



# Developing a shared cluster construct of instructional leadership in TALIS

Jelena Veletić\*, Rolf Vegar Olsen

Centre for Educational Measurement (CEMO), University of Oslo, Norway

## ARTICLE INFO

### Keywords:

TALIS  
Instructional leadership  
Multilevel structural equation modelling  
Shared cluster construct  
Teacher job satisfaction

## ABSTRACT

In the Teaching and Learning International Survey (TALIS), instructional leadership is measured by the self-reports of principals on three items only. When this measure is investigated together with teacher satisfaction with current work environment, no significant associations were found in the Nordic countries participating in the TALIS 2013 round. This paper argues that a potential reason for this might be the severely underrepresented construct of instructional leadership. As an alternative approach, teacher data from the same study are used to establish two important dimensions of instructional leadership at the school level: 1) managing the instructional program and 2) developing the school learning climate. Applying multilevel structural equation modelling (MSEM), we establish two shared cluster constructs at the school level and observe significant modest relationships between these constructs and teacher job satisfaction with current work environment. The paper brings to our attention the different approaches for interpreting, exploring, and making sense of instructional leadership in international large-scale studies, such as TALIS, from the joint perspective of teachers.

## 1. Introduction

School leadership is increasingly viewed as a key factor in education reforms and is currently one of the features of educational systems that is receiving high attention in several international large-scale assessments (Pont, Nusche, & Moorman, 2008; Rutkowski et al., 2013). Out of many competing school leadership conceptualizations (e.g., “transactional,” “distributed,” “transformational”), instructional leadership is one of the most used and investigated. The underlying conceptualization of instructional leadership assumes clear school goals, motivation of staff and students, supervision of progress, and a distinct focus on academic outcomes (Hallinger, 2005). Principals who emphasize high-quality instruction, give instructional feedback to teachers, and support the use of assessment in the classroom are considered to be strong instructional leaders. It is hypothesized that instructional leadership affects teacher attitudes and behaviors as well as student learning outcomes (Hallinger & Wang, 2015; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Liebowitz & Porter, 2019; Robinson, Lloyd, & Rowe, 2008). This framework and perspective on instructional leadership was first established by Hallinger and Murphy (1985) and has subsequently been continuously discussed and revised (Boyce & Bowers, 2018; Hallinger, 2010, 2011).

The Organization for Economic Co-operation and Development (OECD) organizes the Teaching and Learning International Survey (TALIS) to study principals’ and teachers’ working conditions, beliefs,

and attitudes, as well as the larger school environment, including leadership practices. One of the studied constructs is the degree to which instructional leadership is implemented at a school. This construct is represented by a measure that is based on the self-reports of principals on three items (OECD, 2014).

From a conceptual point of view, it is highly unlikely that broad and complex constructs, such as instructional leadership, can be captured well by only three items. Usually, instruments that measure school leadership are composite questionnaires with a considerable number of items. To give an example, the Hallinger’s PIRMS scale for assessing instructional leadership consists of 50 items (Hallinger & Wang, 2015). Compared to such fine-grained and extensive scales, the measure of instructional leadership in TALIS likely exemplifies a severely under-represented construct. Furthermore, instructional leadership can be perceived differently by those who are led and those who are leaders (Urlick & Bowers, 2017). In contrast to previous research, the initial analysis for this paper establishes that there is a non-significant relationship between the existing instructional leadership measure and the teacher job satisfaction with current work environment measure across a range of contexts (Ansley, Houchins, & Varjas, 2019; Qadach, Schechter, & Da’as, 2020).

To remedy this situation, the current study proposes an alternative approach for operationalizing and measuring instructional leadership using the data available from the same study. First, we suggest staying

\* Corresponding author.

E-mail address: [jelena.veletic@cemo.uio.no](mailto:jelena.veletic@cemo.uio.no) (J. Veletić).

closer to the conceptualization found in the literature—the well-established Hallinger-Murphy model. Second, we suggest examining the concept of instructional leadership using teacher data to appropriately reflect their collective perception. Third, we regard collective reports from teachers with shared school characteristic reflecting on instructional leadership practice to be a more trustworthy source than reports provided by principals or other single entities in a school with a leadership role.

In order to establish a new measure of instructional leadership based on responses from teachers, this paper gives an account of how the dimensions of instructional leadership can be conceptualized as a shared perception of teachers using data from TALIS 2013. Furthermore, the new measure is validated empirically using the data from the same study and the association with teacher job satisfaction with the working environment is re-examined. The presented analyses should be regarded as setting up an argument for a principled new approach to operationalizing leadership in large-scale studies. Finally, since cultural features likely affect leadership practices and how individuals report on them (Brewer, Okilwa, & Duarte, 2020; Hallinger, 2018), we conducted analyses in a more homogenous group of countries. Specifically, the analyses use data from the Nordic countries participating in TALIS 2013 (Norway, Sweden, Denmark, and Finland)<sup>1</sup>.

## 2. Theoretical framework

### 2.1. School leadership

School leadership is recognized as an important factor in the area of school development, change, effectiveness, and improvement (Bush, 2009; González-Falcón, García-Rodríguez, Gómez-Hurtado, & Carrasco-Macías, 2019; Hallinger & Heck, 1998; Huber & Muijs, 2010; Leithwood, Harris, & Hopkins, 2008; Robinson et al., 2008). The most common models in education research are instructional leadership, distributed leadership, and transformational leadership (Bush & Glover, 2014; Gumus, Bellibas, Esen, & Gumus, 2018). Each of the models places emphasis on specific leadership features; however, there is also a great deal of commonality between them. Consequently, the dominant theories and models of leadership are not mutually exclusive frameworks for understanding how functions are governed, distributed, and shared at a school.

Studies of the association between school leadership and student learning outcomes provide a complex and inconsistent picture. Some authors report no association between school leadership and student achievement (Krüger et al., 2007; Krüger, Witziers, & Slegers, 2007) while others document small effects (Kyriakides, Creemers, Antoniou, & Demetriou, 2010; Witziers, Bosker, & Krüger, 2003). In this context, it is important to note that it may not be reasonable to assume any direct associations of school leadership to student outcomes because outcomes at the student level are more distal phenomena in comparison to more proximal characteristics, such as observations at the teacher/classroom level (Leithwood et al., 2008). The latter further explains that school leaders can improve teaching and learning indirectly and most powerfully through their influence on staff motivation, commitment, and working conditions. Consequently, research should be designed to study the indirect effects of leadership on student outcomes as mediated through working conditions, teacher well-being, and instructional activities (Ladd, 2009; Pont et al., 2008). As a first step in this chain, our paper examines the relationship between leadership and teacher job satisfaction with current work environment.

<sup>1</sup> Iceland is a part of the Nordic group of countries. However, the country did not authorize the release of the data as a part of the international database. Consequently, Iceland was not included in the analysis.

### 2.2. Instructional leadership

Instructional leadership is the most emphasized model in terms of its potential for fostering student learning outcomes and the quality of teaching and learning (Day, Gu, & Sammons, 2016; Hallinger, 2003; 2019; Heck, Larsen, & Marcoulides, 1990; Louis, Dretzke, & Wahlstrom, 2010; O'Donnell & White, 2005; Robinson et al., 2008). Hallinger and Murphy (1985) developed a conceptual framework and a corresponding scale for measuring instructional leadership—the Principal Instructional Management Rating Scale (PIMRS). This framework describes three dimensions of instructional leadership with ten corresponding functions. The first dimension, *defining the school mission*, involves two functions: framing and communicating school goals. This dimension emphasizes goals concerning academic and learning achievements. A clearly communicated school mission aids teachers and other stakeholders in determining priorities and focusing their attention and activity scope. The second dimension, *managing the instructional program*, involves three functions: curriculum coordination, evaluation and supervision of instruction, and monitoring of student progress. This dimension highlights that what happens in the classrooms is not the responsibility of individual teachers. Successful schools have leaders who take on the responsibility to monitor, supervise, and motivate staff to adopt high-quality curriculum and instructional practices. The third dimension, *developing the school learning climate*, involves five functions: protection of instructional time, provision of incentives for teachers, provision of incentives for learning, promotion of professional development, and continuity of high principal visibility in the school. Altogether, provided by principals and school management teams, these functions build a context in which teachers work, collaborate, and develop towards a set of joint goals.

### 2.3. Measuring instructional leadership in TALIS: principals' and teachers' perspectives

Although listed as a top priority among countries participating in TALIS, school leadership is rather modestly covered by the study. The principal questionnaire includes only a limited number of items, where principals are asked to report on their leadership practices. To be more specific, five items from the principal questionnaire were included in the instrument with the intention to measure instructional leadership (see Table 2). Two items were subsequently excluded, when forming the final scale, resulting in a scale based on three items. It is reasonable to assume that these two items proved to function poorly. One possible reason could be because the three remaining items (PQ2-PQ4) are similarly worded. They all start with the phrase "I took actions to..." which means that they probably cluster together empirically, resulting in a poorly fitted measurement model when all five items are included (Arnulf, Larsen, Martinsen, & Egeland, 2018; Arnulf, Larsen, Martinsen, & Bong, 2014).

Teachers' perspective on instructional leadership is not examined directly in TALIS. This could be challenging for those interested in studying leadership and giving policy recommendations because the evidence shows discrepancies between teachers' and principals' perspectives (Urlick & Bowers, 2017). Without a joint and shared understanding of how leadership is exercised at a school, a lack of responsiveness towards shared goals is likely to occur. Principals may, for instance, report that they invest time and effort in managing the instructional program but that would not be considered to be a trustworthy report of an actual (observable) practice if teachers simultaneously report that they are left on their own in their classrooms. Accordingly, we give preference to developing measures of leadership from collective reports of teachers on school characteristics, environment, and dynamics. We focus on items from the teacher questionnaire that refer to joint school characteristics from which valuable interpretations about school leadership can be drawn. As suggested by Stapleton, Yang, and Hancock (2016), this is the recommended

approach when studying so-called “shared cluster constructs.”

In accordance with this recommendation, indicators are carefully selected to represent evaluations of shared perspectives on leadership. The literature suggests that effective instructional leadership although mostly focused on principals, can be practiced in collaboration with teachers and other administrators (Franz Coldren & Spillane, 2007; Marks & Printy, 2003; Osborne-Lampkin, Folsom, & Herrington, 2015). Inspecting the teacher questionnaire in light of these recommendations and the theory of instructional leadership, two sets of items are identified as potentially relevant indicators of shared school practices that reflect instructional leadership functions. First set of items provides teachers with the opportunity to report on feedback given to them by various entities both within and outside the school. This set of items captures important actions and practices related to how instruction is managed within a school (observations of teaching, student surveys about teaching, access to teacher content knowledge, student test score analyses, teacher self-evaluations, and parent surveys). Second set of items, represents teachers’ reports about their schools more generally in terms of support, mentoring, and professional development. These items reflect how learning at all levels is supported within a school. By providing support, feedback, and training for teachers, a school protects instructional time, promotes professional development, and provides incentives for teachers—all important facets of instructional leadership (Hallinger & Murphy, 1985).

#### 2.4. Teacher job satisfaction: measurement and relevance

From a measurement perspective, job satisfaction has most frequently been studied through a global perspective as a unidimensional construct (Liu & Werblow, 2019; Skaalvik & Skaalvik, 2010, 2011). However, it is important to recognize the complexity of this measure as well as its multidimensional structure (Evans, 1997; Judge, Thoresen, Bono, & Patton, 2001; Weiss, 2002). Hence, this construct is often studied as satisfaction with different facets of work (e.g., satisfaction with salary, satisfaction with supervision) (Stanton et al., 2002). TALIS partially recognizes this by distinguishing between two different measures of teacher job satisfaction: the measure of teacher satisfaction with the profession and the measure of teacher satisfaction with the current work environment. The two scales are only weakly positively correlated in Sweden, Denmark, and Finland having a correlation of .113, .156, and .194; respectively, with non-significant correlation in Norway (OECD, 2014). The satisfaction with profession scale is a more global evaluation of the decision to become a teacher and how the teaching profession is valued in society. The satisfaction with the current work environment scale is focused on satisfaction related to work at a particular school. As such, only the latter dimension reflects a school-level characteristic. In the current study, we propose to use the measure of teacher satisfaction with current work environment as a relevant external criterion for validating measures of school leadership.

The choice is motivated by previous research where the positive relationship between educational leadership and teacher job satisfaction is found across a range of contexts (Benoliel, Shaked, Nadav, & Schechter, 2019; Bogler, 2001; Burkhauser, 2017; Çoğaltay, Yalçın, & Karadağ, 2016; Hariri, Monypenny, & Prideaux, 2012) and across a range of leadership styles (Bogler, 2001; Cerit, 2009; Sun & Xia, 2018). Specifically, instructional leadership is found to be positively associated with teacher job satisfaction through perception of support (Ansley et al., 2019), collective teacher efficacy and shared vision (Qadach et al., 2020), and career and working conditions (Shen, Leslie, Spybrook, & Ma, 2012). A supportive working environment and adequate working conditions are among the most important factors in this relationship (Burkhauser, 2017; Johnson, Kraft, & Papay., 2011; Klassen & Anderson, 2009). The context in which teachers work is also closely associated with teacher job satisfaction (Benoliel et al., 2019; Dou, Devos, & Valcke, 2017; Sebastian & Allensworth, 2012; Sims, 2019). The magnitude of these associations are mostly small (Liu, Bellibaş, &

Gümüş, 2020), which might be caused by teacher job satisfaction being a non-linear function of age and years of working experience (Clark, Oswald, & Warr, 1996; Ma & MacMillan, 1999).

Teacher job satisfaction further relates to teachers’ intention to stay at a school and is consequently an important predictor of teacher retention (Kelly, Cespedes, Clarà, & Danaher, 2019; Skaalvik & Skaalvik, 2011) and teacher turnover (Holtom, Mitchell, Lee, & Eberly, 2008; Ingersoll, 2002; Qin, 2019). It is also found, although not consistently, that teacher job satisfaction is linked to student learning outcomes (Banerjee, Stearns, Moller, & Mickelson, 2017; Caprara, Barbaranelli, Steca, & Malone, 2006; Dutta & Sahney, 2016).

#### 2.5. Importance of a wider context for leadership research

How leadership is perceived and enacted may reflect wider societal norms and values. Hallinger (2018) explores the influence of several school context types (e.g., economic, political, national, cultural) on instructional leadership, showing the importance of a wider context for leadership practice. Thus, when leadership practice at schools is examined and compared across countries, it is important to apply analytical approaches that are sensitive to the societal and cultural contexts within which leadership exists (Hallinger & Leithwood, 1998; Leithwood & Duke, 1998; Walker & Dimmock, 2002). Naturally, the aim of conducting international studies is to make analytical use of variability in policies, practices, and outcomes across countries. Nevertheless, it is well known that in many cases—particularly when based on self-reports—scales may not be invariant across countries (van de Vijver & Tanzer, 2004). In consequence, the present analysis is narrowed down to a selection of more homogenous countries—the Nordic countries taking part in TALIS 2013. As will be returned to in the discussion, a more extensive approach with a focus on measurement invariance is needed to establish the proposed procedure as viable for an international and largely globally targeted survey.

Although these countries share cultural and linguistic similarities, sufficient differences exist in how education is governed (Ahola, Hedmo, Thomsen, & Vabø, 2014). In addition to geographical proximity, Norway, Sweden, and Denmark are also similar with respect to language, historical development, socioeconomic conditions, and wider societal/political/cultural features in general. To be more specific, the Nordic countries have egalitarian societal systems with free access to social services (including schools), strong institutional coordination, general high trust in public/government institutions (including schools), many shared curriculum features, and relatively high (socio)economic and gender equality (Ludvigsen, 2016). Finland also shares many of the same characteristics but has a uniquely different language. The Nordic countries have a long history of framing leadership as a function entrusted to “first among equals.” In this manner, schools developed into relatively flat hierarchies, where the professional identity of school leaders is grounded in the teaching profession, promoting democracy and co-responsibility as fundamental social values (Møller, 2009).

### 3. Research model

How instructional leadership is measured by TALIS is the main issue addressed in this study. As a validation step, the study also examines its association with teacher job satisfaction with current work environment. In the first phase, the study makes use of the instructional leadership scale already developed by OECD and examines its association with teacher job satisfaction with current work environment. In this phase, instructional leadership is measured from the principals’ perspective as suggested by OECD. In the second phase, the study proposes a new measure of instructional leadership based on teacher reports about features of the school environment. Teacher reports are suggested to indicate two dimensions of instructional leadership referred to in the literature: 1) *managing the instructional program* and 2) *developing the school learning climate*. The new measure of instructional leadership is

then applied to re-examine the association with teacher job satisfaction with current work environment. The hypothesized and examined model is presented in Fig. 1. Three latent constructs, represented by ovals in Fig. 1, are measured by 13 indicators (TQ1 - TQ13) representing teacher reports on school characteristics and job satisfaction. Table 3 presents the set of indicators included in the study in detail. In addition, the association with the existing measure of instructional leadership is examined.

The overall aim of the paper is to provide arguments for and to showcase why teacher reports should be considered as a primary source for measures of leadership practices at schools. Given that the current teacher questionnaire was not developed with this purpose, we do not expect to establish perfect measures for use in analyses of data from existing studies. Instead, the paper should be seen as proof-of-concept to be considered for future iterations of TALIS or other international large-scale studies aiming at developing measures of school leadership.

To do so, the study aims to answer four research questions (RQs). Taken together, these RQs and the associated expected outcomes, represent our framework for setting up a validation argument for the new measure of instructional leadership:

- (1) To what degree are principal perception of instructional leadership, as measured by OECD in TALIS 2013, associated with teacher job satisfaction with current work environment?
- (2) What are the measurement properties of the two newly proposed dimensions of instructional leadership based on teacher reports about school features?
- (3) To what degree are the new measures of instructional leadership, developed from the shared perspectives of teachers, associated with teacher job satisfaction?
- (4) To what degree are the new measures of instructional leadership, developed from the shared perspectives of teachers, associated with the instructional leadership as measured by TALIS?

## 4. Methods

### 4.1. Participants

The study performed secondary data analysis of the TALIS 2013 data. TALIS is conducted every five years, beginning in 2008. The target population included lower secondary education (ISCED<sup>2</sup> level 2) teachers and leaders in mainstream schools (OECD, 2014). Data from Norway, Sweden, Denmark, and Finland were used, forming a total sample of 10688 teachers clustered in 676 schools. Table 1 shows the sample sizes across participating countries. Detailed sampling procedures can be found in the TALIS 2013 technical report (OECD, 2014). It may be noted that the total sample sizes of schools were relatively smaller, with a substantially smaller average cluster size in Denmark.

TALIS 2013 is based on a two-stage probability sample design (OECD, 2014). To account for unequal selection probability, sampling weights were used in all analyses. In accordance with Rutkowski, Gonzalez, Joncas, and von Davier's (2010) recommendation and the TALIS user guide (OECD, 2013b) on the usage of sampling weights in multi-level analyses, the final school weight was used at the cluster level. Pure teacher weight, obtained by dividing the final teacher weight with the final school weight, was used at the teacher level.

### 4.2. Measures

#### 4.2.1. Instructional leadership as measured by OECD in TALIS

The instructional leadership scale, like all other scales in TALIS, was built on a confirmatory factor analysis (CFA) framework and the constructs of interest are treated as latent variables (Brown, 2015). As

already discussed, only three items (PQ2–PQ4) out of the five initially intended, were finally used by OECD to build TALIS' instructional leadership scale (see Table 2). The items PQ1 and PQ5 were excluded from the scale due to weak factor loadings (OECD, 2014). Although not uncommon, from a statistical point of view, using only three items to build a scale causes problems with model identification and model fit cannot be evaluated. Another obvious problem with the items from Table 2 is that these are principals' self-evaluations and therefore subject to social desirability bias.

#### 4.2.2. New proposed measures of instructional leadership dimensions from the teachers perspective

Accordingly, we proposed items shown in Table 3 (TQ1–TQ10) as indicators of two dimensions of the Hallinger–Murphy instructional leadership model at the school level.

Items TQ1–TQ6 asked teachers about persons who used certain methods to provide them with feedback about features of their instructional practices. The range of persons listed also included actors external to the school as well as teachers who were not part of the school management team. We recoded the responses to capture feedback practices provided by a person within the school with a defined leadership function (school principals, members of school management team, or an assigned mentor—coded as 1) as an indicator of school leadership, while feedback by external actors and other teachers was coded as 0. The transformed TQ1–TQ6 items were hypothesized to load to a unidimensional latent construct called *managing the instructional program*.

Items TQ7–TQ10 asked teachers to rate their agreement level with statements about the school they worked at. These four items from the teacher level were hypothesized to load to a latent variable called *developing the school learning climate* at the school level.

#### 4.2.3. Teacher job satisfaction with current work environment

Items TQ11–TQ14, presented in Table 3, measure the satisfaction with current work environment construct as suggested by OECD. Since this measure refers to work at a particular school, item TQ4 (which could be understood as a more general evaluation of job satisfaction) is excluded from this scale in our study. The item is also problematic for use at the school level because it does not meet requirements regarding item wording when shared cluster constructs are studied (Stapleton et al., 2016). Furthermore, the notion that this item does not target the current school environment was confirmed in an initial empirical investigation, demonstrating that it had close to zero variance between schools across all included countries. This observation is in line with the work of Zakariya (2020), revealing that this item caused problems in modelling the TALIS job satisfaction scale. Hence, we proposed using a scale for teacher satisfaction with current work environment that consists of three items only. The intention here was not to study the measurement properties of this scale in isolation, but to use this measure as an external criterion for validating two separate instructional leadership measures; thus, the issue with model identification is not critical in this context.

### 4.3. Statistical analysis

Data were first prepared using IDB Analyzer and IBM SPSS 25. Further analyses were done with Mplus version 8.1 (Muthén & Muthén, 1998-2017/Muthén & Muthén, 1998-2017). In the first step, a number of multilevel confirmatory factor analyses (MCFA) were modelled to evaluate each construct separately by country. Subsequently, targeted multilevel structural equation models (MSEM) were conducted (Kline, 2015). These analytical approaches have been developed for analyzing clustered data, where variance at the individual teacher level (within schools) and at the school level (between schools) is properly handled.

In line with Stapleton et al.'s (2016) recommendations, we modelled the aforementioned measures from the teacher questionnaire as shared

<sup>2</sup> International Standard Classification of Education (ISCED, 1997).

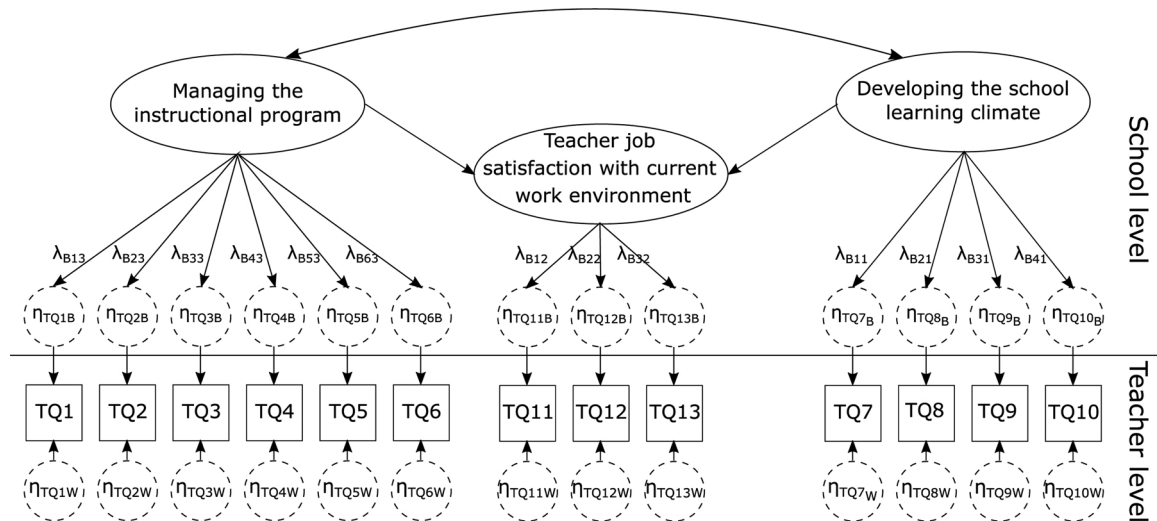


Fig. 1. Hypothesized model of instructional leadership dimensions—*managing the instructional program* and *developing the school learning climate*—and their association with school-level *teacher job satisfaction with current work environment*.

Note 1. At the teacher level, all items correlate. For the sake of simplicity, correlations are not displayed.  
 Note 2. Residuals for the latent variables at the between level are also not displayed.

Table 1  
Sample- and Cluster Sizes.

	Denmark	Finland	Sweden	Norway
Number of teachers	1649	2739	3319	2981
Number of schools	148	197	186	145
Cluster size	10.79	18.44	17.15	19.50

Table 2  
Items From the TALIS 2013 Principal Questionnaire (PQ) Used to Measure Instructional Leadership.

Item	Item wording	Original TALIS code
PQ1	I observed instruction in the classroom.	TC2G21B
PQ2	I took actions to support co-operation among teachers to develop new teaching practices.	TC2G21C
PQ3	I took actions to ensure that teachers take responsibility for improving their teaching practices.	TC2G21D
PQ4	I took actions to ensure that teachers feel responsible for their students' learning outcomes.	TC2G21E
PQ5	I provided parents and guardians with information on the school and student performance.	TC2G21F

cluster constructs at the school level, with a saturated model of covariances at the teacher level. Intraclass correlation 1 (ICC1), as a measure of clustering, and intraclass correlation 2 (ICC2), as a measure of reliability at the cluster level, should be high enough to be considered as evidence that items show acceptable and sufficient degree of clustering (Bliese, 2000). Weighed least squares means and variance adjusted (WLSMV) estimator was used because categorical data with less than five response categories were analyzed (Brown, 2015; Rhemtulla, Brosseau-Liard, & Savalei, 2012). The amount of missing data in this study was not substantial. By default, Mplus with WLSMV does not include cases with missing data on all variables.

Usually, a number of fit indices are reported to evaluate the total (within and between) model fit: the chi-square ( $\chi^2$ ) with corresponding degrees of freedom (df) and its significance (p); the root mean square error of approximation (RMSEA) close to .06 or below; the comparative fit index (CFI) close to .95 and greater; the Tucker–Lewis index (TLI) close to .95 and greater; and the standardized root mean square residual (SRMR) close to .08 or below (Hu & Bentler, 1999). For the models

Table 3  
Items From the TALIS 2013 Teacher Questionnaire (TQ) Used in the Study.

Item	Item wording	Original TALIS code
TQ1	In this school, who uses the following methods to provide feedback to you? [External individuals or bodies; School principal; Member(s) of the school management team; Assigned mentors; Other teachers (not part of the management team); I have never received this feedback at this school.] Feedback following direct observation of your classroom teaching.	TT2G28A
TQ2	Feedback from student surveys about your teaching.	TT2G28B
TQ3	Feedback following an assessment of your content knowledge.	TT2G28C
TQ4	Feedback following an analysis of your students' test scores.	TT2G28D
TQ5	Feedback following your self-assessment of your work (e.g., presentation of a portfolio assessment).	TT2G28E
TQ6	Feedback following surveys or discussions with parents or guardians.	TT2G28F
TQ7	How strongly do you agree or disagree with the following statements about this school? In this school, a development or training plan is established for teachers to improve their work as a teacher.	TT2G31D
TQ8	In this school, feedback is provided to teachers based on a thorough assessment of their teaching.	TT2G31E
TQ9	In this school, measures to remedy any weaknesses in teaching are discussed with the teacher.	TT2G31G
TQ10	In this school, a mentor is appointed to help the teacher improve his/her teaching.	TT2G31H
TQ11	[Finally] We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements? I would like to change to another school if that were possible.	TT2G46C
TQ12	I would recommend my school as a good place to work.	TT2G46G
TQ13	I enjoy working at this school.	TT2G46E
TQ14	All in all, I am satisfied with my job.	TT2G46J

presented here, the overall model fit was largely dominated by the individual level (Ryu, 2014). Given that our models were fully saturated for the within part, the total model fit would be uninformative; hence, we relied only on the SRMR for the between level (SRMR<sub>b</sub>). The SRMR<sub>b</sub> can detect misspecification at the cluster level reasonably well (Kim, Dedrick, Cao, & Ferron, 2016; Ryu, 2014).

It should be noted that these rule of thumb criteria, although frequently used, are quite arbitrary and should not be followed blindly.

Fit indices can be affected by numerous factors, such as small-sample bias, effects of violation of normality and independence, estimation method, model complexity, etc. Sample-size bias especially arises in multilevel models, where samples at the between level are smaller than ideally desired. Accordingly, the cut-off criteria of .08 for SRMR at the between level is generally too strict (Asparouhov & Muthen, 2018). Given these sample characteristics and the complexity of models estimated in this study, this criterion was relaxed. The study was conducted using the following steps:

- (1) Descriptive statistics at the item level, ICC1 and ICC2 were analyzed to test appropriateness for multilevel modelling.
- (2) Association between the current instructional leadership scale available in TALIS and the satisfaction with current work environment was analyzed using MSEM.
- (3) New constructs were proposed and tested using the MCFA, country-by-country: *developing the school learning climate, managing the instructional program*, and teacher job satisfaction with current work environment.
- (4) *Developing the school learning climate* and *managing the instructional program* were investigated as a two-factor measurement model of instructional leadership.
- (5) Bivariate latent correlations between the two newly proposed measures and teacher job satisfaction with current work environment were estimated separately. Similarly, correlations between these two measures and the existing measure of instructional leadership were estimated.
- (6) The final model (Fig. 1), with both dimensions of instructional leadership (*managing the instructional program* and *developing school learning climate*) as predictors of teacher job satisfaction with current work environment, was carried out.

## 5. Results

### 5.1. Descriptive statistics

For most items, ICC1s and ICC2s<sup>3</sup> are acceptable according to common recommendations (Geldhof, Preacher, & Zyphur, 2014; Klein, S.W., J., & Kozlowski, 2001), suggesting that multilevel modelling is meaningful. Only the TQ14 item from the teacher job satisfaction scale shows a low measure of clustering (ICC1 = .01–.05) in Finland and Sweden and, consequently, low reliability at the school level (ICC2 = .02–.44). As already stated, this item was—for this and other substantive reasons—omitted from further analyses.

Teachers in Denmark, Finland, and Sweden, reported lower levels of agreement for most items on the Managing the instructional leadership scale, while Norwegian teachers largely expressed that these forms of feedback occurred in their schools.

### 5.2. Association between instructional leadership and teacher job satisfaction with current work environment

Instructional leadership was first modelled and analyzed according to the measure used by OECD—as a unidimensional scale based on the responses of principals to questions PQ2–PQ4 (see Table 2). The teacher job satisfaction scale originated from responses in the teacher questionnaire. Table 4 presents the outcome of the MSEM analysis, where Teacher job satisfaction with current work environment was regressed on Principals’ instructional leadership. Across the four countries, the model had a good or acceptable model fit (SRMR<sub>b</sub> = .027–.044). However, within each country, the model explains less than 2 % of the variance in teacher job satisfaction at the school level, with R<sup>2</sup> not statistically different from zero. This result is not consistent with most

**Table 4**

MSEM Regression Model Estimates Between Instructional Leadership and Teacher Job Satisfaction With Current Work Environment at the School Level.

	Denmark	Finland	Sweden	Norway
$\chi^2$	4.774	9.084	8.248	6.325
df	8	8	8	8
SRMR <sub>b</sub>	0.027	0.044	0.027	0.029
$\beta$ (S.E.)	−0.04 (0.12)	−0.03 (0.11)	−0.10 (0.09)	−0.00 (0.11)
R <sup>2</sup>	0.002	0.001	0.010	0.000

\*\*\*p < .001; \*\*p < .01; \*p < .05.

Note. The regression coefficients are standardized.

previous research, where instructional leadership is found to be an important predictor of teacher job satisfaction (Burkhauser, 2017; Dou et al., 2017; Ilgan, Parylo, & Sungu, 2015; Johnson et al., 2011). One explanation might, of course, be that there is no actual association between principals’ perception of instructional leadership and teacher job satisfaction with current work environment in the TALIS 2013 data for Nordic countries. An alternative explanation—motivating this study—is that this lack of significant association is caused by severe construct underrepresentation (and other methodological limitations) in the existing measure of leadership.

### 5.3. Building a new measure of instructional leadership using teacher data

As an alternative, this study proposes a new way for measuring instructional leadership using items from the teacher questionnaire. This section presents the analyses conducted to establish this new measure. The first step was to conduct separate confirmatory factor analyses for each of the two proposed sub-dimensions of instructional leadership, modelled as a shared construct.

For the *managing the instructional program dimension*, the models in Sweden, and Norway are evaluated as acceptable, having an SRMR<sub>b</sub> of .080 and .046, respectively (see Table 5). The model fits in Denmark and Finland are somewhat higher but still having a SRMR<sub>b</sub> with an approximate fit. Significantly higher residuals are observed for item TQ4 in all countries—with particular low factor loadings in Denmark and Finland. This item refers to feedback following the analysis of student test scores. It is likely that the item reflects features of assessment policies that differ across countries because 1) grading policies substantially vary across countries (Klette, 2002), and 2) policy frameworks for teacher appraisal and feedback substantially vary across countries (OECD, 2013a). In some countries (e.g., Finland, Denmark, and

**Table 5**

MCFA of the Latent Construct *Managing the Instructional Program* at the School Level.

	Denmark	Finland	Sweden	Norway
$\chi^2$	18.188*	10.567	33.048***	18.883*
df	9	9	9	9
SRMR <sub>b</sub>	0.094	0.096	0.080	0.046
<b>Managing the instructional program</b>				
TQ1	0.480 (0.109)	0.833 (0.082)	0.648 (0.073)	0.717 (0.058)
TQ2	0.902 (0.107)	0.801 (0.085)	0.764 (0.060)	0.850 (0.040)
TQ3	0.927 (0.096)	0.928 (0.101)	0.891 (0.109)	0.945 (0.040)
TQ4	0.554 (0.126)	0.506 (0.271)	0.690 (0.085)	0.701 (0.057)
TQ5	0.822 (0.186)	0.740 (0.099)	0.691 (0.099)	0.980 (0.033)
TQ6	0.709 (0.135)	0.854 (0.113)	0.756 (0.059)	0.953 (0.037)
<b>Residuals</b>				
TQ1	0.769 (0.105)	0.306 (0.137)	0.580 (0.094)	0.486 (0.084)
TQ2	0.156 (0.193)	0.358 (0.136)	0.416 (0.092)	0.287 (0.068)
TQ3	0.142 (0.178)	0.139 (0.187)	0.206 (0.195)	0.107 (0.075)
TQ4	0.693 (0.140)	0.744 (0.274)	0.524 (0.117)	0.508 (0.080)
TQ5	0.325 (0.305)	0.452 (0.147)	0.522 (0.137)	0.040 (0.065)
TQ6	0.497 (0.191)	0.270 (0.194)	0.428 (0.090)	0.092 (0.070)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

Note 1. Table shows standardized factor loadings and residuals with standard errors

<sup>3</sup> For more, see Appendix A.

Norway), a policy for teacher appraisal is not formally established, while the policies differ in other countries where they do exist (e.g. probation period as a form of appraisal in Sweden) (OECD, 2013c). Furthermore, the use of student test results for teacher appraisal is less common or does not exist in some countries (e.g., Finland, Norway, and Denmark) (OECD, 2013c). In addition, item TQ1, which refers to classroom teaching observations, has low factor loadings in Denmark in comparison to other countries. The results suggest that there are differences in how the instructional program is managed across the Nordic group of countries, especially with respect to classroom observations and analyses of student test scores. Those practices might also be indicators of non-observed phenomena at the school level, not captured by this dimension of instructional leadership.

For the *developing the school learning climate* dimension, all countries demonstrated an acceptable model fit, with an SRMR<sub>b</sub> of .046 in Denmark, .023 in Finland, .027 in Sweden, and .018 in Norway (see Table 6). In conclusion, the measurement model provides evidence for the claim that the *developing the school learning climate* dimension, as reported by teachers, captures a potentially useful measure of school characteristic in all investigated countries. In other words, developed training plan for teachers, assessment of teaching followed by feedback, open discussions about weaknesses in teaching, and mentoring are important facets of school climate that nurturing professional development in schools across all countries.

5.4. Measure of teacher job satisfaction with current work environment

Items TQ11, TQ12, and TQ13 from the teacher questionnaire were used to measure the unidimensional latent factor *teacher job satisfaction with current work environment* at the school level. As explained above, this study excluded one of the items included in the official OECD measure. In doing so, other issues arose such as the model being just identified. To resolve this issue, the residual variance for item TQ13 was fixed to a very small value (0.01) at the between level (Brown, 2015). In line with previously reported analyses, a model fully saturated at the teacher level was estimated with factor structure estimated at the school level only. The model fit indices reveal a good model fits across all countries. The respective SRMR<sub>b</sub> values obtained are .000 in Denmark, .004 in Finland, .021 in Sweden, and .025 in Norway.<sup>4</sup> Therefore, a good fitting model of teacher job satisfaction with current work environment is established in all countries suggesting that desire to teach, enjoyment connected with it, and the feeling of being in a good place (all related to

**Table 6**  
MCFA of the Latent Construct *Developing the School Learning Climate* at the School Level.

	Denmark	Finland	Sweden	Norway
$\chi^2$	2.612	2.627	3.733	2.739
df	3	2	2	2
SRMR <sub>b</sub>	0.046	0.023	0.027	0.018
<b>Developing the school learning climate by</b>				
TQ7	0.683 (0.164)	0.813 (0.059)	0.640 (0.090)	0.680 (0.062)
TQ8	0.691 (0.114)	0.799 (0.082)	0.764 (0.074)	0.968 (0.034)
TQ9	0.956 (0.012)	0.932 (0.074)	0.866 (0.076)	0.902 (0.048)
TQ10	0.640 (0.136)	0.678 (0.066)	0.779 (0.069)	0.712 (0.059)
<b>Residual variances</b>				
TQ7	0.534 (0.223)	0.339 (0.096)	0.591 (0.115)	0.537 (0.084)
TQ8	0.523 (0.157)	0.393 (0.128)	0.416 (0.113)	0.064 (0.066)
TQ9	0.086 (0.023)	0.132 (0.137)	0.250 (0.131)	0.186 (0.086)
TQ10	0.590 (0.174)	0.541 (0.089)	0.393 (0.108)	0.493 (0.084)

\*\*\*p < .001; \*\*p < .01; \*p < .05. Note 1. Table shows standardized factor loadings and residuals with standard errors

a particular school) are important indicators of teachers' shared perception of being satisfied with working environment

5.5. Association between teachers' and principals' perceptions of instructional leadership

As a validation step for the newly established dimensions of instructional leadership from the perspective of teachers, we examined their correlation with the instructional leadership measure from the perspective of principals, as proposed by TALIS. The results are shown in Tables 7 and 8. Both dimensions are moderately positively correlated with instructional leadership in Norway and Sweden. *Developing the school learning climate* dimension is also moderately positively correlated to instructional leadership in Denmark, while the relationship is not significant in Finland. Furthermore, the relationship between Managing the instructional program as reported by teachers is not statistically significant with the principals' report on instructional leadership. The model that includes the *developing the school learning climate* dimension has the best overall fit—with either acceptable or approximately acceptable fit in all countries—while the fit for the model with the *managing the instructional program* dimension is more modest.

Very high correlations were not expected given that the measure based on principals' and teachers reports, respectively, captures different aspect of instructional leadership, and moreover, since the measure derived from the principals' responses are prone to be biased due to self-reporting on their own actions. Taken together, the mostly moderate and substantial relationship between the two teacher-based measures and the principal's report is consistent with our hypothesis that the measure developed from the teacher questionnaire captures facets of instructional leadership. On the other hand, the fact that the relationships between the two measures are relatively modest, and even non-significant for both models in one country, also suggests that principals and teachers largely disagree in their perceptions of how instructional leadership is executed in their schools.

5.6. Association between the new measures of Instructional leadership and Teacher Job Satisfaction With Current Work Environment

5.6.1. Managing the instructional program and teacher job satisfaction with current work environment

To answer RQ 3, teacher job satisfaction with current work environment is regressed on managing the instructional program at the school level for all four countries. Table 9 shows model fit evaluation and corresponding statistics. Acceptable model fit is only observed in Norway but, given the complexity of the model and the characteristics of samples, Denmark and Finland, are regarded as having an approximate model fit. Accordingly, correlations should be interpreted with caution. In line with theoretical expectations, the correlation is substantial and statistically significant in Finland and Norway,  $\beta = .474$  and  $.415$ , respectively. For these countries, this implies that when instruction is managed through feedback by principals or other assigned persons at the level of the school, teachers are on average more satisfied. Greater satisfaction can be achieved by providing teachers with feedback after classroom observations, assessments of their content knowledge,

**Table 7**  
SEM Regression Model Estimates Between “Managing the Instructional Program” and “Instructional Leadership”.

	Denmark	Finland	Sweden	Norway
$\chi^2$	42.155*	16.072	45.699**	81.036***
df	26	26	26	27
SRMR <sub>b</sub>	0.114	0.088	0.085	0.093
corr (S.E.)	0.094 (0.163)	0.109 (0.124)	0.306** (0.107)	0.301*** (0.073)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

<sup>4</sup> See Appendix B.

**Table 8**  
SEM Regression Model Estimates Between “Developing the School Learning Climate” and “Instructional Leadership”.

	Denmark	Finland	Sweden	Norway
$\chi^2$	19.208	14.088	20.165	49.208***
df	14	15	13	13
SRMRb	0.088	0.054	0.052	0.081
corr (S.E.)	0.297* (0.144)	0.006 (0.118)	0.329** (0.106)	0.377*** (0.086)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

**Table 9**  
SEM Regression Model Estimates Between “Teacher Job Satisfaction With Current Work Environment” and “Managing the Instructional Program” Dimension of Instructional Leadership at the School Level.

	Denmark	Finland	Sweden	Norway
$\chi^2$	34.900	26.01	64.349***	33.685
df	28	27	29	27
SRMRb	0.107	0.097	0.103	0.055
$\beta$ (S.E.)	0.181 (0.152)	0.474*** (0.133)	0.167 (0.111)	0.415*** (0.096)
R <sup>2</sup> (S.E.)	0.033 (0.055)	0.225* (0.107)	0.028 (0.037)	0.172* (0.079)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

analysis of student test scores, etc. In contrast, although still positive, no statistically significant correlations are found in Sweden and Denmark. This may suggest that more aspects, than who provides the feedback, should be considered. The format, frequency, quality and consequences of the feedback on teachers’ instruction is for instance not directly captured by this measure.

5.6.2. *Developing the school learning climate and teacher job satisfaction with current work environment*

The relationship between the instructional leadership dimension of developing the school learning climate and teachers job satisfaction is modelled in the same fashion. According to the provided model fit indices in Table 10, acceptable model fits are found in all countries, except for Denmark. However, the model fit in Denmark is only marginally higher than the rule of thumb. Statistically significant and substantially meaningful correlations are found in all countries. Teachers feel satisfied when working in schools where they receive support in teaching, whether by having a mentor to help them improve their teaching or other person to discuss potential weaknesses in teaching. Moreover, teachers feel satisfied when working in schools where they receive support in terms of professional development and training. This is in itself not surprising, when considering this phenomenon from an individual teacher’ perspectives. However, this analysis also reveals that this is a systematic relationship characterizing the larger community of teachers within schools.

**Table 10**  
SEM Regression Model Estimates Between “Teacher Job Satisfaction With Current Work Environment” and “Developing the School Learning Climate” Dimension of Instructional Leadership at the School Level.

	Denmark	Finland	Sweden	Norway
$\chi^2$	26.738*	18.817	19.234	22.661
df	15	15	15	15
SRMRb	0.093	0.044	0.045	0.057
$\beta$ (S.E.)	0.553*** (0.128)	0.517*** (0.084)	0.505*** (0.095)	0.609*** (0.072)
R <sup>2</sup> (S.E.)	0.306** (0.142)	0.267** (0.087)	0.255** (0.096)	0.370*** (0.087)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

5.6.3. *Full model of instructional leadership and teacher job satisfaction with current work environment*

In the first step for modelling the full model proposed in Fig. 1, the correlation between the two proposed dimensions of instructional leadership was investigated. This step revealed that the managing the instructional program and the developing the school learning climate dimensions are highly correlated, with correlations higher than .70 in all countries.<sup>5</sup> The model fit is found acceptable in Norway, with approximate fit in Finland and Sweden.

High correlation between these two dimensions of instructional leadership is a signal that the final model, where both dimensions are included as predictors of teacher job satisfaction with current work environment, would run into problems associated with multicollinearity. Accordingly, attempts to run this model resulted in unstable estimation with rather poor model fits and large standard errors. This problem is discussed by Marsh, Dowson, Pietsch, and Walker (2004). To deal with this issue, they demonstrate that constraining the paths from two latent predictors to be equal leads to a more parsimonious fit to the data, reducing standard errors in the path coefficients. Accordingly, we adopted this approach and the solution is reported in Table 11. As for all previous models, the fit is found to be good in Norway. However, the model does not fit very well in any of the other countries, with only Finland and Sweden approximating an acceptable fit.

6. Discussion and conclusion

The main issue addressed in this study is the measurement of instructional leadership as currently implemented in the TALIS 2013 survey. We argued that, even though TALIS does an important job in providing information about educational systems around the world, the concept of instructional leadership is not adequately covered in its instruments. The main purpose of this study was to provide researchers and those interested in leadership with a possible alternative approach to assess, study, and interpret leadership at schools. Hence, this paper should be regarded as being proof-of-concept for motivating and supporting future studies in which measures of school leadership are included. Specifically, we argued that the instructional leadership measure in TALIS is hampered by construct underrepresentation when assessed through self-reports of principals on three items only. Furthermore, we suggested that the measurement of what we perceive to be a school characteristic is not well represented through reports by a single school entity. A multilevel factor structure for the two newly proposed measures was examined and, taking further steps to support

**Table 11**  
SEM Regression Model Estimates Between “Managing the Instructional Program” and “Teacher Job Satisfaction With Current Work Environment”—( $\beta_1$ )—and Between “Developing the School Learning Climate” and “Teacher Job Satisfaction With Current Work Environment”—( $\beta_2$ ).

	Denmark	Finland	Sweden	Norway
$\chi^2$	103.615**	79.091*	131.738***	107.087***
df	65	64	65	65
SRMRb	0.150	0.105	0.108	0.075
Corr (S.E.)	0.971*** (0.130)	0.700*** (0.121)	0.846*** (0.063)	0.918*** (0.041)
$\beta_1$ (S.E.)	0.214* (0.083)	0.242*** (0.064)	0.199*** (0.054)	0.255*** (0.046)
$\beta_2$ (S.E.)	0.109* (0.047)	0.335*** (0.064)	0.162** (0.056)	0.256*** (0.046)
R <sup>2</sup> (S.E.)	0.103 (0.078)	0.285** (0.099)	0.121 (0.065)	0.250** (0.080)

\*\*\*p < .001; \*\*p < .01; \*p < .05.

<sup>5</sup> See Appendix C



their validity, the associations with the existing measure were estimated. Furthermore, the theory proposes that instructional leadership correlates with teacher job satisfaction (Burkhauser, 2017; Sims, 2019). However, no such association was found in the TALIS 2013 data between the existing instructional leadership measure and teacher job satisfaction with the current work environment. Accordingly, as a final validation step for the new measures of instructional leadership proposed, we used a scale representing teacher satisfaction with their current work environment as an external criterion.

As an alternative, we proposed an approach in which items from the teacher questionnaire—modelled at the school level—are used as indicators of instructional leadership in TALIS. This approach is inspired by and parallel to how instructional quality in classrooms is increasingly based on student reports instead of relying on single teacher reports (Wagner et al., 2016). The newly proposed measures have at least three promising features: (a) they represent a joint collective evaluation of practices at the school level; (b) they can be associated with specific dimensions and functions included in the Hallinger–Murphy instructional leadership model; and (c) their indicators cover a wider representation of this model when taken together. To be more specific regarding the latter, the measures included in the new approach represent functions that are part of two out of three dimensions of the Hallinger–Murphy model (*managing the instructional program* and *developing the school learning climate*). Accepting that measurement in the international context is complex, we included data from the Nordic countries participating in the TALIS 2013 survey. We demonstrated that the two newly proposed measures of instructional leadership functioned reasonably well in most countries. *Developing the school learning climate* had a superior fit across countries in comparison to *managing the instructional program*. The new measures were found to be moderately positively correlated with the existing TALIS measure based on instructional leadership reports from principals. Furthermore, we were able to demonstrate that the newly proposed measures were also positively related to an external criterion (teacher satisfaction with their current work environment). Given that these items were not included in the study with the intention to build indicators of leadership at the school level, we find these results to be promising—even if the structural models did not satisfy the frequently used rules of thumb for evaluating fit in all countries.

Of the two suggested dimensions reflecting instructional leadership, the *managing the instructional program* measure was the least successful in terms of model fit. There are at least two possible underlying causes that are consistent with these observations: 1) ambiguities introduced by the item format and 2) culturally situated interpretation of the core “feedback” concept involved in this set of items. Regarding the item format, there are two sources of information: 1) item contexts (e.g., analysis of student test scores or assessment of teacher content knowledge) and 2) information regarding the instruction (whether feedback has been given). In finding a way to respond affirmatively to these items, a phenomenon defined by specific actions must first be evaluated as being present and then, given this, a teacher has to evaluate that feedback is typically provided. This creates ambiguity that could result in different interpretations of what the item is actually asking for. Factor loadings for some specific items on this scale were rather low in some countries, possibly indicating that, even if this function of instructional leadership is executed at schools, specific practices may differ across countries in accordance with accepted norms. For example, classroom observations are widely used as an instrument for appraisal and feedback in many countries. However, in countries where teacher appraisal is more informal and not regulated by law (like Norway, Finland, and Denmark), classroom observations are not something that occurs regularly or systematically. Instead, in these countries, where teachers have a high degree of autonomy, the main form of feedback is often in the shape of less formalized dialogues between colleagues (Nusche, Earl, Maxwell, & Shewbridge, 2011; Shewbridge, Jang, Matthews, & Santiago, 2011). This was confirmed also in previous research where feedback from

school leaders in Nordic countries has shown to be lower than the international average, further suggesting less hierarchical structure in Nordic schools (Ludvigsen, 2016). However, the amount of feedback differs not only across countries but also within countries, and across persons involved in giving feedback, across practices after which feedback is given, and across novice and experienced teachers (Ludvigsen, 2016).

Further to this, a final aim would be to develop new measures that would work across a global context. This leads us into another major issue that possibly leads to invariant properties of this measure—the culturally situated perception of the term “feedback,” which is involved in the question stem. This may not be an easy concept to translate or adapt to different languages/cultures. Feedback is a complex phenomenon that involves not only the act of someone observing and providing constructive reflections but that also manifests a structural relation between the persons who give or receive the feedback, respectively. This expresses a power relationship or a view on authority which is culturally specific (Hofstede, 1984; Inglehart & Welzel, 2005). With this background, it is reasonable to suggest that feedback is seen as a support function in some contexts, while it could be regarded more as a control function in others.

In this paper, the association with teacher job satisfaction with current work environment was examined as a validation step for the newly proposed measures of instructional leadership. The association with this external criterion was particularly strong and stable across countries for the *developing the school learning climate* dimension, while this relationship was weaker and less stable, overall, across countries for the *managing the instructional program* dimension—although still positive and significant in Finland and Norway. In addition, this study demonstrated that the two dimensions of instructional leadership, based on modelling the between-school variation of teacher responses, are highly correlated. This means that schools that score highly on *managing the instructional program* also tend to promote a climate beneficial for learning, as theoretically expected. However, beyond the fact that the dimensions are highly correlated, the current data and design do not allow for a more specific examination of the internal structure of the concept of instructional leadership.

A major limitation of this study is that measures of instructional leadership were developed in a post-hoc fashion from items that were not originally intended to be used for this purpose. Arguments are provided as to why these items are still reasonable indicators of instructional leadership at the school level—the statements reflect school-level phenomena evaluation and represent reasonable reflections of the core concepts found in the Hallinger–Murphy framework. Although we conclude that the measurement and structural models presented provide promising results, the models are far from perfect. Specifically, the *managing the instructional program* dimension did not demonstrate ideal measurement properties in all the countries. With the complexity of the multilevel structural models analyzed in this study, the data are not ideal given the average small cluster size (Asparouhov & Muthen, 2018). Moreover, for pragmatic reasons, we analyzed data in a small group of more homogeneous countries. Further work is needed for validating the proposed measures in a wider international context, in particular with a focus on the analysis of measurement invariance.

Self-reports by principals or other school leaders may reasonably be suspected of bias due to social desirability, personality traits, or other construct irrelevant features. The TALIS 2018 study made some improvements regarding the emphasis on distributed and a collective component of leadership where both, principal and teacher perceptions are available. Given the presented results, we suggest that even further developmental work is essential in order to measure instructional (or other types) of leadership by using teachers’ collective observations instead of relying on one principal’s self-report. Teachers should be asked about specific and observable actions embedded in their school settings. Then, teachers are in a position to provide indicators for measures of “leadership in action.” It goes without saying that the

specific set of actions included in the questions should tightly be linked to the underlying theoretical concept of leadership that is of relevance for the specific research at hand. Although the present study does not provide a complete roadmap to how this may be done, it does provide a case—or proof-of-concept—that such an alternative approach to measuring school leadership is a promising avenue, deserving attention in future developments of large-scale education studies.

**Declaration of Competing Interest**

The authors report no declarations of interest.

**Appendix A**

*Descriptive statistics*

Interclass correlation 1 (ICC1) and interclass correlation 2 (ICC2) as measures of clustering and reliability at the cluster level, respectively.

Item	ICC1 Denmark	ICC2	ICC1 Finland	ICC2	ICC1 Sweden	ICC2	ICC1 Norway	ICC2
<b>Managing the instructional program</b>								
TQ1	0.35	0.85	0.17	0.79	0.36	0.90	0.47	0.95
TQ2	0.09	0.53	0.12	0.71	0.30	0.88	0.20	0.83
TQ3	0.16	0.67	0.07	0.60	0.07	0.54	0.16	0.79
TQ4	0.19	0.71	0.03	0.36	0.11	0.68	0.20	0.83
TQ5	0.06	0.40	0.12	0.71	0.10	0.65	0.18	0.81
TQ6	0.12	0.60	0.05	0.51	0.12	0.69	0.12	0.73
TQ7	0.04	0.31	0.12	0.71	0.08	0.57	0.18	0.81
TQ8	0.06	0.40	0.05	0.49	0.11	0.67	0.20	0.82
TQ9	0.10	0.54	0.09	0.64	0.08	0.60	0.15	0.77
TQ10	0.09	0.49	0.12	0.71	0.09	0.62	0.26	0.87
TQ11	0.12	0.60	0.10	0.66	0.14	0.66	0.10	0.67
TQ12	0.27	0.80	0.19	0.81	0.20	0.81	0.19	0.81
TQ13	0.14	0.63	0.06	0.53	0.10	0.43	0.09	0.66
TQ14	0.12	0.58	0.01	0.18	0.04	0.72	0.10	0.68

**Appendix B**

MCFA of the Latent Construct “Teacher Job Satisfaction With Current Work Environment” and Errors

	Denmark	Finland	Sweden	Norway
$\chi^2$	0.000	0.036	8.912**	7.276**
df	1	1	1	1
SRMRb	0.000	0.004	0.021	0.025
<b>Teacher Job Satisfaction with current work environment</b>				
TQ13	0.967 (0.039)	0.913 (0.023)	0.957 (0.008)	0.953 (0.012)
TQ12	0.988 (0.038)	0.991 (0.048)	0.991 (0.019)	0.984 (0.027)
TQ11	0.954 (0.044)	0.918 (0.063)	0.936 (0.034)	0.965 (0.041)
<b>Residuals</b>				
TQ13	0.064 (0.076)	0.166 (0.041)	0.085 (0.016)	0.091 (0.022)
TQ12	0.025 (0.074)	0.018 (0.096)	0.017 (0.038)	0.031 (0.053)
TQ11	0.089 (0.083)	0.158 (0.116)	0.123 (0.063)	0.069 (0.078)

\*\*\*p < .001; \*\*p < .01; \*p < .05 Note 1. Table shows standardized factor loadings and residuals with standard errors

**Appendix C**

MCFA Results of a Single Two-Factor Measurement Model of “Managing the Instructional Program” and “Developing the School Learning Climate”

	Denmark	Finland	Sweden	Norway
$\chi^2$	51.338*	52.296*	77.173***	65.974***
df	34	35	34	34
SRMRb	0.116	0.109	0.089	0.057
Corr (S.E.)	0.955*** (0.131)	0.708*** (0.119)	0.844*** (0.063)	0.938*** (0.039)

\*\*\* $p < .001$ ; \*\* $p < .01$ ; \* $p < .05$

## References

- Ahola, S., Hedmo, T., Thomsen, J.-P., & Vabø, A. (2014). *Organisational features of higher education. Denmark, Finland, Norway and Sweden*. NIFU [Working Paper] <https://nifu.brage.unit.no/nifu-xmlui/handle/11250/2358908>.
- Ansley, B. M., Houchins, D., & Varjas, K. (2019). Cultivating positive work contexts that promote teacher job satisfaction and retention in high-need schools. *Journal of Special Education Leadership*, 32(1), 3–16.
- Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., & Bong, C. H. (2014). Predicting survey responses: How and why semantics shape survey statistics on organizational behaviour. *PLoS ONE*, 9(9). <https://doi.org/10.1371/journal.pone.0106361>.
- Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., & Egeland, T. (2018). The failing measurement of attitudes: How semantic determinants of individual survey responses come to replace measures of attitude strength. *Behavior Research Methods*, 50(6), 2345–2365. <https://doi.org/10.3758/s13428-017-0999-y>.
- Asparouhov, T., & Muthen, B. (2018). *SRMR in Mplus*. <http://www.statmodel.com/download/SRMR2.pdf>.
- Banerjee, N., Stearns, E., Moller, S., & Mickelson, R. A. (2017). Teacher job satisfaction and student achievement: The roles of teacher professional community and teacher collaboration in schools. *American Journal of Education*, 123(2), 203–241. <https://doi.org/10.1086/689932>.
- Benoliel, P., Shaked, H., Nadav, N., & Schechter, C. (2019). School principals' systems thinking: Antecedents and consequences. *Journal of Educational Administration*, 57(2), 167–184. <https://doi.org/10.1108/JEA-08-2018-0144>.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability. Implications for data aggregation and analysis. In K. J. Kline, & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- Bogler, R. (2001). The influence of leadership style on teacher job satisfaction. *Educational Administration Quarterly*, 37(5), 662–683. <https://doi.org/10.1177/00131610121969460>.
- Boyce, J., & Bowers, A. J. (2018). Toward an evolving conceptualization of instructional leadership as leadership for learning. *Journal of Educational Administration*, 56(2). <https://doi.org/10.1108/JEA-06-2016-0064>.
- Brewer, C., Okilwa, N., & Duarte, B. (2020). Context and agency in educational leadership: Framework for study. *International Journal of Leadership in Education*, 23(3), 330–354. <https://doi.org/10.1080/13603124.2018.1529824>.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition). New York, NY: The Guilford Press.
- Burkhauser, S. (2017). How much do school principals matter when it comes to teacher working conditions? *Educational Evaluation and Policy Analysis*, 39(1), 126–145. <https://doi.org/10.3102/0162373716668028>.
- Bush, T. (2009). Leadership development and school improvement: Contemporary issues in leadership development. *Educational Review*, 61(4), 375–389. <https://doi.org/10.1080/00131910903403956>.
- Bush, T., & Glover, D. (2014). School leadership models: What do we know? *School Leadership & Management*, 34(5), 553–571. <https://doi.org/10.1080/13632434.2014.928680>.
- Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, 44(6), 473–490. <https://doi.org/10.1016/j.jsp.2006.09.001>.
- Cerit, Y. (2009). The effects of servant leadership behaviours of school principals on teachers' job satisfaction. *Educational Management Administration & Leadership*, 37(5), 600–623. <https://doi.org/10.1177/1741143209339650>.
- Clark, A., Oswald, A., & Warr, P. (1996). Is job satisfaction U-shaped in age? *Journal of Occupational and Organizational Psychology*, 69(1), 57–81. <https://doi.org/10.1111/j.2044-8325.1996.tb00600.x>.
- Çoğaltay, N., Yalçın, M., & Karadağ, E. (2016). Educational leadership and job satisfaction of teachers: A meta-analysis study on the studies published between 2000 and 2016 in Turkey. *Eurasian Journal of Educational Research*, 16(62). <https://doi.org/10.14689/ejer.2016.62.13>.
- Day, C., Gu, Q., & Sammons, P. (2016). The impact of leadership on student outcomes: How successful school leaders use transformational and instructional strategies to make a difference. *Educational Administration Quarterly*, 52(2), 221–258. <https://doi.org/10.1177/0013161X15616863>.
- Dou, D., Devos, G., & Valcke, M. (2017). The relationships between school autonomy gap, principal leadership, teachers' job satisfaction and organizational commitment. *Educational Management Administration & Leadership*, 45(6), 959–977. <https://doi.org/10.1177/1741143216653975>.
- Dutta, V., & Sahney, S. (2016). School leadership and its impact on student achievement. *International Journal of Educational Management*, 30, 941–958. <https://doi.org/10.1108/IJEM-12-2014-0170>.
- Evans, L. (1997). Understanding teacher morale and job satisfaction. *Teaching and Teacher Education*, 13(8), 831–845. [https://doi.org/10.1016/S0742-051X\(97\)00027-9](https://doi.org/10.1016/S0742-051X(97)00027-9).
- Franz Coldren, A., & Spillane, J. P. (2007). Making connections to teaching practice: The role of boundary practices in instructional leadership. *Educational Policy*, 21(2), 369–396. <https://doi.org/10.1177/0895904805284121>.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72. <https://doi.org/10.1037/a0032138>.
- González-Falcón, I., García-Rodríguez, M. P., Gómez-Hurtado, I., & Carrasco-Macías, M. J. (2019). The importance of principal leadership and context for school success: Insights from '(in)visible school'. *School Leadership & Management*, 0(0), 1–18. <https://doi.org/10.1080/13632434.2019.1612355>.
- Gumus, S., Bellibas, M. S., Esen, M., & Gumus, E. (2018). A systematic review of studies on leadership models in educational research from 1980 to 2014. *Educational Management Administration & Leadership*, 46(1), 25–48. <https://doi.org/10.1177/1741143216659296>.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education*, 33(3), 329–352. <https://doi.org/10.1080/0305764032000122005>.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, 4(3), 221–239. <https://doi.org/10.1080/15700760500244793>.
- Hallinger, P. (2010). Developing instructional leadership. In B. Davies, & M. Brundrett (Eds.), *Developing successful leadership* (pp. 61–76). Netherlands: Springer. [https://doi.org/10.1007/978-90-481-9106-2\\_5](https://doi.org/10.1007/978-90-481-9106-2_5).
- Hallinger, P. (2011). Leadership for learning: Lessons from 40 years of empirical research. *Journal of Educational Administration*, 49(2), 125–142. <https://doi.org/10.1108/09578231111116699>.
- Hallinger, P. (2018). Bringing context out of the shadows of leadership. *Educational Management Administration & Leadership*, 46(1), 5–24. <https://doi.org/10.1177/1741143216670652>.
- Hallinger, P. (2019). Instructional leadership 1950-2018: Transformation from an American to a global leadership construct. In T. Bush, L. Bell, & D. Middlewood (Eds.), *Principles of educational leadership and management* (3rd ed., pp. 19–38). London, UK: SAGE Publications Ltd.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2), 157–191. <https://doi.org/10.1080/0924345980090203>.
- Hallinger, P., & Leithwood, K. (1998). Unseen forces: The impact of social culture on school leadership. *Peabody Journal of Education*, 73(2), 126–151. [https://doi.org/10.1207/s15327930pje7302\\_6](https://doi.org/10.1207/s15327930pje7302_6).
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *The Elementary School Journal*, 86(2), 217–247. <https://doi.org/10.1086/461445>.
- Hallinger, P., & Wang, W.-C. (2015). *Assessing instructional leadership with the principal instructional management rating scale*. Springer International Publishing. <https://www.springer.com/gp/book/9783319155326>.
- Hariri, H., Monypenny, R., & Prideaux, M. (2012). Principalship in an Indonesian school context: Can principal decision-making styles significantly predict teacher job satisfaction? *School Leadership & Management*, 32(5), 453–471. <https://doi.org/10.1080/13632434.2012.723617>.
- Heck, R. H., Larsen, T. J., & Marcoulides, G. A. (1990). Instructional leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly*, 26(2), 94–125. <https://doi.org/10.1177/0013161X90026002002>.
- Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*. Beverly Hills: SAGE.
- Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). 5 turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *The Academy of Management Annals*, 2(1), 231–274. <https://doi.org/10.1080/19416520802211552>.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Huber, S. G., & Muijs, D. (2010). School leadership effectiveness: The growing insight in the importance of school leadership for the quality and development of schools and their pupils. In S. Huber (Ed.), *School leadership—International perspectives* (pp. 57–77). Netherlands: Springer. [https://doi.org/10.1007/978-90-481-3501-1\\_4](https://doi.org/10.1007/978-90-481-3501-1_4).
- Ilgan, A., Parylo, O., & Sungu, H. (2015). Predicting teacher job satisfaction based on principals' instructional supervision behaviours: A study of Turkish teachers. *Irish Educational Studies*, 34(1), 69–88. <https://doi.org/10.1080/03323315.2015.1005393>.
- Ingersoll, R. M. (2002). The teacher shortage: A case of wrong diagnosis and wrong prescription. *NAESP Bulletin*, 86(631), 16–31. <https://doi.org/10.1177/019263650208663103>.
- Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511790881>.
- Johnson, S. M., Kraft, M. A., & Papay, J. P. (2011). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, 1–39.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127(3), 376–407. <https://doi.org/10.1037/0033-2909.127.3.376>.
- Kelly, N., Cespedes, M., Clarà, M., & Danaher, P. A. (2019). Early career teachers' intentions to leave the profession: the complex relationships among preservice

- education, early career support, and job satisfaction. *Australian Journal of Teacher Education*, 44(3), 93–113. <https://doi.org/10.14221/ajte.2018v44n3.6>.
- Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. *Multivariate Behavioral Research*, 51(6), 881–898. <https://doi.org/10.1080/00273171.2016.1228042>.
- Klassen, R. M., & Anderson, C. J. K. (2009). How times change: Secondary teachers' job satisfaction and dissatisfaction in 1962 and 2007. *British Educational Research Journal*, 35(5), 745–759. <https://doi.org/10.1080/01411920802688721>.
- Klein, K. J., S.W. J., & Kozlowski, S. W. J. (2001). *Multilevel theory, research, and methods in organizations foundations, extensions, and new directions*. San Francisco: Jossey-Bass.
- Klette, K. (2002). Reform policy and teacher professionalism in four nordic countries. *Journal of Educational Change*, 3(3), 265–282. <https://doi.org/10.1023/A:1021234030580>.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (Fourth edition). New York, NY: Guilford Publications.
- Krüger, M., Witziers, B., & Sleegers, P. (2007). The impact of school leadership on school level factors: Validation of a causal model. *School Effectiveness and School Improvement*, 18(1), 1–20. <https://doi.org/10.1080/09243450600797638>.
- Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: implications for theory and research. *British Educational Research Journal*, 36(5), 807–830. <https://doi.org/10.1080/01411920903165603>.
- Ladd, H. F. (2009). *Teachers' perceptions of their working conditions: How predictive of policy-relevant outcomes?* (Working paper No.33.). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research. <https://doi.org/10.1037/e722072011-001>.
- Leithwood, K., & Duke, D. L. (1998). Mapping the conceptual terrain of leadership: A critical point of departure for cross-cultural studies. *Peabody Journal of Education*, 73(2), 31–50. <https://doi.org/10.1207/s15327930pje7302.2>.
- Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management*, 28(1), 27–42. <https://doi.org/10.1080/13632430701800060>.
- Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Executive summary how leadership influences student learning*. New York, NY: The Wallace Foundation.
- Liebowitz, D. D., & Porter, L. (2019). The effect of principal behaviors on student, teacher, and school outcomes: A systematic review and meta-analysis of the empirical literature. *Review of Educational Research*, 89(5), 785–827. <https://doi.org/10.3102/0034654319866133>.
- Liu, Y., & Werblow, J. (2019). The operation of distributed leadership and the relationship with organizational commitment and job satisfaction of principals and teachers: A multi-level model and meta-analysis using the 2013 TALIS data. *International Journal of Educational Research*, 96, 41–55. <https://doi.org/10.1016/j.ijer.2019.05.005>.
- Liu, Y., Bellibaş, M.Ş., & Gümüş, S. (2020). The effect of instructional leadership and distributed leadership on teacher self-efficacy and job satisfaction: Mediating roles of supportive school culture and teacher collaboration. *Educational Management Administration & Leadership*. <https://doi.org/10.1177/1741143220910438>, 1741143220910438.
- Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, 21(3), 315–336. <https://doi.org/10.1080/09243453.2010.486586>.
- Ludvigsen, S. (2016). *Northern lights on PISA and TALIS*. Denmark: Nordic Council of Ministers.
- Ma, X., & MacMillan, R. B. (1999). Influences of workplace conditions on teachers' job satisfaction. *The Journal of Educational Research*, 93(1), 39–47. <https://doi.org/10.1080/00220679909597627>.
- Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370–397. <https://doi.org/10.1177/0013161X03253412>.
- Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, 96(3), 518–522. <https://doi.org/10.1037/0022-0663.96.3.518>.
- Møller, J. (2009). Approaches to school leadership in Scandinavia. *Journal of Educational Administration and History*, 41(2), 165–177. <https://doi.org/10.1080/00220620902808244>.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's Guide* (eighth edition). Los Angeles, CA: Muthén & Muthén.
- Nusche, D., Earl, L., Maxwell, W., & Shewbridge, C. (2011). *OECD reviews of evaluation and assessment in education*. Norway: OECD Publishing.
- O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' instructional leadership behaviors and student achievement. *NASSP Bulletin*, 89(645), 56–71. <https://doi.org/10.1177/019263650508964505>.
- OECD. (2014). *TALIS 2013 technical report*. Paris: OECD Publishing.
- OECD. (2013a). *OECD review on evaluation and assessment frameworks for improving school outcomes—OECD*. Paris: OECD Publishing. <http://www.oecd.org/education/school/oecdreviewonevaluationandassessmentframeworksforimprovingchooloutcomes.htm>.
- OECD. (2013b). *TALIS user guide for the international database*. Paris: OECD Publishing.
- OECD. (2013c). *Teachers for the 21st century: Using evaluation to improve teaching*. Paris: OECD Publishing.
- Osborne-Lampkin, L., Folsom, J. S., & Herrington, C. D. (2015). *A systematic review of the relationships between principal characteristics and student achievement*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.
- Pont, B., Nusche, D., & Moorman, H. (2008). *Improving school leadership, policy and practice* (Volume 1). Paris: OECD.
- Qadach, M., Schechter, C., & Da'sa, R. (2020). Instructional leadership and teachers' intent to leave: The mediating role of collective teacher efficacy and shared vision. *Educational Management Administration & Leadership*, 48(4), 617–634.
- Qin, L. (2019). Factors relating to teachers' intention to change school: A multilevel perspective. *Policy Futures in Education*, 17(3), 318–338. <https://doi.org/10.1177/1478210318822184>.
- Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>.
- Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44(5), 635–674. <https://doi.org/10.1177/0013161X08321509>.
- Rutkowski, D., Rutkowski, L., Bélanger, J., Knoll, S., Weatherby, K., & Prusinski, E. (2013). *The Talis 2013 conceptual framework*. Paris: OECD Publishing.
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142–151. <https://doi.org/10.3102/0013189X10363170>.
- Ryu, E. (2014). Model fit evaluation in multilevel structural equation models. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00081>.
- Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly*, 48(4), 626–663. <https://doi.org/10.1177/0013161X11436273>.
- Shen, J., Leslie, J. M., Spyrbrook, J. K., & Ma, X. (2012). Are principal background and school processes related to teacher job satisfaction? A multilevel study using schools and staffing survey 2003-04. *American Educational Research Journal*, 49(2), 200–230. <https://doi.org/10.3102/0002831211419949>.
- Shewbridge, C., Jang, E., Matthews, P., & Santiago, P. (2011). *OECD reviews of evaluation and assessment in education*. Denmark 2011. Paris: OECD Publishing.
- Sims, S. (2019). Modelling the relationships between teacher working conditions, job satisfaction and workplace mobility. *British Educational Research Journal*, 0(0). <https://doi.org/10.1002/berj.3578>.
- Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, 26(4), 1059–1069. <https://doi.org/10.1016/j.tate.2009.11.001>.
- Skaalvik, E. M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession: Relations with school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education*, 27(6), 1029–1038. <https://doi.org/10.1016/j.tate.2011.04.001>.
- Stanton, J. M., Sinar, E. F., Balzer, W. K., Julian, A. L., Thoresen, P., Aziz, S., et al. (2002). Development of a compact measure of job satisfaction: The abridged job descriptive index. *Educational and Psychological Measurement*, 62(1), 173–191. <https://doi.org/10.1177/001316440206200112>.
- Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, 41(5), 481–520. <https://doi.org/10.3102/1076998616646200>.
- Sun, A., & Xia, J. (2018). Teacher-perceived distributed leadership, teacher self-efficacy and job satisfaction: A multilevel SEM approach using the 2013 TALIS data. *International Journal of Educational Research*, 92, 86–97. <https://doi.org/10.1016/j.ijer.2018.09.006>.
- Urick, A., & Bowers, A. J. (2017). Assessing International teacher and principal perceptions of instructional leadership: A multilevel factor analysis of TALIS 2008. *Leadership and Policy in Schools*, 0(0), 1–21. <https://doi.org/10.1080/15700763.2017.1384499>.
- van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 54(2), 119–135. <https://doi.org/10.1016/j.erap.2003.12.004>.
- Wagner, W., Göllner, R., Werth, S., Voss, T., Schmitz, B., & Trautwein, U. (2016). Student and teacher ratings of instructional quality: Consistency of ratings over time, agreement, and predictive power. *Journal of Educational Psychology*, 108(5), 705–721. <https://doi.org/10.1037/edu0000075>.
- Walker, A., Dimmock, C., et al. (2002). Moving school leadership beyond its narrow boundaries: Developing a cross-cultural approach. In K. Leithwood, P. Hallinger, G. C. Furman, K. Riley, J. MacBeath, & P. Gronn (Eds.), *Second International handbook of educational leadership and administration* (pp. 167–202). Netherlands: Springer. [https://doi.org/10.1007/978-94-010-0375-9\\_7](https://doi.org/10.1007/978-94-010-0375-9_7).
- Weiss, H. M. (2002). Deconstructing job satisfaction: Separating evaluations, beliefs and affective experiences. *Human Resource Management Review*, 12(2), 173–194. [https://doi.org/10.1016/S1053-4822\(02\)00045-1](https://doi.org/10.1016/S1053-4822(02)00045-1).
- Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, 39(3), 398–425. <https://doi.org/10.1177/0013161X03253411>.
- Zakariya, Y. F. (2020). Investigating some construct validity threats to TALIS 2018 teacher job satisfaction scale: Implications for social science researchers and practitioners. *Social Sciences*, 9(4), 38. <https://doi.org/10.3390/socsci9040038>.