even the metric level of invariance but does reach the configural level of invariance, statistical comparisons would be biased to an even greater extent. Differences between education systems could be due to differences in the meaning of the constructs (indicated by varying associations of the items with the scale) and due to differences in how the level of agreement or disagreement is mirrored in the scale.

The use of scale scores for regression-based analysis is justified for all scales that reach the metric or scalar level of measurement invariance. For instance, path models with manifest or latent variables (the latter also called "structural equation models") can be conducted to

[14] In the case of group comparisons of any groups within countries (such as gender or teacher versus assistants), measurement invariance testing of the cross-groups comparability of the scales scores are recommended.

compare the strength of the associations between countries and economies. In such analysis, the standardised beta parameters are comparable across variables and groups. If the analysis uses the latent models calculated for one specific country/economy, the variables of the respective scale have to be included in the model instead of the created scale score. In such case, the model uses the indicator variables to create latent constructs and estimates associations between them. Further, to compare across countries and economies, the analysis can be based on MG-SEM (multiple groups structural equation modelling), an approach that allows the use of latent constructs for defined groups (such as countries and economies) that are estimated with fixed parameters, i.e. they are programmed to be comparable across groups. Yet, at the same time, the associations are specified in a flexible way for each country/economy, meaning that the associations between the constructs are country/economy-specific, but the latent constructs themselves are the same for all groups in the analysis. Please note that such analysis pre-assumes that the comparability of the constructs across countries and economies is justified, i.e. is validated in some way, for example, by the means of measurement invariance testing.

4.5. Specifications for modelling

Multivariate and multilevel modelling

With cross-sectional data such as these, it is not possible to establish causation. However, in efforts to inform policy it is instructive to establish associations between variables. Bivariate correlations in survey data are often used to explain the direction and strength of simple associations between two measures that share a common foundation, such as the overlap of two practices in a classroom or two attitudes about teaching. When both variables are numeric (discrete or continuous), Pearson's *r* can describe the direction and strength of the association. When the variables are categorical, other tests are more accurate. For correlations including at least one nominal variable, such as classroom subject matter or school type, Cramer's *V* is most appropriate. For correlations including ordinal variables, including 4-category Likert measures, Kendall's *Tau* is preferred since Pearson's *r* is prone to inflating the strength of the correlation when the linear assumption is applied to a few-category ordinal variable (Lohr, 2010[96]; Snedecor and Cochran, 1989[97]).

When the goal of the analysis is to establish a complex, multivariate relationship, then regression modelling techniques are commonly used. Ordinary least squares and binary logistic were the predominant regressions used in prior TALIS reports due to their robustness to assumptions and friendliness to non-technical readers. Regression models declare one measure as a "dependent" variable and one or more variables as "independent" or "control" variables. With the cross-sectional survey data collection design of TALIS, the declaration of these variables rests on theoretical logic rather than a causal design related to experimental or longitudinal panel data collections. Goodness-of-fit test statistics (see above, Section 4.2) would be used to check the relevance of the theoretical assumptions that guided the model specifications.

The declaration of the dependent ("*Y*") variable in the regression model in this type of cross-sectional data points to the outcome being explained. The independent variables are the focal measures that are theoretically and logically linked to the outcome. Control variables are traits related to the independent and/or dependent variables that condition the focal relationship between the independent and dependent variables (Aneshensel, 2012[98]). Variables that should be controlled in this quasi-experimental modelling approach are those measures that account for spuriousness, and that suppress or moderate the focal relationship (Aneshensel, 2012[98]). There are some antecedent variables, such as teacher demographics

or initial teacher education information, that temporally precede some other TALIS variables. With antecedent variables, the temporal order can be discussed if it is accompanied by the reminder that temporal order does not assume causation.

The TALIS 2013 report regularly specified teacher controls of gender, years of teaching experience and level of education, as well as the classroom controls of classroom demographics and the systems control of country/economy identifier as control variables in the regression models (OECD, 2014[17]). These control variables are not intended as the focus of the reporting but, rather, are included in efforts to increase the accuracy and precision of the direction and strength of the estimate.

ted focal relationship. For transparency, these controls would be noted in each report table (perhaps at the bottom or under the title) and the regression estimates would be included in web-based appendix tables. Other sociological/contextual variables that would be important to use as controls for various themes are those listed in the teacher- or school-characteristics subsections in Section 3.

When the purpose of an analysis is to establish an "overall" relationship between focal variables – those relationships across the "average TALIS teacher" that are independent of any particular system – model specifications need to adjust estimates in order to account for the differences in mean intercepts on the dependent variable for each country/economy (Gustafsson and Johansson, 1999[99]; Wooldridge, 2010[100]). Fixed effects regression modelling is a technique commonly used to control for the independent country/economy differences (e.g. intercepts) in order to focus on the central relationship (Gustafsson and Johansson, 1999[99]; Wooldridge, 2010[100]). Given the complex sampling employed by TALIS, these fixed effects can be declared by simply using country/economy identifier variables (e.g. country/economy dummy variables). This approach can be easily and accurately applied with the complex sampling weight and replicated with the public dataset. This method can be used with most dependent variables – item or scale constructs – but it cannot be used with scale constructs that do not demonstrate a minimum of metric invariance because the specification assumes parallel slopes on the measures.

A multilevel modelling of TALIS teacher data where the country/economy is one of the levels presents the important drawback of assuming that country differences are random. That is, a three-level modelling where the country/economy was declared as level-3 with schools nested within country/economy at level-2 and teachers within schools at level-1 would assume dependent error structures between country/economy and school (Raudenbush and Bryk, 2002[101]), which is not likely a uniform assumption across TALIS countries and economies. Similarly, a two-level specification of teachers (level-1) within countries and economies (level-2) or schools (level-1) within countries and economies (level-2) also assumes a uniform dependent error structure between country/economy at level-2 and school or teacher at level-1 (Raudenbush and Bryk, 2002[101]). Instead, fixed effects at the country/economy level can stabilise each country's/economy's intercept, or point of entry, while not assuming random variation of the country/economy effect on the individual. The purpose of these fixed effect country/economy indicators is to control for country/economy variation on the estimates. The purpose would not be to compare the country/economy estimates to the reference country/economy indicator that is omitted from the regression specification.

Pragmatically, several statistical packages, such as SPSS, do not allow for complex sampling weights when declaring a multilevel modelling specification, which is problematic in producing population estimates. However, some specialised software is available (HLM, MLwiN, Mplus) but the correct use of the weights would require access to the weights components, which are unavailable on the public-use international file. Any software under consideration would also need to handle balanced repeated replication

(BRR) replication weights. Moreover, multilevel modelling needs sufficient numbers in all cells at all levels, which may also induce a complication for some TALIS participants with lower teacher counts within schools.

A multilevel modelling structure on a per-country/economy basis, e.g. within-country/economy analyses where each model represents only one country/economy, where teachers (level-1) are nested within schools (level-2) could be produced, as long as the software package could also use the complex sample weights. A two-level modelling structure of schools (level 1) within countries and economies (level 2) cannot readily be performed without the complex specifications, as stated in the above paragraphs. In addition, some multilevel analyses conducted simultaneously on several countries and economies may only be possible at the OECD and to those who have access to the national databases; HLM (and possibly other packages) expects to have level-specific survey weights that may not be distributed in the international database for confidentiality reasons, as was the case in TALIS 2008 and 2013.

It is imperative to use the TALIS sampling weights to ensure accuracy of model estimates. TALIS uses the Fay's adjustment BRR weight to adjust for the complex sampling design of TALIS. These weights not only adjust for the unequal probability of sampling, but also provide asymptotically unbiased estimates of the standard errors. Many statistical packages (SAS, Stata, R) can use this weight and appropriately adjust models. For users of SPSS software, it is possible to use the IEA IDB Analyzer to produce accurate estimates. The details of weighting are discussed in the technical report (OECD, 2019[4]). For the time being, rationale and instructions included in the *TALIS 2013 Technical Report* and user guide (Becker, 2014[102]; OECD, 2014[16])provide information that will still be accurate for the 2018 design and weighting approach.

In addition to establishing complex relationships, regression modelling can be a technique used when the purpose is to establish associations between two variables where at least one of the variables is a latent construct that did not reach scalar invariance (and, therefore, cannot be accurately used in cross-country comparisons). Few latent variables in TALIS reach scalar invariance (see above, Section 4.4) however many do reach metric invariance. For these metric-invariant latent constructs, a country-fixed effects regression specification accounts for differences in the mean intercepts for each country/economy and focuses attention on the slope coefficient. The slope coefficient is good at establishing the direction and strength of the association with the dependent variable, no matter the differences in country/economy means and assuming that the assumption of parallel slopes is reasonable.

Another possible approach to analysing data from TALIS, besides the variable-centred focus on proportions, measures of central tendency or variation at the system level, relates to the amount of variation found within and between schools. The significances of school and country/economy differences are less important here; effect sizes may give a better summary of salience of constructs in the between-school variance. For TALIS, though, providing clear recommendations regarding the desirability and feasibility of describing between- and within-school variation is not very straightforward. This relates to two main factors: 1) the sample design, which involves the selection of schools, then the selection of teachers within these schools leading to a clustered sample, and 2) the empirical observation of relatively low amounts of between-school variance from previous TALIS cycles.

The results of variance decomposition analysis, and multilevel models in particular, depend on how schools are defined and organised within countries and economies and how they are chosen for sampling purposes. As in PISA, in some countries and economies, schools may have been defined as administrative units (even if they spanned several geographically separate institutions), as those parts of larger educational institutions that serve students at

the ISCED level concerned, as physical school buildings, as shifts or, rather, from a management perspective (e.g. establishments having a principal). The *TALIS 2018 Technical Report* includes information on how countries and economies defined schools in their respective systems (OECD, 2019[4]). Stratification may also affect the estimate of between-school variation, particularly if variables used for stratification, a process aimed at reducing variation within strata, are associated with between-school differences.

Starting with TALIS 2008, unpublished analyses revealed a relatively low amount of between-school variation, in light of the question of whether aggregation of teacher data at the school level could reasonably be correlated with aggregate student data from the same schools. This low between-school variation is not surprising considering that teachers are not clustered in schools like students. Instead, teachers are clustered at their point of university (or other institutional) training and then dispersed across schools. Students, conversely, often do differ between schools because students' enrolments cluster by demographics and, in some schools, academic abilities, which produce greater between-school variation. TALIS 2013 included tables in which variance components were estimated. This question is still relevant to date in the now formalised TALIS-PISA link international option. One avenue to explaining these relatively low levels of between-school variation relates to the intentionally broad definition of the teacher population, which includes individuals with quite varied backgrounds and responsibilities in terms of gender, age, initial preparation, contract terms, working time, level of experience, satisfaction, subjects taught, practices, beliefs and likely many other demographic, physiological, cultural, professional or social factors not yet observed in TALIS participants.

In the field, some conventional wisdom or rules of thumb would likely consider the between-school variance to be substantial if it amounts to more than 10% of the total variance. Unfortunately, such a value has rarely been exceeded in the case of TALIS but, to the best of our knowledge, this between-school variance has not been evaluated with a larger range of variables from different thematic areas. Studies of student achievement often find larger amounts of variance between schools. In these cases, different thresholds were suggested, e.g. that 25% or less be considered low – see Foy (2004[103]). Values can be expected to vary both by topic (and related measures) and by country/economy. Previous sections in this plan may provide guidance on which variables could be reviewed for substantial amounts of variation between schools, e.g. related to characteristics of the school administration that could be expected to vary more substantially in light of autonomy allocated at the school level for leadership, certain unbalanced characteristics related to student intake and related policies (e.g. on diversity, or a notion such as school climate). For now, it can be noted that teachers appear to vary considerably more than schools.

It is usual to express the decomposition of the total variance into the between-school variance and the within-school variance by the coefficient of intra-class correlation, also denoted *Rho*. Mathematically, this index is equal to the following in the simple case:

$$Rho = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}}$$

The index provides an indication of the percentage of variance that lies between schools. A low intra-class correlation indicates that teachers within schools are "performing" (or, in the context of TALIS, "behaving", "observing" or "perceiving") similarly to each other in the school, while higher values point towards larger differences between schools. However, the sample designs used in TALIS were optimised to derive useful estimates at the system level or, in light of countries and economies reporting interests, at the level of explicit strata,

e.g. regions or other entities. Any variance decomposition analysis would need to be configured carefully to account for the complex, stratified, multi-stage sample designs in TALIS to arrive at proper interpretation of the calculated intra-class correlations, especially if sample designs may violate the underlying assumptions supporting the analysis of variance models. Applying the theoretical aspects in Foy (2004[103]) to TALIS, it might be appropriate to consider three variance components: 1) between strata, 2) between schools within strata, and 3) between teachers within schools. TALIS has no component equivalent to the between classrooms within-school component.

With regard to reporting, total variation in a measure of interest could be reported as the square of the standard deviation within each country or economy. However, this would make for estimates on a scale that, in our view, cannot be easily and directly understood by readers. Further, due to the stratified and clustered nature of the sample design(s), the sum of the between- and within-school variation components, as an estimate from a sample, may not necessarily add up to the total variance.

For the purpose of TALIS, which may be to estimate and highlight larger or lower proportions of variation within and between schools – presumably in relation to a hypothesis, research interest or a policy framework aimed at consistent implementation within schools, across schools, or both – could be reported as the intra-class correlation directly as a percentage with appropriate guidance for interpretation in the text. The analysis suggested in previous sections of this plan do not rely on the analysis of variance components.

Modelling of changes across TALIS cycles

This third data collection of TALIS presents a strong desire to discuss the changes in teacher perceptions, teaching behaviours, school conditions, and principal leadership between the repeated surveys of TALIS 2008, 2013 and 2018. Since these data are not longitudinally sampled where the same teachers or schools are tracked and re-surveyed every five years, observed changes in the variables can only be discussed at the systems-level, e.g. by country/economy. Analyses can test whether a systems-level change is significantly distinct between survey year cycles (see above, Section 4.2). In countries and economies where data collection was mixed – both paper and electronic collections in the 2018 cycle – inclusion of a variable for the delivery method could test change due to the data collection method.

Analysis of change using cross-sectional survey data can identify overall "net" change across the aggregate of teachers and schools (Firebaugh, 1997[104]; Lynn, 2009[105]). To discern among the net change if and how different individuals improved or worsened on a measure or if any group changes within a measure, including filtering the differences of age, cohort, or period effects, can only be tested with longitudinal data that followed teachers across cycles (Firebaugh, 1997[104]). This longitudinal type of data collection is not the design of TALIS.

For example, if a change was found in the aggregate responses to the survey item regarding the barrier to professional development due to family responsibilities, models would likely want to be constructed to identify whether the change is true and distinct from the changing age demographics of the teachers in the surveys (age effect), the attitudes related to changing norms for different stages of the life course (cohort effect), or the evolving definition of "family" (period effect). The cross-sectional data collection of TALIS cannot discern these differences but the selection of respondent data can limit the data only to the consistent aged population among all cycles by excluding the aging-in and aging-out populations (Firebaugh, 1997[104]). The cohort effect can be specified with recoding linear

age into cohort dummy variables in the pooled dataset (Firebaugh, 1997[104]).[15] The period effect can be controlled with survey year dummy variables (Firebaugh, 1997[104]). With these age, period and cohort accommodations, a regression model can specify the 2018 measure as the dependent variable and the prior cycle measures as the focal independent variables. Care must be exercised to prevent over-identification of the model (Firebaugh, 1997[104]). These models would not be able to distinguish if change comes from age, cohort, or period effects. Rather, these careful specifications act to constrain the "noise" in each of the full datasets and, thus, confirm if there is true net change and, if so, the direction of the change between surveys.

If the model does identify a significant net change, then subsequent models can establish whether specific variables in TALIS significantly and substantively relate to the net change. Analyses attempting to explain reasons for the change between cycles would only be able to specify dependent, independent and control variables that have remained consistently measured across the snapshots of data collection cycles. There are few TALIS items that meet this criterion. To specify the model, the dependent variable would either need to be recoded to measure the change from one survey year to the next or specify the prior survey year variable as a control variable (Firebaugh, 1997[104]). Further, if the data collection within a country/economy changed wholly from paper in 2013 to electronic in 2018, there will be no way to test if the change is due to collection type since this binary variable will be perfectly correlated with the cycle/year variable. Similarly, the independent and control variables need to model change – this is most commonly done by calculating the change per variable between years (Firebaugh, 1997[104]). An interaction effect (cycle *X) could test if there is a significant change in the impact of an independent variable on the dependent variable over time.

Section 2, above, listed any items where this type of consistency has been maintained. The constructs of self-efficacy, classroom disciplinary climate, job satisfaction, and professional development needs that have remained consistently asked across survey years would require exhaustive testing and anchoring to assess change across TALIS cycles. When comparing items across survey cycles, data users are reminded that the 2018 averages of Belgium and the United Arab Emirates need to be subset to match 2013. The list of countries and economies for which such a situation applies was revised after the MS data collection.

If the goal of analyses is to establish "international change" on key measures, then the overall averages for each survey year need to be subset to ensure an apples-to-apples comparison. A simple comparison of the "international average" from each TALIS cycle will not produce accurate statistics since participation fluctuated across data collections. If the purpose is to plot and compare three time points for 2008 to 2013 to 2018, then calculations would be limited to those three-time participant countries and economies. Weights need to be applied per TALIS year. As discussed in Section 4.2, reporting some measure of direction and size of change along the model parameters, in addition to the overall *p*-value to evaluate the magnitude of the change between survey years, is recommended.

[15] An alternative would be to create dummy variables for the aging-in and aging-out populations, but the multi-collinearity of these dummy variables with the cohort dummy variables prohibit this option.

References

Ainley, J. and R. Carstens (2018), *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework*, OECD Publishing, Paris, http://dx.doi.org/10.1787/799337c2-en. [3]

Anderson, N. and M. West (1998), "Measuring climate for work group innovation: Development and validation of the team climate inventory", *Journal of Organizational Behavior*, Vol. 19/3, pp. 235-258, http://dx.doi.org/10.1002/(SICI)1099-1379(199805)19:3<235::AID-JOB837>3.0.CO;2-C. [80]

Andrews, P. (2007), "Barriers to innovation", *Leadership Excellence*, Vol. 24/10, p. 19. [76]

Aneshensel, C. (2012), *Theory-based Data Analysis for the Social Sciences*, Sage Publications, Inc., Thousand Oaks, CA. [98]

Bakker, A. and E. Demerouti (2007), "The Job Demands-Resources model: State of the art", *Journal of Managerial Psychology*, Vol. 22/3, pp. 309-328, http://dx.doi.org/10.1108/02683940710733115. [77]

Baumert, J. et al. (2010), "Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress", *American Educational Research Journal*, Vol. 47/1, pp. 133-180, http://dx.doi.org/10.3102/0002831209345157. [48]

Becker, A. (ed.) (2014), *TALIS 2013 User Guide for the International Database List*, OECD, Paris, http://www.oecd.org/education/school/TALIS-2013-User-guide.pdf. [102]

Berliner, D. (2002), "Comment: Educational research: The hardest science of all", *Educational Researcher*, Vol. 31/8, pp. 18-20, http://dx.doi.org/10.3102/0013189X031008018. [5]

Bethlehem, J. (2009), *Applied Survey Methods: A Statistical Perspective*, Wiley, New York, NY. [85]

Blömeke, S. et al. (2018), "Effects of job motives, teacher knowledge, and school context on beginning teachers' commitment to stay in the profession: A longitudinal study in Germany, Taiwan, and the United States", in Akiba, M. and G. LeTendre (eds.), *International Handbook of Teacher Quality and Policy*, Routledge, Oxford. [46]

Boyd, D. et al. (2011), "The influence of school administrators on teacher retention decisions", *American Educational Research Journal*, Vol. 48/2, pp. 303-333, http://dx.doi.org/10.3102/0002831210380788. [41]

Butt, G. et al. (2005), "Teacher job satisfaction: Lessons from the TSW Pathfinder Project", *School Leadership & Management*, Vol. 25/5, pp. 455-471, http://dx.doi.org/10.1080/13634230500340807. [35]

Caprara, G. et al. (2003), "Efficacy beliefs as determinants of teachers' job satisfaction", *Journal of Educational Psychology*, Vol. 95/4, pp. 821-832, http://dx.doi.org/10.1037/0022-0663.95.4.821. [64]

Chapman, C. et al. (eds.) (2015), *Routledge International Handbook of Educational Effectiveness and Improvement Research: Research, Policy, and Practice*, Routledge, Abingdon; New York, NY. [67]

Chen, F. (2007), "Sensitivity of goodness of fit indexes to lack of measurement invariance", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 14/3, pp. 464-504, http://dx.doi.org/10.1080/10705510701301834. [95]

Cheung, G. and R. Rensvold (2002), "Evaluating goodness-of-fit indexes for testing measurement invariance", *Structural Equation Modeling*, Vol. 9/2, pp. 233-255, http://dx.doi.org/10.1207/S15328007SEM0902_5. [111]

Cheung, G. and R. Rensvold (1999), "Testing factorial invariance across groups: A reconceptualization and proposed new method", *Journal of Management*, Vol. 25/1, pp. 1-27, http://dx.doi.org/10.1177/014920639902500101. [112]

Clarke, D. and H. Hollingsworth (2002), "Elaborating a model of teacher professional growth", *Teaching and Teacher Education*, Vol. 18/8, pp. 947-967, https://doi.org/10.1016/S0742-051X(02)00053-7. [56]

Coburn, C., W. Penuel and K. Geil (2013), *Research-Practice Partnerships: A Strategy for Leveraging Research for Educational Improvement in School Districts*, William T. Grant Foundation, New York, NY. [11]

Collie, R., J. Shapka and N. Perry (2012), "School climate and social-emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy", *Journal of Educational Psychology*, Vol. 104/4, pp. 1189-1204, http://dx.doi.org/10.1037/a0029356. [61]

Creemers, B. and L. Kyriakides (2008), *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*, Routledge, Abingdon, https://lib.ugent.be/nl/catalog/rug01:001240853. [10]

Crossman, A. and P. Harris (2006), "Job satisfaction of secondary school teachers", *Educational Management Administration and Leadership*, Vol. 34/1, pp. 29-46, http://dx.doi.org/10.1177/1741143206059538. [36]

Darling-Hammond, L. (2000), "Teacher quality and student achievement: A review of state policy evidence", *Education Policy Analysis Archives*, Vol. 8/1, pp. 1-44, http://dx.doi.org/10.14507/epaa.v8n1.2000. [52]

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: Journal of the European Survey Research Association*, Vol. 2/1, pp. 33-46, http://dx.doi.org/10.18148/SRM/2008.V2I1.365. [113]

Decristan, J. et al. (2015), "Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding", *American Educational Research Journal*, Vol. 52/6, pp. 1133-1159, http://dx.doi.org/10.3102/0002831215596412. [21]

Desimone, L. (2009), "Improving impact studies of teachers' professional development: Toward better conceptualizations and measures", *Educational Researcher*, Vol. 38/3, pp. 181-199, http://dx.doi.org/10.3102/0013189X08331140. [24]

DeYoung, C., J. Peterson and D. Higgins (2002), "Higher-order factors of the Big Five predict conformity: Are there neuroses of health?", *Personality and Individual Differences*, Vol. 33/4, pp. 533-552, https://doi.org/10.1016/S0191-8869(01)00171-4. [78]

Dinham, S. and C. Scott (1998), "A three domain model of teacher and school executive career satisfaction", *Journal of Educational Administration*, Vol. 36/4, pp. 362-378, http://dx.doi.org/10.1108/09578239810211545. [37]

Dumont, H., D. Istance and F. Benavides (eds.) (2010), *The Nature of Learning: Using Research to Inspire Practice*, Educational Research and Innovation, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264086487-en. [81]

Farley-Ripple, E. et al. (2018), "Rethinking connections between research and practice in education: A conceptual framework", *Educational Researcher*, Vol. 47/4, pp. 235-245, http://dx.doi.org/10.3102/0013189X18761042. [6]

Fauth, B. et al. (2014), "Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes", *Learning and Instruction*, Vol. 29, pp. 1-9, http://dx.doi.org/10.1016/j.learninstruc.2013.07.001. [22]

Firebaugh, G. (1997), *Analyzing Repeated Surveys*, Quantitative Applications in the Social Sciences, Sage Publications, Inc., Thousand Oaks, CA. [104]

Firestone, W. and J. Pennell (1993), "Teacher commitment, working conditions, and differential incentive policies", *Review of Educational Research*, Vol. 63/4, pp. 489-525, http://dx.doi.org/10.2307/1170498. [51]

Foy, P. (2004), "Intraclass Correlation and Variance Components as Population Attributes and Measures of Sampling Efficiency in PIRLS 2001", in *Paper presented at the 1st IEA International Research Conference, Nicosia, Cyprus, 11-13 May 2004*, https://www.iea.nl/sites/default/files/2019-03/IRC2004_Foy.pdf. [103]

Fulton, K., I. Yoon and C. Lee (2005), *Induction Into Learning Communities*, National Commission on Teaching and America's Future, Washington, DC, https://files.eric.ed.gov/fulltext/ED494581.pdf. [74]

Goddard, Y., R. Goddard and M. Tschannen-Moran (2007), "A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools", *Teachers College Record*, Vol. 109/4, pp. 877-896. [25]

Gouzi, N. and J. Dumais (2017), *Impact of changing the definition of 'participating teachers' using TALIS 2013 ISCED Level 2 data (unpublished)*. [91]

Gustafsson, B. and M. Johansson (1999), "In search of smoking guns: What makes income inequality vary over time in different countries?", *American Sociological Review*, Vol. 64/4, pp. 585-605, http://dx.doi.org/10.2307/2657258. [99]

Hallinger, P. (2008), "Methodologies for Studying School Leadership: A Review of 25 Years of Research Using the PIMRS", *Paper presented at Research on Schools, Neighborhoods, and Communities: Toward Civic Responsibility, American Educational Research Association Annual Meeting, New York NY, 24-28 March 2008*. [69]

Harris, A., D. Muijs and M. Crawford (2003), *Deputy and Assistant Heads: Building Leadership Potential*, National College for School Leadership, Nottingham. [71]

Hattie, J. (2009), *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*, Routledge, London. [32]

Holzberger, D., A. Philipp and M. Kunter (2014), "Predicting teachers' instructional behaviors: The interplay between self-efficacy and intrinsic needs", *Contemporary Educational Psychology*, Vol. 39/2, pp. 100-111, http://dx.doi.org/10.1016/j.cedpsych.2014.02.001. [44]

Honig, M. and C. Coburn (2008), "Evidence-based decision making in school district central offices", *Educational Policy*, Vol. 22/4, pp. 578-608, http://dx.doi.org/10.1177/0895904807307067. [12]

Hoy, W. and A. Woolfolk (1993), "Teachers' sense of efficacy and the organizational health of schools", *The Elementary School Journal*, Vol. 93/4, pp. 355-372, https://doi.org/10.1086/461729. [75]

Hsu, M., S. Hou and H. Fan (2011), "Creative self-efficacy and innovative behavior in a service setting: Optimism as a moderator", *Journal of Creative Behavior*, Vol. 45/4, pp. 258-272, http://dx.doi.org/10.1002/j.2162-6057.2011.tb01430.x. [57]

Huber, S. and D. Muijs (2010), "School leadership effectiveness: The growing insight in the importance of school leadership for the quality and development of schools and their pupils", in Huber, S. (ed.), *School Leadership: International Perspectives*, Springer, Dordrecht. [70]

Hu, L. and P. Bentler (1999), "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives", *Structural Equation Modeling: Multidisciplinary Journal*, Vol. 6/1, pp. 1-55, http://dx.doi.org/10.1080/10705519909540118. [106]

Hulpia, H., G. Devos and Y. Rosseel (2009), "The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment", *School Effectiveness and School Improvement*, Vol. 20/3, pp. 291-317, http://dx.doi.org/10.1080/09243450902909840. [72]

Hurt, H., K. Joseph and C. Cook (1977), "Scales for the measurement of innovativeness", *Human Communication Research*, Vol. 4/1, pp. 58-65, http://dx.doi.org/10.1111/j.1468-2958.1977.tb00597.x. [79]

Ingvarson, L., M. Meiers and A. Beavis (2005), "Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy", *Education Policy Analysis Archives*, Vol. 13/10, pp. 1-28, http://dx.doi.org/10.14507/epaa.v13n10.2005. [33]

Isoré, M. (2009), "Teacher Evaluation: Current Practices in OECD Countries and a Literature Review", *OECD Education Working Papers*, No. 23, OECD, Paris, http://dx.doi.org/10.1787/223283631428. [30]

Jensen, B. et al. (2016), *Beyond PD: Teacher Professional Learning in High-Performing Systems*, National Center on Education and the Economy, Washington, DC, http://ncee.org/wp-content/uploads/2015/08/BeyondPDDec2016.pdf. [53]

Karabenick, S. and T. Urdan (eds.) (2010), "The decade ahead: Applications and contexts of motivation and achievement", *Advances in Motivation and Achievement*, Vol. 16, Part B, http://dx.doi.org/10.1108/S0749-7423(2010)000016B008. [39]

Kelchtermans, G. (2006), "Teacher collaboration and collegiality as workplace conditions: A review", *Zeitschrift für Pädagogik*, Vol. 52/2, pp. 220-237. [28]

Klassen, R. et al. (2009), "Exploring the validity of a teachers' self-efficacy scale in five countries", *Contemporary Educational Psychology*, Vol. 34/1, pp. 67-76, http://dx.doi.org/10.1016/j.cedpsych.2008.08.001. [65]

Klassen, R. and M. Chiu (2010), "Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress", *Journal of Educational Psychology*, Vol. 102/3, pp. 741-756, http://dx.doi.org/10.1037/a0019237. [59]

Klieme, E., C. Pauli and K. Reusser (2009), "The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms", in Janík, T. and T. Seidel (eds.), *The Power of Video Studies in Investigating Teaching and Learning in the Classroom*, Waxmann, Münster. [23]

Kyriakides, L., C. Christoforou and C. Charalambous (2013), "What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching", *Teacher and Teacher Education*, Vol. 36, pp. 143-152, http://dx.doi.org/10.1016/j.tate.2013.07.010. [43]

Kyriakides, L. and B. Creemers (2008), "Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: a study testing the validity of the dynamic model", *School Effectiveness and School Improvement*, Vol. 19/2, pp. 183-205, http://dx.doi.org/10.1080/09243450802047873. [49]

Ladhari, R. (2010), "Developing e-service quality scales: A literature review", *Journal of Retailing and Consumer Services*, Vol. 17/6, pp. 464-477, http://dx.doi.org/10.1016/j.jretconser.2010.06.003. [94]

Leithwood, K. et al. (2004), *How Leadership Influences Student Learning*, The Wallace Foundation, New York, NY, https://www.wallacefoundation.org/knowledge-center/Documents/How-Leadership-Influences-Student-Learning.pdf. [13]

Leithwood, K. and C. Riehl (2003), *What We Know about Successful School Leadership. A Report by Division A of AERA*, National College for School Leadership, Nottingham. [73]

Little, J. (1990), "The persistence of privacy: Autonomy and initiative in teachers' professional relations", *Teachers College Record*, Vol. 91/4, pp. 509-536, https://www.tcrecord.org/content.asp?contentid=406. [27]

Lohr, S. (2010), *Sampling: Design and Analysis, Second Edition*, Brooks/Cole, Boston, MA. [96]

Lynn, P. (ed.) (2009), *Methodology of Longitudinal Surveys*, Wiley, Chichester. [105]

Miller, M., M. Brownell and S. Smith (1999), "Factors that predict teachers staying in, leaving, or transferring from the special education classroom", *Exceptional Children*, Vol. 65/2, pp. 201-218, https://doi.org/10.1177/001440299906500206. [62]

National Academy of Education (2008), "Attracting and keeping quality teachers", *White Papers Project: Education Policy Briefing Sheet*, National Academy of Education, Washington, DC. [66]

OECD (2019), *TALIS 2018 Technical Report*, OECD, Paris, http://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf. [4]

OECD (2017), *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264279889-en. [86]

OECD (2017), "Proposal for the TALIS 2018 Reporting Plan", No. EDU/TALIS/GB(2017)2/REV4 (internal document), Directorate for Education and Skills, OECD, Paris. [1]

OECD (2016), *PISA 2015 Results (Volume I): Excellence and Equity in Education*, PISA, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264266490-en. [92]

OECD (2015), "Guiding the Policy and Content Focus of TALIS 2018", No. EDU/INES/TALIS(2015)3 (internal document), Directorate for Education and Skills, OECD, Paris. [2]

OECD (2014), *TALIS 2013 Results: An International Perspective on Teaching and Learning*, TALIS, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264196261-en. [17]

OECD (2014), *TALIS 2013 Technical Report*, OECD, Paris, http://www.oecd.org/education/school/TALIS-technical-report-2013.pdf. [16]

OECD (2013), *Education at a Glance 2013: OECD Indicators*, OECD Publishing, Paris, https://dx.doi.org/10.1787/eag-2013-en. [107]

OECD (2013), *Teaching and Learning International Survey, TALIS 2013: Conceptual Framework*, OECD, Paris, http://www.oecd.org/education/school/TALIS%20Conceptual%20Framework_FINAL.pdf. [34]

OECD (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, Education and Training Policy, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264018044-en. [31]

OECD (2004), "Definitions and Classifications of OECD International Education Statistics", in *OECD Handbook for Internationally Comparative Education Statistics: Concepts, Standards, Definitions and Classifications*, OECD Publishing, Paris, https://dx.doi.org/10.1787/9789264104112-5-en. [90]

Pant, H. (2016), *Wie Wissen aus der Bildungsforschung für die Praxis nutzbar gemacht werden kann [How knowledge from educational research can be used in practice]*, Presentation at the Transfertagung "Nutzung und Nutzen von Evidenz zur Schul- und Unterrichtsentwicklung", Johannes Gutenberg-Universität Mainz, 20 July 2016, http://www.stebis.de/news/Transfertagung-2016_Vortraege-und-Poster/Pant_Vortrag_EviS_20072016_PDF.pdf (accessed on 6 May 2018). [9]

Pashiardis, P., S. Brauckmann and D. Muijs (2011), "Leadership and organisational performance: From research to prescription?", *International Journal of Educational Management*, Vol. 25/1, pp. 45-60, http://dx.doi.org/10.1108/09513541111100116. [42]

Patterson, M. et al. (2005), "Validating the organizational climate measure: Links to managerial practices, productivity and innovation", *Journal of Organizational Behavior*, Vol. 26/4, pp. 379-408, http://dx.doi.org/10.1002/job.312. [83]

Rao, J. and A. Scott (1987), "On simple adjustments to chi-square tests with sample survey data", *The Annals of Statistics*, Vol. 15/1, pp. 385-397, http://dx.doi.org/10.2307/2241089. [93]

Raudenbush, S. and A. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Thousand Oaks, CA. [101]

Reynolds, D. et al. (2014), "Educational effectiveness research (EER): A state-of-the-art review", *School Effectiveness and School Improvement*, Vol. 25/2, pp. 197-230, http://dx.doi.org/10.1080/09243453.2014.885450. [47]

Richardson, P. and H. Watt (2010), "Current and future directions in teacher motivation research", in Urdan, T. and S. Karabenick (eds.), *The Decade Ahead: Applications and Contexts of Motivation and Achievement (Advances in Motivation and Achievement, Volume 16 Part B)*, Emerald Group Publishing Limited, Bingley. [40]

Rowe, K. (2003), "The importance of teacher quality as a key determinant of students' experiences and outcomes of schooling", in Meiers, M. (ed.), *Building Teacher Quality: ACER Research Conference 2003: Proceedings, Melbourne, 19-21 October 2003*, Australian Council for Educational Research, https://research.acer.edu.au/cgi/viewcontent.cgi?article=1001&context=research_conference_2003. [54]

Sahlberg, P. (2011), "Paradoxes of educational improvement: The Finnish experience", *Scottish Educational Review*, Vol. 43/1, pp. 3-23, http://www.scotedreview.org.uk/media/microsites/scottish-educational-review/documents/318.pdf. [18]

Scheerens, J. and R. Bosker (1997), *The Foundations of Educational Effectiveness*, Pergamon, Oxford. [68]

Scherer, R. and T. Nilsen (2016), "The relations among school climate, instructional quality, and achievement motivation in mathematics", in Nilsen, T. and J. Gustafsson (eds.), *Teacher Quality, Instructional Quality and Student Outcomes*, Springer International Publishing, Cham. [50]

Schermelleh-Engel Karin, H. Moosbrugger and H. Müller (2003), "Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures", *Methods of Psychological Research*, Vol. 8/2, pp. 23-74. [109]

Seashore Louis, K. et al. (2010), *Learning from Leadership: Investigating the Links to Improved Student Learning. Final Report of Research to the Wallace Foundation*, University of Minnesota, Minneapolis, MN. [14]

Skaalvik, E. and S. Skaalvik (2010), "Teacher self-efficacy and teacher burnout: A study of relations", *Teaching and Teacher Education*, Vol. 26/4, pp. 1059-1069, http://dx.doi.org/10.1016/j.tate.2009.11.001. [58]

Slavin, R. (2002), "Evidence-based education policies: Transforming educational practice and research", *Educational Researcher*, Vol. 31/7, pp. 15-21, http://dx.doi.org/10.3102/0013189X031007015. [7]

Snedecor, G. and W. Cochran (1989), *Statistical Methods, Eighth Edition*, Iowa State University Press. [97]

Statistics Canada (2009), *Statistics Canada Quality Guidelines, Fifth Edition*, Minister of Industry, Ottawa, https://www150.statcan.gc.ca/n1/en/catalogue/12-539-X2009001. [84]

Stearns, E. et al. (2015), "Collective pedagogical teacher culture and teacher satisfaction", *Teachers College Record*, Vol. 117/8, pp. 1-32, https://www.tcrecord.org/content.asp?contentid=18017. [38]

Steenkamp, J. and H. Baumgartner (1998), "Assessing measurement invariance in cross-national consumer research", *Journal of Consumer Research*, Vol. 25/1, pp. 78-107, http://dx.doi.org/10.1086/209528. [114]

Steiger, J. (1990), "Structural model evaluation and modification: An interval estimation approach", *Multivariate Behavioral Research*, Vol. 25/2, pp. 173-180, http://dx.doi.org/10.1207/s15327906mbr2502_4. [108]

Teddlie, C. and D. Reynolds (eds.) (2000), *The International Handbook of School Effectiveness Research*, Routledge, New York, NY. [8]

Timperley, H. et al. (2007), *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration [BES]*, New Zealand: Ministry of Education, Wellington. [26]

Tschannen-Moran, M. and A. Hoy (2001), "Teacher efficacy: Capturing an elusive construct", *Teaching and Teacher Education*, Vol. 17/7, pp. 783-805, https://doi.org/10.1016/S0742-051X(01)00036-1. [60]

Tschannen-Moran, M. and A. Woolfolk Hoy (2007), "The differential antecedents of self-efficacy beliefs of novice and experienced teachers", *Teaching and Teacher Education*, Vol. 23/6, pp. 944-956, http://dx.doi.org/10.1016/j.tate.2006.05.003. [45]

UNECE (2009), *Making Data Meaningful. Part 2: A Guide to Presenting Statistics*, United Nations, Geneva. [88]

UNSD (2005), "Designing Household Survey Samples: Practical Guidelines", *Series F*, No. 98, Department of Economic and Social Affairs, Statistics Division, United Nations, https://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf. [87]

Wainer, H. (1984), "How to display data badly", *The American Statistician*, Vol. 38/2, pp. 137-147, http://dx.doi.org/10.2307/2683253. [89]

Wallace Foundation Staff (2011), *Research Findings to Support Effective Educational Policies: A Guide for Policymakers*, The Wallace Foundation, New York, NY, https://www.wallacefoundation.org/knowledge-center/Documents/Findings-to-Support-Effective-Educational-Policy-Making.pdf. [15]

Watt, H. and P. Richardson (2008), "Motivations, perceptions, and aspirations concerning teaching as a career for different types of beginning teachers", *Learning and Instruction*, Vol. 18/5, pp. 408-428, http://dx.doi.org/10.1016/j.learninstruc.2008.06.002. [19]

Watt, H. et al. (2012), "Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale", *Teaching and Teacher Education*, Vol. 28/6, pp. 791-805, http://dx.doi.org/10.1016/j.tate.2012.03.003. [20]

Weiss, E. (1999), "Perceived workplace conditions and first-year teachers' morale, career choice commitment, and planned retention: A secondary analysis", *Teaching and Teacher Education*, Vol. 15/8, pp. 861-879, https://doi.org/10.1016/S0742-051X(99)00040-2. [63]

Wenglinsky, H. (2002), "How schools matter: The link between teacher classroom practices and student academic achievement", *Education Policy Analysis Archives*, Vol. 10/12, pp. 1-30, http://dx.doi.org/10.14507/epaa.v10n12.2002. [55]

Wooldridge, J. (2010), *Econometric Analysis of Cross Section and Panel Data, Second Edition*, MIT Press, Cambridge, MA. [100]

Yi, M., K. Fiedler and J. Park (2006), "Understanding the role of individual innovativeness in the acceptance of IT-based innovations: Comparative analyses of models and measures", *Decision Sciences*, Vol. 37/3, pp. 393-426, http://dx.doi.org/10.1111/j.1540-5414.2006.00132.x. [82]

Yoon, K. et al. (2007), "Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement", *Issues & Answers Report, REL 2007*, No. 033, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest, https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf. [29]

Yu, C. (2002), "Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes", *A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Education*, University of California, Los Angeles, https://www.statmodel.com/download/Yudissertation.pdf. [110]