Challenges and Opportunities in Measuring School Leadership

An analysis of data from the Teaching and Learning International Survey (TALIS)

Jelena Veletić

Dissertation submitted for the degree of Ph.D.

Centre for Educational Measurement (CEMO)
Faculty of Educational Sciences
University of Oslo

2023

Series of dissertations submitted to the
Faculty of Educational Sciences, University of Oslo
No. 366

ISSN 1501-8962

Cover: UiO.
Print production: Graphic center, University of Oslo.

Acknowledgments

The very first time I spoke with my supervisor, Professor Rolf Vegar Olsen was about the announced position for this doctorate. During our conversation, he asked about my reason for coming to Oslo. My response was simple “Love”. And indeed, being with my husband was the main trigger to seek my PhD. I had never considered getting a PhD before, and neither has anyone in my extended family (except my husband). It turns out that the five-year period here in Oslo was a time of major changes but also the “time of my life”. I moved to a foreign country, learned a foreign language (which I continue to learn), met many people from all over the world, some of whom became very close friends, and most importantly, became a mother. I matured both academically and personally. This would have not been possible without other people who supported me, everyone in their own domain, alternating and complementing one another.

Professor Rolf Vegar Olsen (later just Rolf) embraced my humble background and motivated me to believe that I am capable of much more than I had previously thought. His approach to learning was accommodating and human, which washed away many of my insecurities and made me feel comfortable as a PhD student. I respect his balanced, patient, and unobtrusive style of mentoring and supervision. I am also grateful for the text edits, academic discussions, and efforts he made to polish my ideas and thoughts and to help me express my opinions in a more bold and more straightforward way. I appreciate the ideas he was sharing without any hesitation. I am thankful for his care, not only for my academic progress, but also for my personal growth. Thank you, Rolf, for introducing me and guiding me throughout the academic world. I hope and wish you learned at least a tiny bit from (and with) me as I have learned from you.

I would also like to thank Professor Sigrid Blömeke for helping me “get off the ground” with my PhD and with my academic upbringing. When times were challenging, Sigrid devoted her precious time to discuss papers, models, and alternatives, as well as to read thoroughly my writings. Although I used to think she is bit firm, I have learned over the time I really appreciate her straight and precise manner. She is a prime example of successful and fair academic, leader, and teacher. Thank you, Sigrid.

Many thanks go to my secondment host, Professor Leonidas Kyriakides, from the University of Cyprus who helped frame my first paper ever, and Professor John Jerrim, University College London, who gave extensive and valuable feedback during my final

reading. The collaboration with Professor Heather E. Price on my second article, expanded my experience, knowledge, and views on the topic from the non-measurement perspective. I would like to thank Heather not only for being patient while I was on my parental leave and for respecting my sometimes slow pace, but also for always emphasising friendly side of our relationship. I would also like to thank Professor Isa Steinman for the valuable comments at the later stages of my writing.

They were all members of a larger OCCAM group, as were many others. I consider myself extremely fortunate for being part of such a large international and interdisciplinary academic network. I appreciate training, experience, expertise, and adventure. Foremost, I am grateful for the people I met. Ana Maria, Elisa, Evi, Edwin, thank you for making long days not only about work but also about life. I want to specially thank Ana Maria who was my Cyprus companion, conference cheerleader, English dictionary, and much more.

CEMO has been an extraordinary environment in every sense. I was supported with technicalities but also with great working atmosphere. Thanks go to colleagues with whom I shared my everyday life Wangqiong, Henrik, Håkon, Saskia, KJ, Kseniia, Diego, Jose, Jarl, Maoxin, Waldir, Denise, Stefan, and Alexandra who were nothing but supportive. I also thank administrative staff at CEMO Anne-Catherine, Øystein, Gunnhild, Tara, and Amanda who made paperwork easy. Special thanks go to Professor Ronny Scherer, who I was blessed to have across the corridor, to selflessly help with multi-level, multi-group, multi-respondents, multi-hard questions.

Then, it is my husband Mladen. Academic himself, Mladen pushed me into his world of writing, re-writing, editing, polishing, discussing, reviewing, applying, accepting, and rejecting. Seriously. If I only knew. All jokes aside, thank you, Mladen, for being there. Thank you for recognising and encouraging my bravery, determination, and hard work. You pushed me outside of my comfort zone, and that made me grow. I don’t think non-academic can completely understand a hardship of doing a PhD and I am eternally grateful for having you by my side. Hand by hand with Mladen is our daughter Sonja. Sonja brought a completely new perspective on joy, priorities, determination, and fun in our lives. I learned with her about the most important things in life. Sonja and Mladen are my world above and beyond everything else.

Finally, I want to thank my friends and family in my home country. Dijana and Sanja for endless chats and texts, ultimate love, and support in everything I do. My mother for

unconditional love, safe base, and help. My sister and her family for always bringing so much joy and loudness into my life. My parents in law for supporting our family throughout this journey.

Oslo,

March 2023

Summary

Literature on school leadership is relatively scarce in terms of measurement and international comparisons, but fairly straightforward in terms of the importance of school leadership for the improvement of student learning. To properly measure and account for school leadership in a larger framework of schooling is therefore paramount. Through the lenses of an overarching theoretical, conceptual, and methodological framework, the thesis identifies five core challenges in measuring leadership in the international context, strongly reflecting on the Teaching and Learning International Survey (TALIS) design and data. The identified challenges referred to in this thesis vary from construct under-representations and the reliance on principals' self-reported measures of school leadership, through the (in) comparability of leadership measures across populations and respondents, to the issues of whether we can quantitatively capture the heterogeneity of leadership in a reliable and valid way at the country level to answer whether leadership is a school or country level phenomena. Accordingly, this work suggests alternative approaches for improvement of the current practices by showing how the measure of school leadership can be constructed from a joint perspective of teachers, and how measurement properties of the construct of school leadership vary depending on whom within the school reports about them. The thesis also shows that leadership is more homogeneous in certain countries, but relatively universal in its various manifestations within countries internationally. Thus, inferences and comparisons at the level of countries are made and discussed considering the data limitations. The empirical studies included in this thesis utilise data from teachers and principals participating in TALIS in the years 2013 and 2018, a quinquennial study administered by the Organization for Economic Co-operation and Development (OECD).

The thesis consists of two parts, the extended abstract and three articles. The extended abstract discusses the background, purpose, and relevance of the thesis by presenting and discussing the relevant school leadership literature and literature on international large-scale assessments, in particular on TALIS. Therefore, the overarching research questions engage with the current conceptualization and measurement of school leadership in TALIS and future developments of the measurement of school leadership in the context of ILSA.

Article 1 investigates the measurement properties of the instructional leadership scale as used in TALIS 2013. Further, it explores the possibility to measure leadership from the teachers' collective view by proposing two dimensions of leadership at the level of school,

managing the instructional program and developing the school learning climate. As a robustness and validation check, the separate analyses are conducted in four Nordic countries. Article 2 expands the argumentation of Article 1 by examining how teachers' and principals' perceptions of school climate fit the framework of leadership for learning across 37 countries in TALIS 2018. Different perceptions of school climate are further investigated in relation to instructional and distributed leadership. Article 3 complements the findings from Article 2 by providing insights into leadership for learning at the system level. Clusters of schools with certain characteristics are summarised at the country level informing us about the relevance of systems for leadership research. With this approach, valid and relevant inferences about leadership are drawn in the context of countries. All three articles are based on advanced statistical analyses of TALIS data that are conducted in a multilevel setting but interpreted at the level of school.

Overall, the findings presented in this thesis point to the current discussion of construct under-representation and the measurement invariance issues related to school leadership constructs in TALIS. The findings further suggest that different perceptions of school leadership within school matter for final measurement decisions. This thesis informs those who work in the field of comparative leadership and international education that within-country leadership studies are more informative for actual leadership practice. Moreover, what matters greatly is from whom we collect data about leadership within the school encouraging a "more people involved" approach. The thesis also shows that leadership clearly is a school level phenomenon. However, the examination of the heterogeneity of leadership practices across countries, which reflects the broader cultural characteristics of societies, was not supported with the TALIS data. Therefore, we would rather talk about leadership as a global practice finely shaped in local schools and societies, than about leadership as a feature of educational systems and higher-level units. Finally, as a lateral finding, the thesis provides some more substantial knowledge on the associations between school leadership and other relevant factors at the school level.

Table of Contents

List of Articles

Article 1

Veletić, J., & Olsen, R. V. (2021). Developing a shared cluster construct of instructional leadership in TALIS. *Studies in Educational Evaluation*, 68, 100942. https://doi.org/10.1016/j.stueduc.2020.100942.

Status: Published.

Article 2

Veletić, J., Price, E.H., & Olsen, R. V. (under review). Teachers' and principals' perception of school climate: The role of principals' leadership style in organizational quality. *Educational Assessment, Evaluation and Accountability.*

Status: Re-Submitted.

Article 3

Veletić, J., & Olsen, R. V. (2021). Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018. *International Journal of Leadership in Education*. Scopus. https://doi.org/10.1080/13603124.2021.1953612.

Status: Published**.**

Part I

Extended Abstract

1. Introduction

This doctoral thesis lies at the intersection of three educational research areas- research on school leadership, research on international large-scale assessments, and research on measurements in education. This is achieved in an overarching endeavour of analysing, describing, and critically accessing the current conceptual and methodological perspectives and solutions regarding the construct of school leadership in TALIS.

Although separate articles included in this thesis also discuss more specific and substantive research questions, such as an association of school leadership and teacher job satisfaction or school climate, the main point and focus of this thesis relates to measurement opportunities and challenges involved when studying leadership. More precisely, the thesis identifies five core challenges in measuring leadership in the international context heavily focusing on the TALIS study. The thesis discusses construct representation of school leadership in TALIS, the sources of information about leadership within schools, and the comparability of the construct across countries and across actors of leadership (teachers and principals) within schools, as well as the representation of the school leadership measures in a multilevel educational setting. The main rationale for focusing on the challenges in measuring leadership is twofold. First, given the multilevel nature of the leadership construct and the growing understanding that school leadership is something that is accomplished jointly by many actors, the thesis sheds light on the difficulty of measuring school leadership, as such construct, in the broader sense. The thesis also discusses how cross-cultural data, collected across as many as 47 countries, using a complex sampling design, further complicates these already challenging measurement issues. Second, the thesis suggests alternative methods for capturing school leadership by using teacher collective reports that, from the perspectives of research and practice, might be more informative and helpful. Because the TALIS study is constantly being improved, I hope that at least some of these suggestions can be considered and incorporated in the future TALIS cycles without jeopardizing the core study features. In addition to the main purpose to address measurement challenges, the separate articles also add to the more substantive evidence about the relationship between school leadership and teacher job satisfaction, and the relationship between school leadership and school climate.

In the introduction chapter, I first provide a short background and rationale for the thesis by discussing the relevance and importance of the topic in the larger framework of school leadership and by presenting the argument why this research is needed. In the next section, I

further elaborate on the overarching aim and research scope of the thesis. I conclude the chapter with a conceptual overview and a description of how the three included articles are connected.

1.1 Background and relevance of the thesis

Before moving deeply into the immediate leadership theory upon which this thesis is based, I would like to provide an overview of two distinct perspectives on school leadership as given by Simkins (2005). The author distinguishes between traditional and emerging educational leadership perspectives (see Table 1). According to the traditional view, educational leaders are the primary actors in leadership and the practice of leadership centres on leaders. In contrast, the emerging perspective considers leadership as a property of the system, practised by many, and influenced by its context. The distinction is important and helpful in the context of this thesis because the thesis is heavily focused on the measurement of leadership as described in the emerging perspective. However, the traditional perspective should not be disregarded as many of the characteristics of the traditional views on leadership have been embodied in contemporary leadership thinking and practice and most importantly in the ways how leadership is measured. Thus, much of the criticism about how leadership has been measured today is rooted in a critique of the traditionalist perspective (focus on individuals rather than a system, focus on leaders and followers rather than their mutual joint efforts, and neglection of the context in which leadership is executed, to mention few).

Table 1

Depiction of the Simkins' view on the traditional and emerging leadership views (p.12, Simkins, 2005)

Traditional leadership	An emerging leadership
Leadership resides in individuals	Leadership is a property of social system
Leadership is hierarchically based and linked to office	Leadership can occur anywhere
Leadership occurs when leaders do things to followers	Leadership is a complex process of mutual influence
Leadership is different from and more important than management	The leadership/management distinction is unhelpful
Leaders are different	Anyone can be a leader
Leaders make a crucial difference to organizational performance	Leadership is one of many factors that may influence organizational performance
Effective leadership is generalizable	The context of leadership is crucial

The larger theoretical framework that this thesis rests on is the Leadership for Learning conceptualization. Leadership for learning could be placed on the right-hand side of Table 1 as an emerging form of leadership and an integrated framework that assembles the core characteristics of earlier established leadership models such as instructional leadership (that could be seen as more traditional), as well as transformational and distributed leadership (Aas & Brandmo, 2016; Bowers, 2020; Boyce & Bowers, 2018a; Hallinger, 2009, 2010, 2011; Harris et al., 2007; Printy & Liu, 2021; Southworth, 2002). The specific features of leadership for learning revolve around learning at all levels, including not only student learning but also teacher and principal learning (MacBeath, 2019). This framework recognizes that both teachers and principals have active leadership roles within schools, simultaneously acknowledging the role of the system and features outside of the school (Boyce & Bowers, 2018b; Halverson et al., 2014; Murphy et al., 2007; Porter et al., 2008). To properly account for the model of leadership for learning that goes beyond the individual principal or individual teacher (that would be a more traditionalist way) the "new" measures of leadership must include different sources and diverse actors of leadership (Ahn et al., 2021). Failing to account for different perspectives on leadership within schools, causes the loss of information about teachers' (or principals') experiences regarding leadership within schools. This further causes the loss of information about their (in)congruence (Ahn et al., 2021; Kelley & Halverson, 2012; Park & Ham, 2016; Urick & Bowers, 2017; Wang et al., 2019). Additionally, neglecting to take into account how interactions and actions among school members shape leadership practices, leads to unsuccessful attempts to capture the emerging leadership forms (see for example, Spillane et al., 2007).

The thesis builds on this argument by studying how teachers' and principals' reports can be combined in measuring school leadership in TALIS. In parallel, the thesis demonstrates to what extent the above-mentioned larger leadership for learning framework is represented in the TALIS study. There have been limited efforts in previous research to connect the leadership for learning framework with the TALIS study, with only a few studies attempting to do so (Ahn et al., 2021; Bowers, 2020), including one of the studies that constitutes a part of this thesis (Veletić & Olsen, 2021).

TALIS is an international study that collects data from teachers and principals giving them an opportunity to describe their individual and joint practices and to express their opinions, perceptions, and beliefs in relation to a number of features known to be of importance for the organizational quality and ultimately for the quality of schooling. TALIS stands out as

unique in this regard and is a commonly used resource for a variety of studies that examine teachers and principals as school members, as well as the structure and administration of schools. The objectives of TALIS are 1) to help in reviewing and developing educational policies; 2) to help teachers, principals, and other school stakeholders reflect on and improve their daily practices; 3) to build on existing research evidence to create a strong and dependable foundation for future research. Therefore, the three main groups that benefit from TALIS are policymakers, educational practitioners, and researchers (OECD, 2019b).

Although it is widely accepted in educational effectiveness literature that school leadership constitutes one of the crucial determining factors of school quality (Bellibas et al., 2021; Hallinger & Huber, 2012; Hendriks & Scheerens, 2013), one should keep in mind that TALIS is not a leadership study. It is a widely scoped survey including and studying working life in schools in general, covering a broad range of topics such as schools' physical environment, interactions among school stakeholders, job satisfaction, teaching practices and professional development, to mention a few (OECD, 2019b). In a larger venture to capture and describe various relevant constructs that operate at the different educational levels, and across as many as 47 countries, TALIS, like many other ILSA, must achieve a balance between including measures of as many relevant constructs as possible and keeping the time burden to the respondents low enough. Thus, a common concern is that the included constructs are measured by too few items, which is another core issue that this thesis investigates.

The concept of school leadership is quite challenging to operate in the context of international comparative studies, for reasons that relate to 1) the existing knowledge about the leadership in a broader educational literature and 2) requirements and limitations of approaches to measuring relevant constructs on a large-scale. In brief, the former challenge relates to the fact that a range of partly overlapping definitions of school leadership exist in the literature, including a vagueness in the definitions about whom is included in the leadership practice, and furthermore, limited knowledge about the relevance of cultural features for the leadership practice across the world. The challenge of measurement includes concerns with designs exclusively relying on self-reported data, the measurement properties of the instruments with respect to cross-cultural comparability, and the decisions about who constitutes the best source of information on leadership in schools. Overall, the thesis seeks to contribute to a better understanding of these challenges as well.

1.2 The overarching aim and research scope

Due to the perceived limitations of quantitative studies in capturing the complexities of leadership practices, researchers in the field of school leadership heavily rely on qualitative studies to investigate relevant phenomena (Klenke, 2008). Nevertheless, there is also a substantial body of literature on school leadership based on quantitative data (Condon & Clifford, 2012; Klenke, 2008). However, the impression is that the current tools used in leadership studies rely greatly on traditional ideas about leadership, concentrating predominantly on leaders, hierarchical structures, and roles rather than on the practice of leadership as a characteristic of the group or a process of reciprocal influences (Condon & Clifford, 2012; Spillane & Zuberi, 2009). Thus, instruments to measure school leadership as a group phenomenon that has been achieved by a joint force of teachers, principals, parents, students, and other school members, in particular in a cross-cultural setting, are scarce. Consequently, since the first cycle of TALIS, the data has been extensively used as a resource by researchers with an interest in studying school leadership practices worldwide (Bellibas et al., 2021; Berkovich & Bogler, 2020; Gumus, 2013; Ham & Kim, 2015; Liu, 2020a; Printy & Liu, 2021; Xia & O'Shea, 2022). This was also confirmed in an ongoing systematic review of studies that use TALIS data for secondary analysis (see section 2.2.2 in this thesis), in which one-third of all articles that were selected for the review included leadership as a construct of interest. Under these circumstances, TALIS as a potential source to study leadership, receives increasingly more attention. Thus, it is crucial to have precise, reliable, and valid scales and measures of school leadership that provide an opportunity to capture leadership as a complex process of mutual influence, that can occur anywhere in the school and that is not linked exclusively to what principals do. This is not only important for TALIS, but also for the broader research community interested in understanding and contributing with new knowledge about school leadership.

In this thesis, I study separate issues related to the measurement of school leadership in TALIS and refer to them as challenges in measuring school leadership (See Figure 1). The first challenge refers to construct representation that deals with how well the construct of leadership is represented by the leadership scales in TALIS. This issue is mostly covered in Article 1. The second challenge refers to leadership perceptions, that is how leadership is perceived by different school stakeholders focusing exclusively on teachers and principals. This challenge also includes discussion on whom we should question about leadership within individual schools. Both Article 1 and Article 2 cover this issue. The next challenge, cross-respondent

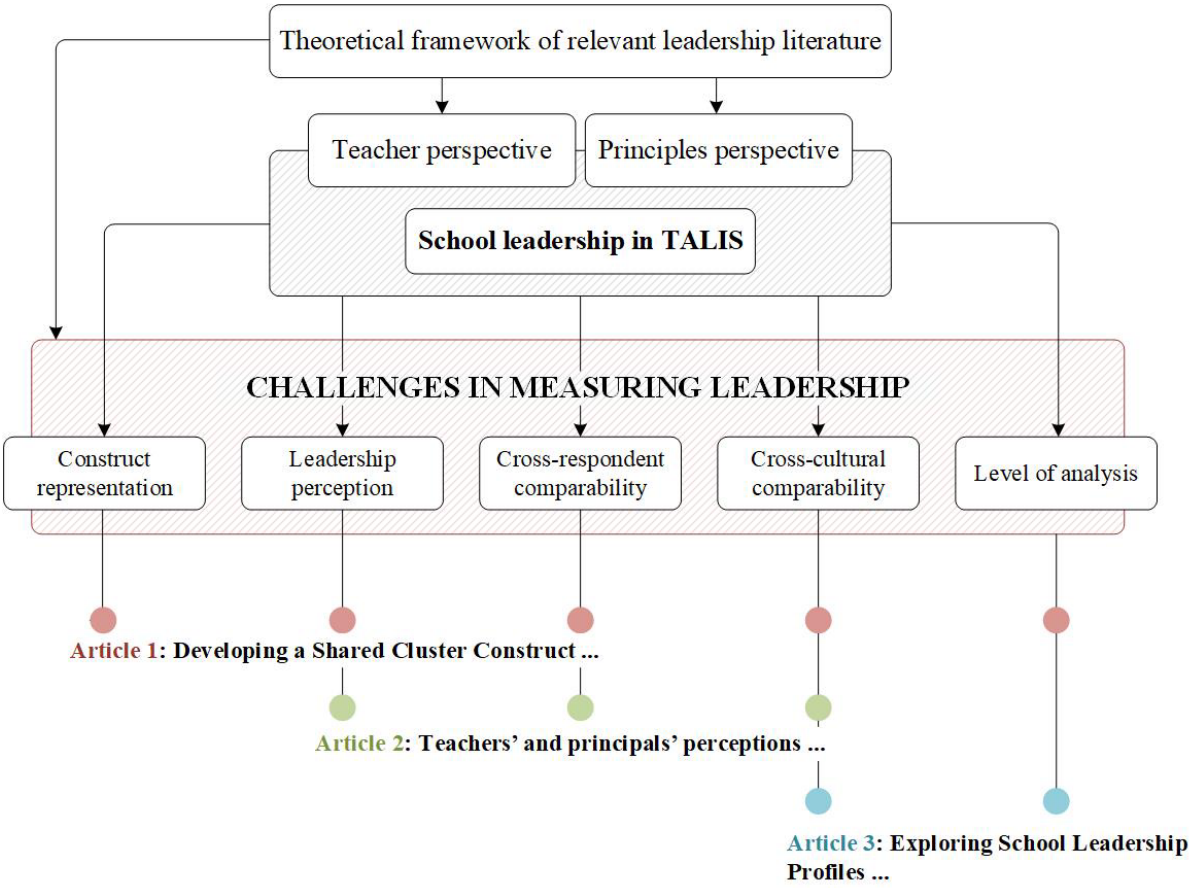
comparability, applies to the comparability of the responses collected across teachers and principals. This challenge is mostly addressed in Article 2 in terms of measurement invariance between teachers and principals in TALIS. Similarly, the fourth challenge, cross-cultural comparability, refers to the comparability of the leadership construct in TALIS across countries. Because all three articles include multiple countries this challenge is addressed in all of them. The final challenge, pertaining to the level of analysis, is linked to determining the most suitable and relevant level of analysis for investigating school leadership. Specifically, it raises questions about whether the heterogeneity in leadership across countries can be effectively captured with TALIS data, and whether it is meaningful to generalize leadership at the system level. Article 3 addresses this as a core issue. Overall, I reflect on and demonstrate how these challenges can influence the way data from TALIS are used in studies that explore more substantive research questions on school leadership.

Therefore, Article 1 shows why the current approach to measuring leadership in TALIS where principals self-report about their practices might not be completely reliable and valid. The study demonstrates the issue by comparing how the association with teacher job satisfaction differs based on whether leadership is measured from a teacher or principal perspective. The teachers' collective view on leadership is discussed and proposed as a recommended method. As a consequence of the first study's findings, Article 2 seeks to study a measure of the school climate with scales constructed from information from both teachers and principals. This study focuses on the congruence between teachers' and principals' perspectives on school climate. From a measurement perspective, the study demonstrates why it is crucial to collect data on such jointly experienced phenomena from both teachers and principals, while from a substantive viewpoint, it analyses how these differences in perception are associated with leadership practices. Finally, Article 3 illustrates that most of the variation in leadership measures lies between schools, and that distinctive leadership profiles at the country level cannot readily be identified.

In Figure 1, I show how these three articles connect in their larger effort to address the five main challenges when measuring school leadership in TALIS. I also show how these challenges are shaped by the attention that school leadership receives in the TALIS study, but also in a larger framework of school leadership research. I give an emphasis on teachers' and principals' perceptions of school leadership as those two groups represent key school actors that constitute leadership practice.

Figure 1

An overview of the core five challenges as identified in this thesis and how the three articles are connected with respect to them



1.3 Overview of the articles

In this section I first give a summary of each of the articles followed by Table 2 in which I summarize specific details related to data, sample, analysis, and research questions in each of them.

Article 1

Veletić, J., & Olsen, R. V. (2021). Developing a shared cluster construct of instructional leadership in TALIS. *Studies in Educational Evaluation*, 68, 100942. https://doi.org/10.1016/j.stueduc.2020.100942. **Published.**

The main purpose of this study was to demonstrate how teacher data from TALIS can be used in a multilevel setting to measure instructional leadership at the level of school. The study first shows no association between instructional leadership as measured in TALIS and teacher job satisfaction. The severe construct underrepresentation of the instructional leadership scale (measured by three items only) in TALIS was discussed as a potential reason for that. Consequently, teacher data are used to build two dimensions of instructional leadership (managing the instructional program and developing a school learning climate) at the level of school and the association with teacher job satisfaction was re-examined. The newly established two dimensions of instructional leadership reported by teachers are found to be not only moderately positively correlated with the existing instructional leadership scale in TALIS, but also with teacher job satisfaction that served as an external criterion. As a validation step, the analyses were conducted in four Nordic countries. The article suggests that using teachers' collective observations, rather than principal self-reports only, is a promising direction in further developments of leadership measures in large-scale education studies.

Article 2

Veletić, J., Price, E.H., & Olsen, R. V. (under review). Teachers' and principals' perceptions of school climate: The role of principals' leadership style in organizational quality. *Educational Assessment, Evaluation and Accountability.* **Re-Submitted.**

This study used parallel items from the teacher and principal questionnaires in TALIS 2018 to build a new measure of school climate that at the same time represents an aspect of leadership for learning achieved jointly by teachers and principals. The main purpose was to study the coherence between teachers' and principals' perceptions of school climate at the level of school. The article illustrates that school climate is perceived differently in all countries and

in most cases, these differences reflect that principals rate the school climate as better than the teachers in the same school. The article also shows that both principals' and teachers' views of school climate are associated with the principal's leadership style. In particular, a higher emphasis on distributed leadership is positively associated with the perceptions of school climate. This association is particularly stable for the principals' perceptions. The article also discusses that multiple views on the same phenomena are informative and a necessary feature of surveys that deal with school leadership.

Article 3

Veletić, J., & Olsen, R. V. (2021). Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018. *International Journal of Leadership in Education*. Scopus. https://doi.org/10.1080/13603124.2021.1953612. **Published.**

The aim of this exploratory study was to establish clusters of countries shaped with specific leadership for learning practices. However, when data from teachers and principals are examined in an unconditional three- or two-level hierarchical model, most of the variation in school leadership lies between schools, whereas the between-country variance is very small. Therefore, five different clusters of schools were established at the school level and their distribution was examined at the country level. The article finds that the distribution of schools that belong to a specific cluster at the country level did not reflect easily identifiable geographical or cultural similarities. The article further shows that some countries have a more homogeneous leadership practice than others. Moreover, most countries are not dominated by one or two leadership types, rather all clusters representing distinctly different leadership practices appear in each individual country. Finally, the article also concludes that there are no substantial differences across clusters with respect to the background characteristics of schools and principals.

Table 2

Overview of the articles that constitute this thesis

	Article I	**Article II**	**Article III**
Title	Developing a shared cluster construct of instructional leadership in TALIS	Teachers' and principals' perception of school climate: The role of principals' leadership style in organizational quality?	Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018
Main Research question	To what degree are two proposed measures of subdimensions of instructional leadership as reported by teachers associated with teacher job satisfaction and instructional leadership as measured in TALIS?	To what extent do teachers' and principals' views of school climate differ within and across schools? To what extent are principals' leadership styles associated with these perceptions of school climate?	What characterizes different profiles of leadership for learning across countries and schools? To what extent countries (and schools) be classified into groups based on leadership for learning practice?
Overarching theoretical framework	Instructional leadership	Leadership for learning	Leadership for learning
Sample and data	Principals and teacher data from TALIS 2013. Nordic countries included (Denmark, Norway, Sweden, and Finland)	Principals and teacher data from TALIS 2018. The intention was to include the whole sample (48 countries) but due to different reasons only 37 countries were analysed.	Principals and teacher data from TALIS 2018; The intention was to include the whole sample (48 countries) but due to missing data only 43 countries were analysed.
Analyses	Multilevel confirmatory factor analysis, Linear regression	Multigroup multilevel confirmatory factor analysis, Linear regression; Fixed effects	Multilevel analysis, Cluster analysis

1.4 Outline of the thesis

This PhD thesis consists of two parts. The first part, the extended abstract, summarizes and gives a bird's-eye view of the three articles, whereas the second part contains each of the three co-authored articles. Part I consists of five larger chapters. The introduction chapter consists of three parts discussing the background of the thesis, the research scope, and the thesis outline. Chapter 2 refers to the theoretical framework and consists of three larger sections. The first section gives an overview of the school leadership literature in general; the second section gives an overview of the school leadership as represented in ILSA with a special reference to the TALIS study. Finally, in the third part of the theoretical framework, the discussion is narrowed to the core challenges related to the measurement of leadership as identified in this thesis. Chapter 3 outlines the methods and methodological choices that guided the work in the articles. Chapter 4 provides a summary of the three articles greatly focusing on the findings whereas Chapter 5 discusses the findings of the three articles but in relation to each other giving the emphasis on the overall thesis contributions to the approaches in measuring school leadership, implications for the TALIS study, and larger theory of leadership.

2. Theoretical framework

This chapter consists of three larger theoretical sections. The first section (2.1) gives an overview of school leadership literature with an emphasis on the definitions and conceptualizations of specific relevance for the constructs and measures implemented in TALIS. Thus, it heavily relies on instructional and distributed leadership as these concepts are explicitly referred to in the framework for the TALIS study. Building on this, the leadership for learning (LFL) conceptualization is discussed more in-depth as a promising approach to synthesize both instructional and distributed leadership theories while keeping learning in the focus. Because ILSA by nature are international comparative studies, and leadership in the broader, emerging literature is considered as dependent of system features, the chapter further discusses the context of leadership and its relevance for leadership practice. The chapter is concluded with a more general reflection on how the evidence from both quantitative and qualitative approaches is important for the comprehensive study of leadership with short notes on foreseeing issues. The second section (2.2) introduces international large-scale assessments and how the current TALIS study relates to the wider theoretical framework of school leadership. The importance of TALIS data for the wider study of leadership is discussed through the lenses of the findings from a systematic literature review of peer-reviewed journal articles that used TALIS data to study leadership. The section further demonstrates the development of a framework for and measures of school leadership in TALIS since the first study in 2008. Finally, in the third part (2.3), the five core challenges in measuring leadership as identified in this thesis are discussed against the theoretical foundations upon which they are based. Throughout, I refer to the school principal as a person who is responsible for the operation of the school. In most cases, individual schools have one principal. In very small schools, principal may also teach part-time, whereas in larger schools there may be more than one assistant principal (Lunenburg, 2010).

2.1 School leadership

Leadership is a fickle and difficult construct to define. Literarily, there are more than three hundred definitions of leadership available in the literature (Cuban, 1988). They differ in terms of the source, process, and outcome of leadership. Broadly speaking, leadership is frequently defined as a form of influence exercised not only by principals but also teachers and other school members (Harris, 2009a; Leithwood, 2001; Leithwood & Duke, 1998; Ogawa & Bossert, 1995), frequently shaped by personal and professional values (Begley, 2001; Brooks

& Mutohar, 2018) and concerned with some kind of vision (Bush & Glover, 2014; Hallinger, 2010; Southworth, 1993). For example, Cuban (1988, p.191.) defined leadership as "a way of organizing followers and manipulating settings to produce desired results".

Leadership has long been seen as one of the key factors in improving school effectiveness (Hallinger & Heck, 1998; Huber & Muijs, 2010; Leithwood et al., 2020; Mincu, 2022; Robinson et al., 2008). In their revisited article on seven strong claims about successful school leadership, Leithwood and colleagues (2020) conclude that leadership has a significant, moderate effect on the organization of the schools, which in turn influences the quality of teaching and learning. Even though the mechanisms through which leadership influences schools are not always completely clear, it is evident that it must work indirectly through actors and actions affecting the content and quality of instruction and students' efforts. A significant body of literature identifies such features, e.g. actions to establish goals and expectations, to take care of teachers' well-being, to maintain good working conditions, and to shape instructional practices (Bellibas et al., 2021; Burkhauser, 2017; González-Falcón et al., 2019; Hallinger & Huber, 2012; Ladd, 2009; Sims, 2019). Thus, the indirect effects of leadership on learning are most reported in the literature (Dumay et al., 2013; Hendriks & Scheerens, 2013; Ladd, 2009; Ross & Gray, 2006) surpassing the studies about the direct effects (Kyriakides et al., 2010; Witziers et al., 2003). That is either because of the inconsistent evidence about its direct effects (Krüger et al., 2007) or because the effects are only found to be significant for certain groups, such as students with special needs or language difficulties (Nettles & Herrington, 2007). All in all, this literature suggests that leadership affects numerous aspects of schools, some of which are more closely and directly linked to leadership (e.g. teacher job satisfaction), and others more distal and influenced by multiple factors with leadership being one of many in a web of mediation causal agents (e.g. student learning outcomes).

One more important attribute when we discuss the effects of school leadership on student achievement is how leadership itself is defined. Therefore, we frequently find leadership conceptualizations[1] such as "instructional leadership", "distributed leadership", and "leadership for learning" to mention a few. This list is not even close to being exhaustive, but

[1] Leadership conceptualizations are sometimes referred to as leadership models or styles in the education literature. In this thesis when I talk about a general idea or a concept that represents a specific leadership practice, I use the term conceptualization. First, the term "model" can allude to or easily be confused with the mathematical representation of data in statistical models. Second, the term "style" connotates with the immediate and exterior behaviour of a person. However, I sometimes use the terms "model" or "style" to keep the original authors wording. Thus, those can be read interchangeably.

it gives a glimpse of the idea behind the need to conceptualize leadership differently. Thus, conceptualizations of leadership are heavily influenced by its core focus which is either instruction, organization, or management (Agasisti et al., 2019). As an illustration, leadership that is conceptualized as instructional is the most emphasized and studied in relation to student learning outcomes (Day et al., 2016; Karadag, 2020; Louis et al., 2010; Özdemir et al., 2022; Robinson et al., 2008) focusing on the factors closely related to learning such as instruction, learning climate, and academic and learning goals (Hallinger, 2003, 2005).

2.1.1 School leadership conceptualizations

The main leadership conceptualizations emerging from the current literature are nicely summarized in Özdemir et al. (2022) systematic review of literature on the relationship between school leadership and student achievement in which at least 12 conceptualizations were identified (instructional-, school-, principal-, educational-, distributed-, shared-, collaborative-, transformational-, transactional-, integrated-, teacher-, organizational-, supportive-, visionary- leadership). Some of these conceptualizations were grouped into smaller clusters as they indicate similar practices. For example, distributed leadership, shared leadership, and collaborative leadership are often used to describe closely related leadership conceptualizations where several actors enact roles and functions of leadership. Similarly, transformational and transactional leadership were grouped together to represent conceptualizations of leadership centring on leaders establishing new norms, changing the attitudes of those who are led, and making fundamental changes to the culture of the organization. Interestingly, a large proportion of studies included in this review chose to use terminology referring to leadership as a more generic function, e.g., school leadership, principal leadership, or educational leadership rather than restricting it to a specific conceptualization. Thus, several of these studies reflect a rather vague and unspecific notion of leadership, which in the end is a challenge when seeking to interpret and synthesize evidence in existing literature (Wang & Ahn, 2023).

Instructional leadership

The most commonly referred conceptualization when examining student learning outcomes is instructional leadership (Hallinger, 2005; Louis et al., 2010; Özdemir et al., 2022; Robinson et al., 2008). Instructional approaches to leadership give school leaders a main role and importance in managing and leading schools (Blase & Blase, 2000; Bossert et al., 1982; Hallinger, 2005). Instructional leadership predominantly emerged and was maintained in the

context of the US (Blase & Blase, 2004; Gumus et al., 2018; Heck et al., 1990; Krug, 1992). However, after Hallinger and Murphy (1985) proposed their conceptualization of instructional leadership, accompanied by a corresponding instrument for measuring it, the Principal Instructional Management Rating Scale (PIMRS), the model raised in popularity worldwide. This model synthesizes three main dimensions of instructional leadership (defining the school mission, managing the instructional program, and developing the school learning climate) and has been the most fully tested and widely adopted instrument in studies of instructional leadership (Southworth, 2002). According to this model, instructional leaders directly engage with the coordination of curriculum and instruction in the classroom. Strong instructional leaders further protect instructional time, promote professional development, provide incentives for teachers and learning in general and keep themselves visible in everyday school life. Framing and communicating school goals falls into their responsibility as well. This conceptualization describes a set of ideal responsibilities, and to some extent, it is fair to say that, collectively, they describe ambitions that are unrealistic for one person to achieve. Accordingly, school leaders frequently report difficulties to balance administrative, managerial, human resource, and institutional tasks with instructional and curriculum functions (Hallinger, 2005; Hoy & Hoy, 2006). In line with such findings, the emerging views on school leadership have increasingly been regarded as functions shared among some, many, or maybe all school stakeholders (Gronn, 2002; Harris, 2004; Marks & Printy, 2003; Spillane et al., 2004). Therefore, in the context of the USA, the Vanderbilt Assessment of Leadership in Education (VAL-ED) was developed to assess the effectiveness of principals' behaviours that influence teachers' outcomes and in turn student learning outcomes, using a multi-rater approach, thus acknowledging that "the information about leadership resides within the shared experiences of these individuals" (Goldring et al, 2015, p.180, Porter et al., 2008,). The VAL-ED framework suggests that school leadership should be assessed at the intersection between core components (high standards for learning, rigorous curriculum, quality instruction, culture of learning and professional behaviour, connection to external communities, and performance accountability) and key leadership processes (planning, implementing, supporting, advocating, communicating, and monitoring) (Goldring et al., 2015)

Distributed leadership

The scope of distributed leadership expanded from a notion of sharing tasks to more complex conceptualizations of joint collaborations and interactions, including individual and collective endeavours in both more or less hierarchical leadership forms (Tian et al., 2016).

Thus, multiple actors share responsibility, participate in decision-making and are held responsible for joint organizational goals (Tian et al., 2016). Despite this general knowledge, studies on distributed leadership still encounter similar limitations as those found in research on school leadership overall. Strictly speaking, researchers have been unable to simply and directly conceptualize distributed leadership and empirically outline its application (Hairon & Goh, 2015; Hulpia et al., 2009; Liu & Werblow, 2019; Lumby, 2016). Some authors conceptualize distributed leadership as a decision-making process that involves collaboration among school stakeholders (Hallinger & Heck, 2010), as a process of shared accountability (Hulpia et al., 2012) or process of professional learning communities (Marks & Printy, 2003). The most helpful and applicable conceptualizations of distributed leadership for this thesis are those proposed by Spillane (2004) and Gronn (2000), which are also some of the most commonly referenced and used in the broader leadership literature. Spillane's model of distributed leadership emphasizes the importance of not only shared leadership among multiple individuals, but also how such leadership is being generated through individuals' interactions and practices. Therefore, *leadership practice* is one of the central concepts in his theory. Additionally, the interactions between leaders, followers, and their situations are crucial in generating such practice. In such a system, the situation not only defines the leadership but is also shaped by leadership, as explained by Spillane et al. (2001). On the other hand, Gronn (2009) build on practice-centred leadership by acknowledging the importance of both shared leadership forms and forms of leadership solely in hands of principals and named it the hybrid model. The discussion about leadership as a shared practice, and leadership as an accumulation of practices executed by different people, has continuously been deepened in the work of some other authors as well (see for example Devos et al., 2014; Harris et al., 2007; Hulpia et al., 2012).

Overall, the study of school leadership embraced a distributed perspective and it became a preferred and advocated method of leading schools (Harris, 2004; Klar et al., 2016; Lumby, 2016). The attractiveness of distributed leadership lies in its capacity to include more people in leadership roles, using their skills and expertise which consequently creates wider opportunities for all. However, there is some theoretical evidence showing that leadership practice that is informally distributed can negatively influence the quality of outcomes and that fewer leaders are preferable, as explained by Harris (2009b). Further drawbacks of distributed leadership have been noted, including the difficulties in establishing priorities, goals, and timelines. Therefore, Harris (2009) suggests that the research needs to provide more insights into the

barriers, unintended consequences, and limitations of distributed perspectives on leadership in order to give recommendations about the "optimal" leadership practice.

Leadership for learning

In the complexity and competitiveness of different leadership conceptualizations, incomplete definitions, and the overlapping models of leadership practices (Bush & Glover, 2014; Daniëls et al., 2019; Gronn, 2003) the conceptualization of learning-centred leadership started to gain more attention recently. Leadership for learning is seen as a concept that integrates the above-mentioned leadership conceptualizations, i.e. instructional leadership, distributed or shared leadership, and transformational leadership (Daniëls et al., 2019; Hallinger, 2011; Tulowitzki et al., 2021). A distinctive feature of leadership for learning is its wider, developmental view of learning. Such learning is not exclusively associated with student outcomes through the improvement of instruction, but encompasses professional, organizational and leadership learning. In other words, learning at all levels of the educational system (MacBeath & Dempster, 2008). Additionally, the position of formal leaders is de-emphasized and interactions at the lower levels (e.g. among teachers) are seen as a source of leadership as well (Harris, 2009a). Moreover, interactions rather than actions of individuals are seen as critical for leadership practice, so leadership for learning is considered a group activity or group level phenomena (MacBeath & Townsend, 2011).

This model emerged not only in the US (Boyce & Bowers, 2018a; Hallinger, 2011; Murphy et al., 2007), but also in Europe, mainly through the comparative leadership project Carpe Vitam that involved several countries (Norway, Sweden, Denmark, Austria, Greece, UK, Australia, and the US) (MacBeath et al., 2005). Although both groups acknowledged the importance of learning in the learning-centred leadership, the ideas resting on instructional leadership practices shaped the development of leadership for learning framework in the US, so those have frequently been used as synonyms (Hallinger, 2011; Murphy et al., 2007) or preceding practices (Boyce & Bowers, 2018a; Hallinger, 2009). For instance, the VAL-ED instrument uses both terms instructional leadership and learning-centred leadership in their conceptual framework (Goldring et al., 2015; Porter et al., 2008). The emergence of the Comprehensive Assessment of Leadership for Learning (CALL) in the US had a significant effect on the development of leadership for learning research. CALL is a multi-source assessment of leadership practice, tested over thousands of educators in the US (principals, teachers, school administrators) at all educational levels (primary, secondary, district)

(Halverson et al., 2014). The leadership for learning conceptualization that emerged from this assessment is five-dimensional. The main dimensions are 1) focus on learning, 2) monitoring teaching and learning, 3) building nested learning communities, 4) acquiring and allocating resources, and 5) maintaining safe and effective learning environments (Halverson et al., 2014).

In parallel, the Carpe Vitam project with its base in Europe, had two main goals. First, it aimed to address the demands of a culture that emphasizes the importance of measurable outcomes by analysing how school leadership influences learning in diverse cultural contexts. Second, the project intended to support a set of democratic principles grounded in a decentralized model of leadership (Frost & Swaffield, 2008). This project identified five key leadership for learning principles: sharing leadership, maintaining a focus on learning, sustaining a learning dialogue, creating an environment for learning, and reframing accountability (MacBeath, 2019).

Both projects, Carpe Vitam and CALL made some important calls for research on leadership in schools. First and foremost, leadership capacity should be built among all members of the school community, and not only through principals (Harris, 2004; Lambert, 2002). Second, leadership is essentially not about leaders, but rather about processes that emerge from the reciprocal interactions among participants (Lambert, 2002). Third, a focus on learning does not only imply students' learning but also teachers', principals', schools' and system learning, enabling leadership capacity to arise out of those learning experiences (MacBeath, 2020). Fourth, school conditions, context, and culture are of the main importance for creating an environment in which learning and leadership can develop (Mitchell & Sackney, 2011; Thoonen et al., 2012). Moreover, leadership for learning involves a shared sense of accountability within schools and beyond, e.g. accountability to external agencies (MacBeath & Dempster, 2008). Lastly, leadership for learning is about shared leadership that is sometimes understood as a delegation of tasks, spontaneously shared practice, and/or teamwork. The understanding of "sharing leadership" heavily varies across contexts and situations in which leadership is enacted. Therefore, how leadership is executed to a large extent depends on specific situational and wider contextual features (MacBeath, 2020; Oc, 2018).

2.1.2 The context for leadership

There is a universal agreement about the importance of context for leadership practice (Clarke & O'Donoghue, 2016; Hallinger, 2018; Harris, 2020) with a considerable amount of research supporting this claim. However, the larger body of the existing literature lack precision

and a more systematic overview of what constitutes the context for leadership (Oc, 2018). For instance, Braun et al. (2011) in their framework of factors that influence differences in policy enactments between schools, differentiate between situated, professional, material, and external contexts. Therefore, the context of leadership sometimes refers to national and system-level features, such as national culture, institutional organizations, the composition of social groups or economic conditions, and in other cases, it refers to organizational school culture, organizational dynamic, school, and class compositions, physical- and human resource environments, to mention a few. For the purpose of this thesis, it is useful to distinguish between these two main categories of contextual factors that exist at both the macro and micro levels.

Contextual factors at the macro level

The evidence surrounding culturally, and contextually embedded leadership mostly relates to the macro level, and most typically this macro context is included as a lens in studies demonstrating that successful educational policies for school leadership and governance cannot always be successfully transferred from one educational system to another (Gurr, 2014; Harris, 2020; Hooge, 2020). Hallinger (2018), for example, highlighted three distinct contextual dimensions that shape leadership practice in schools: 1) Economic (the level of economic development of society), 2) Political (prevailing power structures and relationships as well as political actors), and 3) Sociocultural (prevailing values and norms). His framework emerged as an extension of the Bossert et al. (1982) instructional management model that focuses on micro factors such as leaders' personal characteristics, institutional context, school climate, and instructional organization.

Møller and Schratz (2009) discuss the socio-cultural, historical, and political context of England, Scandinavian, German-speaking, and Eastern European countries, highlighting the differences and similarities between them. The authors emphasise the importance of cultural heritage and language, as well as the cultural and social norms that shape these societies. Ševkušić et al., (2021), documented similarities and differences in school leadership in the Balkan countries, concluding that in most of the selected countries there is a need for higher professionalization of leadership, more autonomy, and a need to relieve principals from administrative overload. The authors further showed that principals' preparation and training varied widely across these countries. While Slovenia has a long practice of school leaders' preparation at the national level, Serbia only recently produced legislation and training for

school leaders. In Bosnia and Hercegovina, the extensive complexity of the educational system prevents national-level actions on school leadership. Flessa et al., (2018) reviewed published articles and policy documents on school leadership in Latin American countries and concluded that countries differ greatly in the explicit emphasis on instructional leadership for principals. In Chile, the supervision of classroom practice directly by principals is part of the policy, in contrast to other Latin American countries. When it comes to the professional development of school leaders, Costa Rica has public (ministry-based) regulated practice, whereas Colombia, Ecuador, and Mexico offer training in both public and private sectors. Chile and Peru do not have any official regulations and the professional development of principals is mostly private.

From such examples, one can conclude that even though countries operate within their own broader context that is shaped by the same or similar language, regional culture and history, substantial differences still exist. Therefore, successful principals and other actors of leadership need to adjust their behaviours, approaches to leadership, and everyday practice to the particular micro context (Brauckmann & Schwarz, 2014; Goldring et al., 2008; Gurr, 2014; Miller, 2018). This points to another important issue in the current literature, and that is a huge reliance upon mostly Western leadership models and research evidence, that cannot be easily translated and used elsewhere (Hofstede, 2001; Oplatka & Arar, 2017; Walker & Dimmock, 2002).

Contextual factors at the micro level

At the micro school level, it is important to distinguish between school climate and school culture (Hoy, 1990; MacNeil et al., 2009; Schoen & Teddlie, 2008). While some authors subsume school culture under school climate (Van Houtte, 2005), others see school climate as a level of school culture (Schoen & Teddlie, 2008). School culture refers to shared beliefs and values that connect the multiple actors of a school community (Deal & Kennedy, 1983; Hoy, 1990; Scheerens & Bosker, 1997) or to unwritten rules and traditions, norms and expectations that permeate everything (Deal & Peterson, 1999). Distinctively, school climate refers to patterns of people's experiences of school life and reflects norms, goals, interpersonal relationships, teaching and learning practices, and organizational structures (Thapa et al., 2013). Wang and Degol (2016) define climate as consisting of four interrelated elements: academic (academic atmosphere, curricula, instruction, teacher training and professional development), community (interpersonal relationships), safety (physical and emotional security), and institutional (organizational and structural features) climate.

Schoen and Teddlie (2008) describe both concepts as very similar yet coming from different research traditions. School climate is typically studied from a psychological perspective, whereas school culture is viewed from an anthropological perspective. Thus, climate is defined in terms of behaviours, and culture in terms of values and norms. Moreover, the term climate is a preferable term in the quantitative literature on school effectiveness, whereas the term culture is more used in qualitative research (Schoen & Teddlie, 2008). According to Schein (2010) school culture has three levels: 1) basic underlying assumptions (unconscious, taken-for-granted beliefs, thoughts and feelings, 2) espoused values (strategies, goals, philosophies), and 3) artifacts (visible organizational structures and processes). The second level, espoused values, involves participant perceptions that are typically aggregated at the level of school and used to describe the psychosocial construct of school climate. Certainly, school climate has been defined as shared perceptions of the work environment and behaviours (Ashforth, 1985; Hoy, 1990), connected to what happens in everyday school life and daily practice.

Defined as that, school climate associates with numerous factors at all educational levels. At the student level, it is one of the most recognized factors relevant for students' learning and wellbeing (Gustafsson & Nilsen, 2016; Hoy et al., 2006; Scherer & Nilsen, 2016). At the teacher level, school climate associates with teacher job satisfaction and self-efficacy (Aldridge & Fraser, 2016; Collie, 2012; Katsantonis, 2020), teacher beliefs, commitment, and engagement (Collie, 2012; Dickhäuser et al., 2021; Muijs & Reynolds, 2002). At the school level, among other factors, school climate associates with the construct of school leadership and their tight connection is not clear in the literature with respect to what constitutes what, that is if school leadership is part of school climate construct or vice versa (Griffith, 1999; Kelley, 2005; Kozlowski & Doherty, 1989). An examination of reciprocal relations between the two would give answer that is more accurate, but such studies are limited in the literature (Kozlowski & Doherty, 1989). As of now, we have certainly learned that leadership processes have important implications for formation and maintenance of climate perceptions. That is, the nature and quality of interactions with principals provide the basis for perception of climate by those who are led (Kozlowski & Doherty, 1989).

2.1.3 The qualitative versus quantitative evidence about leadership

This section approaches the evidence presented previously in the study of leadership from a different perspective. It supplements knowledge about leadership by providing a detailed

examination of how qualitative and quantitative approaches complement each other in the study of leadership. It also highlights the nature of the instruments used in leadership research and broader issues in framing leadership as a multilevel phenomenon with foreseeable reciprocal associations at the school level.

The evidence regarding school leadership comes from various sources and study types. While quantitative studies have the great potential to bring theoretical advancements and clarifications about leadership practice, they have been criticized for complex methodology and the inability to inform the audience about deeper structures of the phenomena (Klenke, 2008). Consequently, qualitative approaches in the study of leadership gained more attention. Supporters of this research strand argue that qualitative approaches offer more opportunities to study leadership beyond quantities and provide more in depth answers to the "why" type of questions rather than "what" and "how" (Klenke, 2008). In her review of research method trends in school leadership journals between 2013 and 2017, Jackson (2019) showed that qualitative approaches were the most common, accounting for 45% of all studies, whereas quantitative, non-empirical and mixed methods studies, shared approximately 20% each. Similarly, Thomson (2017) in his review of 208 studies from six educational leadership journals, showed that only 26 % of the total corpus had some form of quantitative approach-surveys, model testing or secondary data analyses. Gumus et al. (2018) conducted a systematic review of studies on leadership models in educational research from 1980 to 2014 and identified 183 studies that were directly related to the leadership of which as many as 81 or 41% were qualitative.

Nevertheless, quantitative approaches to the study of leadership still constitute an important and necessary source of information about the phenomenon (Bellibas & Liu, 2017; Hallinger & Leithwood, 1998, p. 40; Hallinger & Murphy, 1985; Hulpia & Devos, 2009; Kelley, 2005; Kılınç et al., 2022; Xia & O'Shea, 2022). The quantitative approaches typically employ correlational designs with the purpose of explaining relationships (or making cautious causal claims) between carefully developed measures of constructs (Firestone, 1987). The typical study of leadership involves self-reported surveys or questionnaires that are administered either to leaders or those who are led (Hunter et al., 2007). In doing so, researchers make numerous and rather bold assumptions, e.g. that a group of subordinates witnessed and reports about the same leadership behaviours in questionnaires, and that leadership affects each of them equally. Hunter et al. (2007) conclude that, to remedy this, a balance must be struck among sources, providing information from multiple and varied perspectives. This thesis

addresses this issue by showing how the existing large-scale leadership surveys can be improved by including both teacher and principal perspectives and why none of them alone can give a comprehensive picture of leadership in schools. The thesis also emphasizes the value of quantitative approaches and shows why measurement is of crucial importance for both theory development and improvement of practice.

Another potentially useful trend in the study of leadership is the increase in meta-analyses and systematic reviews in the field of school leadership (Hallinger, 2014, 2019; Robinson et al., 2008; Tan et al., 2020; Tian et al., 2016). These types of studies help to synthesize the evidence about findings in individual studies. For example, Tan et al. (2020) employed second-order meta-analysis to synthesize results from 12 first-order meta-analyses examining which models or practices are more influential for various outcomes. They showed that specific leadership practices such as motivating and providing professional development for teachers, can build up teacher capacity and determine the contribution of principal leadership to school effectiveness (Tan et al., 2020). They further conclude that school leaders should focus on people rather than tasks or resources. Another meta-analysis on the impact of school leadership on student achievement showed that the school leadership influence was strongest in studies that examined the indirect effects of leadership on student outcomes (Hendriks & Steen, 2012).

Special consideration should be given to the nature of the measures of school leadership. The measures of leadership nowadays are frequently designed to capture the traditional role of principals, instead of focusing on leadership practice achieved by many (Condon & Clifford, 2012; Ogawa & Bossert, 1995; Spillane et al., 2007). Moreover, the reliability and validity of these instruments are not always adequately reported. This is obvious from Condon & Clifford's (2012) synthesis of commonly used principal performance instruments. The authors identified 20 instruments of which 12 were dropped from the review due to a lack of transparency or psychometric testing. The traditionalist perspective and its influence on measurement is clearly noticeable in this review, in which eight instruments measure the degree to which principals complete different roles. Two of the most popular instruments focus on instructional practices. First, Principal Instructional Management Rating Scale (PIRMS) developed by Hallinger and Murphy (1985) determines the degree to which principals serve as instructional leaders. Second, the Vanderbilt Assessment of Leadership in Education (VAL-ED) accesses principal performance by gathering information from principals, principal's supervisors, and teachers (Porter et al., 2008).

Hallinger & Heck (2011) expand the discussion on the conceptual and methodological issues in studying leadership. They identify that the predominant approaches to study leadership are most probably incomplete because they rely on studies of one-way associations, while concurrent dominant theories of leadership increasingly describe leadership as enacted in joint practices consisting of reciprocal processes. On top of that, many studies do not account for the inherent multilevel nature of leadership data. School leadership, as a rule, involves interactions between leaders and followers. Regardless of the direction of interactions (top-down, bottom-up, or reciprocal), the hierarchy constitutes the core nature of leadership (Day & Harrison, 2007; Dyer et al., 2005; Yammarino et al., 2005). In closing, it is important to consider Thomson's (2017) suggestion that emphasizes the undervalued nature of methodological tools in the field of school leadership, management, and administration and the need for more studies engaging with methodological issues.

2.2 School leadership in the Teaching and Learning International Survey

This section introduces International Large-Scale Assessments (ILSA) by presenting a broad overview of the common ILSA features that to a large extent are also reflected in the TALIS study. The section on TALIS then presents a condensed systematic overview of the published peer review journal articles that used TALIS data to study school leadership, showing the importance of the TALIS dataset for the overall leadership research. This review supports the theoretical and methodological sections of this thesis by presenting details on how the data used in this study are utilized in other research studies to explore various research questions. The chapter also shows the development of leadership constructs from both a conceptual and a measurement perspective within the TALIS framework. By contrasting, connecting, and comparing different approaches and views on leadership over time, the chapter demonstrates how TALIS 2008 differs from TALIS 2013 and 2018, not only in terms of conceptual underpinnings, but also in terms of theoretical developments. The measures and conceptual framing of the leadership theme in TALIS partly reflect the ongoing and parallel development of broader leadership theory. In addition, the chapter points to the nontrivial interchangeable use of terms, scales, and items to address different concepts in TALIS. Finally, the chapter describes the main steps and approaches used by TALIS to estimate leadership measures and scales.

2.2.1 International large-scale assessments (ILSA)

Since the late 50s, there have been a growing number of comparative international studies in education. The objectives of ILSA are to a large extent similar across studies and typically refer to the contribution to the quality and improvement of learning and teaching by providing data and indicators on relevant educational factors at the system level (Rutkowski et al., 2014). The most prominent indicators are measures of students' achievement in specific learning domains. In addition, ILSA include measures of contextual factors, various affective-behavioural measures, and measures of other so-called noncognitive outcomes. ILSA collectively include a wide range of participating high-, middle- and low-income countries, as well as school- and outside-of-school populations. It is worth mentioning that the two main international organizations responsible for organizing and conducting ILSA are the International Association for the Evaluation of Educational Achievement (IEA), and the Organization for Economic Co-operation and Development (OECD).

Each assessment or survey that is placed under the umbrella of ILSA, has some distinct characteristics such as domains of investigation, types of instruments, participating countries, or target populations. However, the core features of ILSA are to some extent the same or at least similar. For instance, ILSA have cyclical designs with varying periods between studies, but typically 3 to 6 years. Most of them employ a multistage probability sampling design, where first a sample of schools is drawn, followed by a next step where students and/or teachers within the sampled schools are selected. Further, they typically include 20 or more countries. The design and quality monitoring procedures typically ensure that data are representative at the country or system level, allowing for the generalization of findings. Moreover, across different large-scale assessments efforts are put into developing internationally comparable measures. Heavily standardized and extensive administrative procedures for data collection and quality monitoring are implemented, and strict psychometric frameworks are employed to evaluate the quality of the instruments. TALIS certainly shares all these characteristics.

Data from ILSA are actively used by both researchers - to investigate a large number of research questions involving educational policy and practice- and policymakers - to revive interest in educational improvement and to benchmark their jurisdictions against international standards (Loeb & Byun, 2022). It is reasonable to claim that findings from ILSA has achieved a prominent position in political, professional and public dialogue on education and that the studies play an important role in evidence-based policy and outcomes-based accountability

(Addey et al., 2017; Grek, 2009; Johansson, 2016; Loeb & Byun, 2022). Moreover, the future of ILSA appears promising for people relying on those sources of information, with predictions of increased participation, coverage of areas, and increased volume and depth of studies supported by ILSA data (Hastedt & Sibberns, 2020).

2.2.2 Teaching and Learning International Survey (TALIS)

TALIS is an OECD study that primarily accesses descriptive contextual information and perceptions of the school environment, as well as affective-behavioural outcomes. It is the only large-scale survey that aims to obtain representative samples of teachers and principals at the national level. Even though studies such as TIMSS and PISA also have teacher or school questionnaires they mainly sample students and are only representative of certain groups of teachers or principals. Consequently, TALIS constitutes the largest international study oriented towards gaining more knowledge about conditions of teaching, learning environments, teachers' and principals' attitudes, school leadership, professional development and satisfaction (Ainley & Carstens, 2018; Rutkowski et al., 2013). Teachers and principals in lower secondary education are the populations of main interest in TALIS. Additionally, countries can choose to sample teachers and principals from primary or upper secondary education, and teachers and principals from schools that participated in PISA. The latter is commonly known as the PISA-TALIS link.

The first round of TALIS was conducted in 2008 and involved 24 participating countries and economies, followed by the second round in 2013 that involved 34 countries and economies (plus 4 additional in 2014 and 2015), whereas the last round of TALIS conducted in 2018 collected data across 48 countries (Ainley & Carstens, 2018; OECD, 2010; Rutkowski et al., 2013). The next round of TALIS is planned for 2024, changing the study occurrence from five to six years to match every second cycle of PISA (Schleicher et al., 2020). OECD has announced that TALIS 2024 will represent an extension in terms of scales, depth, and scope. Therefore, we can expect even more participating countries; more questionnaire items related to teachers' pedagogical practices, use of technology, professional development and special education; as well as expansion in terms of construct coverage e.g. including educational sustainability and issues related to climate change (OECD, 2022). Undoubtedly, the study is gaining in size and popularity across the world following the trend of other ILSA.

This claim has been supported in an ongoing, not yet published systematic review of the literature that uses data from TALIS[2]. Using the terms "TALIS" and "Teaching and Learning International Survey", between 2008 and June 2022, we selected 164 peer-reviewed journal articles that were published in English, of which 55[3] articles explicitly focused on educational leadership or examined leadership in relation to other variables. Specifically for the group of articles that investigated leadership, we found an increase in publications over the last years (from 5 articles in 2018 to 13 articles in 2021). In the following, some core features of the studies are briefly presented to establish the scope of the literature. Numbers in parentheses refer to the number of articles addressing the said feature.

It is interesting to note that the vast majority of articles were published in journals that come from the broader educational discipline (49), rather than journals with a specific scope on leadership. There is a group of authors with several publications who invest time to study specifically school leadership through the use of TALIS data (see for example, Bellibas & Liu, 2017; Gumus & Bellibas, 2016; Liu et al., 2021). In terms of countries included, one-third of the articles (18) analysed data from a single country. Data from Turkey were most frequently analysed in a single-country analysis (5) followed by data from the United States (3). The rest of the studies (37) included multiple countries, either in a full sample of countries (22), OECD or European countries group of countries (4), or within an otherwise defined smaller sample of countries selected based on geographic proximity such as Nordic, Asian, or based on countries representing contrasting system-level features (11). Most multi-country studies included some Asian countries (e.g. Korea, Shanghai China, Singapore) or English-speaking countries (the US, the UK, Australia).

Only a few of the studies that analysed multiple national samples tested for measurement invariance or reported the level of invariance for the TALIS scales (9). In other cases, the results were reported separately country by country (9), or on a pooled data set (10), for example, in so-called fixed effects setting (3). The majority of studies (31) applied multilevel study design with school/principal being modelled at the highest level. Latent variable modelling, regression analysis, and hierarchical linear modelling were the dominant approaches to analyse the data. In the sample of leadership studies, many constructed the scales for the purposes of their own

[2] Veletić, J., Rodriguez-Mejia A.M., Olsen R.V., Blömeke, S. (in progress). A systematic literature review of the Teaching and Learning International Survey (TALIS) peer-reviewed journal articles

[3] Appendix provides the list of articles included in the review of which most has already been cited in the main body of this extended abstract

study (22) rather than to include the scales available from TALIS (13). The remaining articles used single items in combination with both self- and TALIS-constructed scales.

There is a close balance between articles framed within instructional and distributed perspectives, respectively. Instructional leadership was investigated in 10 studies, the vast majority of which modelled leadership as an independent variable in relation to teacher professional practices such as collaboration and professional development, as well as teacher characteristics such as teacher self-efficacy and job satisfaction. The distributed perspective on leadership was frequently studied in association with teacher job satisfaction (6 studies). Among the six studies on the association between distributed leadership and teacher job satisfaction, three checked for mediating effects of teacher autonomy, professional collaboration, and teacher self-efficacy.

2.2.3 School leadership conceptualization and measures in TALIS

After demonstrating the importance of the TALIS data for the wider study of leadership, this section exhaustively describes the position of school leadership in the larger framework of TALIS. It focuses on the development and evolvement of leadership constructs and specific leadership scales over time. The purpose of Table 3 and the accompanying text are to demonstrate an apparent inconsistency in terminology and construct operationalizations across and within cycles of TALIS. Moreover, the section seeks to clearly identify terms and concepts that are used in this thesis.

Table 3

The construct of school leadership in TALIS over cycles

	TALIS 2008	**TALIS 2013**	**TALIS 2018**
School leadership position in the priority rating exercise	School leadership was placed at the second place in a priority rating exercise.	School leadership was rated as the top prioritized theme.	School leadership was rated as a top priority among OECD countries, and 4th place among all other countries.
Conceptual framework of TALIS	TALIS conceptual framework was organized around policy issues and related indicator domains. Thus,	The conceptual framework adopted for TALIS 2013 was based on a model in terms of inputs, processes, and outcomes at different	The themes were divided into two dimensions: focus (professional characteristics or pedagogical

	TALIS 2008	TALIS 2013	TALIS 2018
	school leadership was investigated under the school policies and effectiveness issues.	levels of educational actors. School leadership was placed at the school level as a process.	practices), and level (institutional or teacher). School leadership was placed at the institutional level with a focus on pedagogical practices.
Theoretical framework of school leadership	The sections of the TALIS questionnaires that relate to school leadership and management were in part framed around the Principal Instructional Management Rating Scale (PIMRS). In addition, items intended to record different forms of management were based on work by Quinn et al. (1996).	The focus is kept on instructional and administrative leadership with repeating indicators on the profile of schools, principals, and school leadership and management team from 2008 and new indicators on distributed and team leadership that were introduced from both principals and teachers' perspective.	Instructional leadership was an ongoing interest. The TALIS framework follows the evolving leadership theory with the distributional trend (teacher leadership and leadership beyond the school). Thus, the theory rests on instructional, distributed, teacher, and system leadership.
School leadership theme coverage	1) Principal background characteristics 2) Principals work distribution 3) Instructional leadership 4) Administrative leadership	1) Characteristics and distribution of principals 2) Instructional leadership 3) Distributed leadership 4) School leadership and school climate	1) Qualification, recruitment, and development of principals 2) Role, function, and actions of principal, instructional leadership, workload, autonomy, function, and action. 3) Distributed leadership and teacher leadership and teachers' perceptions of these types of leadership

	TALIS 2008	TALIS 2013	TALIS 2018
			4) Principal job satisfaction 5) System leadership and leadership in networks of schools
Scales reflecting instructional leadership	1) Management of school goals 2) Instructional management 3) Direct supervision of instruction in the school Reported by principals.	1) Instructional leadership scale Reported by principals	1) School leadership scale Reported by principals.
Scales reflecting distributed leadership	NA	Distributed leadership scale reported by principals Participation among stakeholders' scale reported by teachers.	Participation among stakeholders scale reported by both teachers and principals
Items used to measure instructional leadership (2013) /School leadership (2018) scale	Not reported here because the focus of the thesis is on the later cycles of TALIS. There were 16 items representing 3 scales that measured instructional leadership but none of the items appeared in the later cycles.	[frequency] • I took actions to support co-operation among teachers to develop new teaching practices • I took actions to ensure that teachers take responsibility for improving their teaching skills • I took actions to ensure that teachers feel responsible for their students' learning outcomes	
Items used to measure distributed leadership/ participation among stakeholders scale **[**teachers responded to all	NA	• [agreement] • *This school provides staff with opportunities to actively participate in school decisions* • *This school provides parents or guardians with opportunities to actively participate in school decisions*	

	TALIS 2008	TALIS 2013	TALIS 2018
five items in TALIS 2013 and 2018, whereas principals responded to only three items (in italic) in 2013]		• *This school provides students with opportunities to actively participate in school decisions* • This school has a culture of shared responsibility for school issues • There is a collaborative school culture which is characterized by mutual support	

Source: TALIS technical reports (2008, 2013, 2018), TALIS conceptual frameworks (2013, 2018)

School leadership has certainly been one of the key issues addressed in TALIS. In a rating exercise prior to the main survey, TALIS asks participating countries about which themes they prioritize for inclusion in the survey. This practice helps to determine the final structure and content of the study (Ainley & Carstens, 2018). Leadership has constantly been among the top priority themes for inclusion in all three cycles of TALIS as rated by participating countries. Overall, it was the top priority in 2013, and the top priority among OECD countries in 2018. With this in mind, I have to remind the reader that TALIS is not a leadership study but a larger endeavour in which equally important educational topics are studied (OECD, 2010, 2014b, 2019a). However, the coverage of themes related to school leadership in TALIS has been relatively rich and varied; from those that are more principal-oriented (principal background characteristics, principals' job satisfaction, principal's workload, or work distribution) to those more explicitly related to leadership themes (instructional and distributed leadership).

The conceptual framework of TALIS has developed from defining a selection of constructs intended to reflect a set of policy issues in 2008, to a more clearly defined theoretical model of educational inputs, educational processes, and educational outputs in 2013. In 2018, one can find the even finer conceptual mapping of themes that considers two dimensions: *focus*-based on the extent to which a theme is mainly concerned with professional characteristics or pedagogical practices; *level*-based on whether the theme mainly operates and references institutions or teachers. Overall, school leadership was placed under the umbrella of school policies and school effectiveness issues, as a process that operates at the school (institutional) level and has a focus on pedagogical practices.

In general, TALIS follows global trends and developments in the field of education, including the development and evolvement of school leadership theory and practice. According to the TALIS documentation, the conceptualization of instructional leadership, which has long

been the main construct in the overarching leadership theme, is given particular attention. The distributed perspective on leadership was first introduced in TALIS in 2013, and according to TALIS even more emphasized in 2018. Hence, in 2008, the scales were developed on the basis of Hallinger and Murphy (1985) PIRMS instrument, and the emphasis was on instructional leadership only. In the first cycle, three scales measuring various aspects of instructional leadership were available (see Table 3). All three scales were part of the principal questionnaire whereas the teachers' perspective on leadership was not captured at all.

In the consecutive cycles in 2013 and 2018, the instructional leadership scale from the principal perspective was the only indicator of this kind of leadership. Three items used to measure instructional leadership remained unchanged from 2013 to 2018. Nevertheless, the name of the scale changed from "Instructional leadership" to "School leadership". Consequently, in the published literature the renamed scale has still been used as an instructional leadership scale (see for example, Eryilmaz & Sandoval Hernandez, 2021; Holzberger & Prestele, 2021).

The scales that reflect distributed leadership were first introduced in two different forms in TALIS 2013. Principals reported the distributed leadership scale, while teachers reported the participation among stakeholders scale. Three of the five items that constitute the participation among stakeholders scale as reported by teachers were also indicators of distributed leadership scale from the principal's perspective (see Table 3 in italic). Thus, in practice, both scales have been used as indicators of distributed leadership (see for example, García Torres, 2019; Liu & Werblow, 2019). Consequently, in TALIS 2018 the distributed leadership scale was removed from the TALIS technical report with an explanation that the scale corresponds to the participation among stakeholders scale, but that the latter is a broader representation of the same construct. Therefore, in TALIS 2018, both teachers and principals responded to the items that constituted participation among stakeholders scale. However, in the published literature, the participation among stakeholders scale is still commonly referred to as the distributed leadership scale (see for example, Çoban & Atasoy, 2020; Liu et al., 2021).

Even though it is stated in the TALIS documentation that the 2018 survey includes richer indicators of school leadership in general, and school leadership from a teacher perspective in particular, I only found two additional single items. That is one additional item in the principal questionnaire that the TALIS analytical framework refers to as an indicator of "system leadership" [PQ22j- "I collaborated with principals from other schools on challenging work

tasks"], and one additional item on distributed leadership in the teacher questionnaire [TQ48h- "This school encourages staff to lead new initiatives"] (Price & Carstens, 2020). However, these items were not part of the existing or any other scales, but indicators included as single items that can be analysed separately. The teacher perspective on leadership was only reflected in a single participation among stakeholders scale as it was in TALIS 2013 (Price & Carstens, 2020).

In terms of theoretical underpinnings and scales that assess leadership, TALIS 2008 is clearly distinct from subsequent cycles. The theoretical framework in TALIS 2008 clearly referred to Hallinger and Murphy (1985) conceptualization of school leadership, and the scales were adapted from the PIRMS instrument (developed by the same authors). In contrast, the theoretical underpinnings, here understood as documented by reference to literature in the framework text, are less clear in the 2013 and 2018 surveys (TALIS 2013: Hallinger & Heck, 2010; Ogawa & Bossert, 1995; Pont et al., 2008; Spillane et al., 2004; TALIS 2018: Grubb & Flessa, 2016; Harris, 2009; Muijs, 2011). However, several documents from OECD suggest that Leadership for Learning could be an overarching framework that connects instructional, distributed, and system leadership (Ainley & Carstens, 2018; Bowers, 2020; OECD, 2016). This demonstrates the challenge researchers are confronted with when using data from TALIS to conduct secondary analyses related to substantial issues of leadership. Wang & Ahn (2023) reached a similar conclusion with the TALIS data in a parallel study on construct content validity in the school leadership literature. Specifically, the authors found, as also demonstrated above, that the same measures were used in relation to differently labelled constructs, and vice versa. Hence, they concluded that there is a lack of alignment between theoretical definitions and operationalizations in the literature using data from the study.

Other school- and principal-related indicators, such as principal demographics, or the mapping of different principal responsibilities, are quite stable and can be linked back to 2008. This does not imply that comparisons across cycles can be easily made. Firstly, the country coverage and samples differ across cycles and across populations. More specifically, the number of countries participating in TALIS is increasing over time. Moreover, in TALIS 2008 special education needs teachers were not part of the target population but they were included in TALIS 2013 and 2018. Secondly, the variables of interest change through changes in questionnaires. Finally, the overall environment in which education is happening is rapidly changing, e.g. the increased use of technology in education. The latter has likely been further boosted by the recent COVID-19 pandemic, and future TALIS surveys will have to evolve

accordingly to maintain the overall relevance of the study. On the other hand, TALIS, like all ILSA, aims to provide policymakers with indicators of changes in core educational qualities/phenomena, which suggests that substantial parts of the instruments should be kept constant over time. As hard priorities of the content in the questionnaires are made, the balance needs to be maintained so that the instruments are sensitive to changes while at the same time, they should provide indicators of change.

2.2.4 Scale scores estimation in TALIS

In the first phase of the instrument development, TALIS starts with a theoretical identification of items that could represent specific latent constructs. The choices are made based on established theory, knowledge from earlier studies, as well as prior experience and expert knowledge. The TALIS International Consortium with advice from the Questionnaire Expert Group makes the final decisions about the items that enter the main study. Typically, all ILSA will conduct a large field trial prior to the main study where data are used to evaluate the latent constructs and to modify them for the main survey.

The TALIS instruments consist of two questionnaires, administered to the teacher and principal samples, respectively. The questionnaires include items to learn about principals' and teachers' perceptions, attitudes, beliefs, and practices. Single items can be treated as a source of information in itself, or several items can be combined using specific statistical methods to create scales or indices. The procedure of combining the responses to multiple items into a single measure is called scaling. In TALIS, there are two different types of item combinations: 1) simple indices (e.g. ratios) that are calculated using simple arithmetical transformations, and 2) scale scores that are calculated using latent variable modelling within the framework of confirmatory factor analysis (CFA). This modelling approach is used to evaluate and form latent constructs that cannot be directly observed, such as the degree of leadership in schools or the degree of teacher job satisfaction.

The widely used factor model considers the relationships between observed, manifest variables (responses to questionnaire items) and latent variables (unobservable traits, constructs, such as school leadership). This is accomplished by statistically relating covariation between observed variables to latent variables in the analysis of interindividual differences. Thus, if a latent variable underlies a number of observed variables, those will "move together", and conditioning on that latent variable will cause the rest of the variation in the observed indicators to be independent (Borsboom et al., 2003).

Recent developments in TALIS include the test of measurement invariance across countries and across the International Standard Classification of Education (ISCED) levels in order to test the cross-country and cross-levels comparability of constructs, respectively. According to the TALIS technical report, all leadership scales (e.g. Instructional leadership, Distributed leadership, Participation among stakeholders) reached the metric level of invariance across countries and across levels (OECD, 2019a). This means that the unit of the scale is the same across groups and allows for comparisons of for instance unstandardized regression coefficients or covariances across groups. It should be noted that it does not allow for the comparison of latent means, for which a scalar level of invariance is needed. Hence, serious limitations are imposed on the comparison of the construct across different populations that are discussed more in-depth in the next section.

Finally, the TALIS international reports consist of descriptive statistics and estimates of the internal consistency for all constructs, the results from CFA modelling (factor loadings and model fit) and the results from measurement invariance testing. I return to the core features of confirmatory factor analysis in the method section.

2.3 Five core challenges in measuring leadership in TALIS

This section is organized around five core challenges in measuring leadership in TALIS as identified in this thesis and presented in Figure 1. The section presents and elaborates on the theoretical foundations and origin of the challenges while the findings and potential solutions are presented in the discussion section.

2.3.1 Challenge 1: Construct (under) representation

The validity of constructs included in different large-scale assessments is often discussed to be threatened by their design. The main challenge comes from the need to address the heterogeneity of populations, contexts, and languages (Wagemaker, 2020). Moreover, practical considerations, such as time limitations, knowledge of respondents, and research fragmentations place additional challenges to the whole process (Blömeke et al., 2022). Although the identification and definition of the constructs is the fundamental first step in developing instruments in ILSA (Lyons-Thomas et al., 2022), the development of the constructs should be viewed as a collaborative and largely pragmatic exercise involving all participating countries (Blömeke et al., 2022). Consequently, it is too frequently found that

constructs are represented by a small number of items, which represents a serious threat to construct validity (Messick, 1995).

Since 2013, TALIS has been concerned with two different conceptualizations of leadership, instructional and distributed leadership (see Table 3). The following subsections explain the extent to which TALIS's instructional and distributed leadership measures may not be fully representative of the constructs they are designed to measure. The issue of construct representativeness is discussed in terms of 1) theoretical definitions of leadership 2) too few questionnaire items that are used to measure leadership, and 3) inconsistent use of the items in their representation of the construct of leadership.

Instructional leadership scale

Instructional leadership in both TALIS 2013 and 2018 refers to "… supporting and developing high-quality instructional practices, developing and implementing policies that support student achievement, developing learning communities, providing feedback on instruction, modelling effective instruction, and supporting the use of assessment data." (OECD, 2019a, p.34), or to "… the actions that a principal takes to promote growth in student learning*",* (OECD, 2019a, p.35). The latter definition is taken from the Alberta Teachers' Association Magazine (Flath, 1989). While this definition may reflect some of the core features of instructional leadership (see for example, Hallinger, 2005; Hallinger & Murphy, 1985; Hoy & Hoy, 2006; Krug, 1992; Murphy, 1988), it is still unclear which of the existing instructional leadership conceptualizations as defined in the well-established theory, represent the foundation for it. Consequently, a precise definition of the instructional leadership construct within the framework of TALIS is unavailable.

When a solid theoretical foundation is lacking, the consequence may often be that the decision leading to developing scales is primarily informed by a pragmatic and empirical investigation of the data. The instructional leadership scale in TALIS is a prime example of this issue. Initially, five items were included to measure instructional leadership in TALIS 2013, but two were subsequently excluded due to poor model fit. One possible reason for poor model fit, as discussed in Article 1, might be the item wording of the three included items that all start with the phrase "I took actions to…". Based solely on the fact that the full model with five items fits poorly, two items were dropped from the scale. Consequently, instructional leadership was assessed using only three items (See Table 3 for more details). This leads to the second issue of too few items that represent the construct. Although it is obvious that those

items target actions that principals take to promote student learning through teachers and instruction, it is less obvious that such a broad and complex construct can be captured by these three items only. In contrast, Hallinger's PIRMS scale for measuring instructional leadership consists of 50 items (Hallinger & Murphy, 1985). Not only the PIRMS instrument was richer, but also TALIS 2008 itself used to have 3 different scales (including 17 items) to measure instructional leadership.

Additionally, the instructional leadership scale was renamed to the school leadership scale in TALIS 2018 without any specific explanation. Despite appearing trivial, the change is significant from both a theoretical and a secondary research perspective. As shown previously, scholars either use more general leadership conceptualizations, like educational or school leadership, or they refer to specific conceptualizations of leadership using the same scale.

Distributed leadership scale

Distributed leadership in TALIS 2013 and 2018 is defined in terms of leadership practices including "… interactions with other leaders, teachers, staff, parents, and students, rather than formal leaders' traits, roles, functions or organizational structures", (Ainley & Carstens, 2018, p.37). According to the TALIS technical report, three main aspects of distributed leadership are "…making collaborative decisions; emphasizing school governance that empowers staff and students and encourages shared accountability for student learning; and emphasizing school-wide participation in efforts to evaluate the school's academic development, (Ainley & Carstens, 2018, p.37). Thus, distributed leadership as defined in TALIS largely focuses on one aspect of the construct (shared decision-making and participation), consequently neglecting other aspects (e.g. development of people or direction setting). Moreover, the conceptual framework of TALIS involves different models/ conceptualizations of distributed leadership that partially overlap but differ in some core aspects (e.g. Hallinger & Heck, 2010; Harris, 2013; Spillane et al., 2001). Thus, TALIS proposes a definition despite the fact that the general theory of distributed leadership is not well defined, neither in terms of the major concepts that comprise leadership nor in terms of operationalizations (Lumby, 2016). Consequently, a claim that distributed leadership conceptualization in TALIS rests on a specific theoretical foundation is not fully supported.

Distributed leadership was assessed by asking principals (and teachers) to indicate how strongly they agree or disagree with five (or three) statements as applied to their school (OECD, 2019a) (See Table 3 for details). The items assess the extent to which different school

stakeholders participate in decision-making and whether the school has a culture of shared responsibilities and mutual support. Normally, distributed leadership assessments take much longer forms (Halverson et al., 2014; Porter et al., 2008) that consider not only shared decision-making but also other important aspects of distributed leadership, such as quality and distribution of support and supervision or cooperation in the leadership team (Hulpia et al., 2012).

In addition to that, the last two items that constitute the participation among stakeholders scale clearly refer to school culture, which represents a distinguishably different construct (Deal & Peterson, 1999; Hallinger & Leithwood, 1998; Van Houtte, 2005). Consequently, this scale is frequently used by researchers to assess different constructs such as school climate or school culture, or school environment in general (Liu et al., 2022; Türker & Kahraman, 2021; Wang et al., 2019). A similar conclusion has been reached in a parallel study that investigated to what extent constructs in the school leadership literature overlap using TALIS data (Wang & Ahn, 2023). Using systematic literature review and network analysis the authors demonstrated that the same items were used to measure different constructs and that different items were used to measure the same constructs in TALIS.

2.3.2 Challenge 2: Different perceptions of leadership

This challenge discusses the importance of including the relevant individuals when seeking to establish an informative measure of leadership in schools. TALIS mostly relies on principal self-reports about their own actions and perceptions. This is only partially true for distributed leadership because it also includes the teachers' perceptions. A large number of these responses are indeed perceptions requiring that respondents provide an estimation or subjective judgment of the environment or situation (Favero & Bullock, 2015). Consequently, respondents from two different groups or even within the same group might describe the same phenomenon or characteristic differently. Regardless, we frequently depend on those views to provide averages about particular characteristics or behaviour (Favero & Bullock, 2015).

There is quite a lot of evidence demonstrating how different people perceive the same phenomena. For example, the expectations about what constitutes effective leadership practice may be quite different across principals and teachers within the same school, across teachers with various experiences, or across principals and teachers from different parts of the world. The perception of teachers' autonomy across teachers and principals constitutes a good example. Pitt (2010) suggested that, although more autonomy for teachers is advocated widely,

some teachers view increased autonomy as a lack of assistance in the classroom. Frase and Sorenson (2016) supported this claim by providing an argument why teacher autonomy is sometimes perceived as a way for leadership to avoid their duties. Other examples of different perceptions of the same phenomena include Brezicha et al., (2020) work that demonstrates how principals perceive that they include teachers in decision-making while at the same teachers perceive they are excluded. Yet another example uses TALIS 2008 data to show how teachers and principals perceive instructional leadership as a single or multidimensional phenomenon (Urick & Bowers, 2017). Therefore, different perceptions are not only crucial for the complete understanding of the phenomena of interest, but they can also serve as a main study focus. For instance, in Article 2, we investigate how different perceptions of school climate can co-exist within the school environment.

When research relies solely on self-reports, the data reflects either individuals' perceptions and/or their familiarity with a particular situation. However, individual perceptions may be biased due to different reasons including their organizational position, knowledge, self-awareness, and leniency towards socially desirable reporting, to mention a few (Atwater et al., 1998; Fisher & Katz, 2000). For instance, principals are more likely to give socially desirable responses because they feel directly responsible for leadership practice in their schools, whereas teachers likely perceive themselves as less directly accountable. Moreover, the dissonance in how things are perceived may also be caused by a lack of information. Rutkowski and Rutkowski (2010) therefore argue that background questionnaires in ILSA for various reasons might be ill-fitted to the target populations giving the example of student and parent mismatched reports about the number of books at home.

Another difficulty that goes in line with the question of whom should report phenomena in schools arises when the measure is constructed from responses from one group and examined in relation to other measures also constructed from the responses of that same group (e.g. teachers report about leadership and their job satisfaction in the same questionnaire). Such results can be affected by common source bias. Common source bias arises when both variables, predictor and outcome, share the common source, e.g. questionnaire and therefore have correlated measurement errors. In other words, teachers may respond to a questionnaire in a certain way due to their affective state or organizational culture in a particular school and, consequently, we might find positive relationships when they do not exist, but also insignificant relationships when an actual relationship exists (Favero & Bullock, 2015).

2.3.3 Challenges 3 and 4: Question of comparability

In this subsection, I have combined challenges 3 and 4 because the fundamental question in both challenges is the comparability of the measures across different groups. Because there are important differences in measurement approaches when measures are compared between teachers and principals versus between countries in TALIS, I kept them as distinct challenges in Figure 1. The method section contains a more in-depth technical discussion of the distinctions between these two. However, the comparability problem draws from the same theoretical positions that have been discussed here.

The question of comparability has been one of the main goals and a recurring challenge in ILSA (Avvisati et al., 2019; Ercikan & Koh, 2005; Glassow et al., 2021; He et al., 2019; Pokropek et al., 2017; Rutkowski & Rutkowski, 2018; Scherer et al., 2016). This issue is closely linked to measurement invariance (Millsap, 2012; OECD, 2014a). Measurement invariance ensures that constructs of interest can meaningfully be compared across the groups. In general, a group can be defined in several ways, but for this thesis we focus on the cultural background (participating countries in TALIS) and organizational positions (teachers and principals within the school). The issue of comparability arises from the fact that the TALIS survey is offered in different languages and cultural contexts in which the same terms or constructs may be interpreted differently. Moreover, when teachers and principals are asked about the same phenomena in parallel items, it must be ensured that they have a joint understanding of the items, which is not self-evident given their different position and functions. To exemplify how this issue of measurement invariance is challenging, the example of including indicators using the term "feedback" is useful. Giving and receiving feedback is one of the crucial elements of leadership. However, across countries, the same amount of feedback could be perceived either as a control and monitoring function, or as a support through which principals and teachers collaborate and engage in joint endeavours. Moreover, while principals may perceive they are providing extensive quality feedback, teachers may believe the feedback is poor or not credible. Thus, this concept can hardly be translated across different cultures but also across teachers and principals. Consequently, instruments that include the term "feedback" may function differently across the groups. Therefore, rather than being actual variations in the construct of interest, the differences that are found across the groups may simply be due to items that are perceived differently.

Traditionally, measurement invariance approaches were developed in the context of comparisons between two groups, e.g. to establish whether comparisons across genders can be done. However, the complexity of testing and establishing invariance comes as a real challenge in the context of ILSA, frequently including as many as 50 countries and three educational levels. Indeed, the TALIS report from 2018 stated that only two scales reached scalar measurement invariance, that is invariance needed to compare means between all 47 educational systems. (OECD, 2019a). To add to the argument, in TALIS 2013 no scales reached scalar measurement invariance (OECD, 2014b). Although it might be unrealistic to expect that such a large number of countries are comparable across every single indicator, the results of the testing are at least to say discouraging. The conclusion that we can draw from the TALIS report is that a valid comparison of latent means across all countries is not possible (OECD, 2014b, 2019a).

2.3.4 Challenge 5: Level of analysis

The fifth and last challenge considers if leadership should be generalized at the school or country level. Moreover, it discusses whether we can at all quantitatively capture the heterogeneity of leadership across countries with the TALIS data. Although intuitively one might think that leadership is a school level phenomenon, the evidence from the broader literature is not that straightforward.

Much of the qualitative evidence on school leadership emphasizes the importance of the context for leadership practice (Møller & Schratz, 2009; Vedøy & Moller, 2007). Johnson et al., (2008) analysed the differences in the Successful School Principalship Project (ISSPP) across the contexts of Norway, the USA, and China and concluded that "…the theory and practice in educational leadership and management are socially constructed and more contextually bound than some are prepared to admit" (p. 419, Johnson et al., 2008). The same authors pointed to the importance of a political and cultural heritage as well as a common language. Because a lot of the literature, as well as instruments that are constructed to measure leadership, come from the English-speaking context, the important information and nuances are simply lost in translation. The examples include not a such easy translation of the English words accountability, benchmarking or management to a Scandinavian or Chinese language (Johnson et al., 2008). Moreover, Printy & Liu (2021) highlighted the importance of the context for leadership in the discussion on levels of "control" over school including the local school-, national-, state-, regional-, and municipal level, and that "the amount of influence each level

can exert varies by policy" (Printly & Liu, 2021, p.318). A good indicator of how much control or authority the school or principal has over how the school is operated is the degree of decentralization. Schools that operate in highly centralized systems might not have much autonomy or power in decision-making which consequently influences their approach to leadership in schools. Moreover, the degree of autonomy may vary with respect to different functions, e.g. principals might not have that much autonomy with respect to finance and stuffing in contrast to instruction (Bush, 2016).

The fact that much of the literature and debate about school leadership has its roots in Western-oriented studies, which distorts the global view of leadership as a contextually dependent phenomenon, is the other problem that served as the basis for this challenge. Although there is a better understanding of the significance of context for leadership today, there is still a lack of international research to support this claim.

3. Methodology and methodological considerations

This chapter describes the methods and methodological choices applied in the three empirical articles. The chapter is organised around four main sections 1) data and sample, 2) measures, 3) analytical approach, and 4) ethical considerations. As all three articles are based on secondary analyses of TALIS data, the sample and its variation across the articles are presented and discussed first. The measures section compares and contrasts different approaches taken to operationalize leadership across the three articles and gives a brief presentation of the teacher job satisfaction scale that is used in Article 1. Additionally, the section on analysis summarizes the advocated approach to how leadership should be measured and justifies the separate methodological choices across articles. No further presentation of cluster analysis is given since this approach is fully accounted for in Article 3. Lastly, some ethical considerations associated with the secondary analysis of TALIS data are described.

3.1 Data and sample

Because this thesis investigates leadership in the international context, the TALIS dataset was a logical choice representing the most comprehensive source of data on school level phenomena (including more than 20 broader themes), across 48 countries (last cycle), and involving both teachers' and principals' perspectives. The data used in this thesis are from the two most recent cycles, TALIS 2013 and TALIS 2018, the most recent data available at the time when the research was conducted.

In Article 2 and Article 3 that analyse TALIS 2018 data, the intention was to use complete datasets (all countries included). However, the datasets were reduced due to missing data. Article 2 included only countries leaving out provinces and cities. In addition to that, countries that did not administer some of the key variables were excluded. In Article 3, data from 42 countries were included, while 6 countries were excluded due to a large number of missing data on key items. Article 1 is of a different nature, restricting the sample to the more homogeneous group of Nordic countries. The selection was motivated by the fact that context might influence leadership practice and that comparisons across completely different countries may be unreasonable. Because the main goal of the analysis in Article 1 was to develop and demonstrate the feasibility of a principally alternative way to measure instructional leadership using teacher reports, a limited set of more homogeneous countries was a vital first step. The main reason for this selection was to include countries with which the authors were familiar.

3.2 Measures

3.2.1 School leadership

School leadership is the main variable investigated in all three articles. However, different approaches are used to represent these constructs across the different analyses. This demonstrates the flexibility and strength of the TALIS data, which enables the secondary analysis of the same construct (e.g. leadership) from a variety of perspectives using a variety of methodologies by individual researchers and countries.

In Article 1, the teacher data are used in a multilevel setting to construct two dimensions of instructional leadership (managing the instructional program and developing the school learning climate) at the level of school. These two measures were evaluated and discussed in terms of measurement properties. As a validation step, these two measures were inspected in relation to the TALIS instructional leadership scale. Moreover, the measures were associated with the teacher job satisfaction scale that served as an external criterion to concurrent validity. Thus, Article 1 uses the instructional leadership scale already available from the TALIS documentation, but also constructs and suggests the usefulness of two new measures of specific dimensions of instructional leadership from the teachers' perspective.

Article 2 uses the instructional and distributed leadership scales reported by principals already available in TALIS as predictors of the school climate as perceived by teachers and principals. On the other hand, the measure of school climate was constructed by a set of items included in both the teacher and principal questionnaires. The scale is constructed to be comparable across teachers and principals and to capture important aspects of the leadership for learning framework. In a similar fashion to Article 1, Article 2 uses teacher data to measure school climate at the school level and combines this data with principal reports.

In Article 3, a holistic and descriptive approach to leadership is taken, using as much data as possible from both teachers and principals to explore leadership for learning at the school level in a comprehensive manner. In addition to instructional and distributed scales from TALIS, the article made use of various available scales that are not originally described as indicators of school leadership. The article provides an argument for why these scales (e.g., academic pressure, team innovativeness, stakeholder involvement) are indicators capturing dimensions of leadership for learning. In total, six scales from the principal questionnaire and

six scales from the teacher questionnaire were analysed as indicators of leadership for learning at the level of school.

3.2.2 Teacher job satisfaction

To evaluate the criterion validity of two proposed dimensions of instructional leadership, Article 1 estimates the correlation between them and teacher job satisfaction. Previous research led us to hypothesise that any measure of instructional leadership should be positively correlated with teacher job satisfaction. Moreover, the correlation between the newly proposed measures and the existing measure of instructional leadership in TALIS were estimated, also hypothesized to be moderately positive. The measure of teacher job satisfaction used in Article 1 was adapted from the TALIS *Teacher job satisfaction with current work environment* scale. The main modifications included using three instead of four items and evaluating and interpreting the measure at the level of school rather than at the teacher level. Both changes were justified with the fact that measures should represent shared cluster construct (to be returned to below). Therefore, as with leadership, we focused on the satisfaction with the work environment which should refer to a phenomenon shared by all group members.

3.3 Analytical approach

The general approach in this thesis regarding the methodological choices is based on the intention to keep the analyses as close as possible to the TALIS framework. Accordingly, the methods of choice are closely related to the latent variable modelling framework also used to establish, evaluate, and report the scales in the official TALIS reports. Thus, the methodological considerations presented and discussed below are seen from the lenses of confirmatory factor analysis (CFA), multigroup CFA (MGCFA), and structural equation modelling (SEM). Moreover, the additional level of complexity is added by conducting these analyses in a multilevel setting that has been routinely applied to analyse data from ILSA. Because both Article 1 and Article 2 employed a so-called saturated model on level 1 (teacher level) in a multilevel setting (teachers nested in schools), a separate subsection is dedicated to justifying this choice. Furthermore, the approach used to evaluate measurement invariance is reviewed in a separate subsection.

3.3.1 Multilevel modelling

Multilevel modelling has been the method of choice when data with a hierarchical (clustered) structure are analysed (Dyer et al., 2005; Hox et al., 2017; Snijders & Bosker, 1999).

Research in education mostly explores such data, because individuals are naturally nested in groups. For instance, pupils are nested in classrooms, classrooms are nested in schools, schools in regions or countries and so on. Multilevel modelling has numerous applications, e.g. in survey data that are not simple random samples, in repeated measures models when individuals are assessed at a number of fixed time points, in studies of twins, families or other naturally occurring clusters of individuals, or in meta-analyses (Leeuw & Meijer, 2008). International large-scale assessments consist of data collected in surveys with complex designs. TALIS is based on a two-stage sampling design with unequal selection probability at different stages (Meinck & Vandenplas, 2021). In the first stage of sampling in TALIS, schools are selected, and then a sample of teachers is drawn within the selected schools. This was done to ensure that all subgroups are represented in the sample. Consequently, teachers within individual schools share some common characteristics with other teachers in the same school, and some distinct characteristics with teachers from other schools. This further means that the observations at the individual teacher level are not independent (those who work at the same school tend to share some commonalities). Failing to account for these dependencies violates key assumptions of standard statistical analysis, such as regression, which assumes that all the respondents in a sample are independent units, resulting in variances being generally underestimated. Therefore, when confronted with such clustered samples, multilevel modelling is the method of choice allowing to directly model school level effects.

When using ILSA data, the resampling method (using replicate weights) is another technique for accurately estimating standard errors (Jakubowski & Gajderowicz, 2022). In practice, there will be smaller or larger differences between estimates of standard error obtained from these two approaches depending on the country and sample that is being analysed (Jakubowski & Gajderowicz, 2022). However, my research question, which operationalizes leadership as a group level phenomenon at the school level, cannot be addressed by methods using replicate weights. I thus employed multilevel models. Regardless, the differences between the two approaches should not matter significantly as shown by Jakubowski & Gajderowicz (2022).

3.3.2 Latent variable modelling within the multilevel framework

The strength of multilevel modelling is the possibility to address how much of the variance lies between individuals (e.g. teachers), and groups (e.g. between schools or/and countries) and to examine this variability at each of the levels. Thus, research questions are

addressed at the individual level, group level, or across levels. Although multilevel modelling has been developed for single score scales or single indicators, recent developments allow for integrating multilevel modelling with factor analysis. Simply stated, factor analysis and regression are united into a single SEM framework. Thus, SEM facilitates latent variables, multiple indicators, measurement errors, and complex structural relationships (Kline, 2015). In SEM one can distinguish between the measurement part of the model (the relationship between observed indicators and latent variables, e.g. CFA), and the structural part of the model (the relationships between constructs e.g. regressions). This thesis makes use of CFA, in order to incorporate multiple indicators of a single unobserved construct that causes the covariance between the indicators, while simultaneously controlling for measurement error. Finally, by using multilevel SEM, we control not only for measurement error at the individual and group level, but also for sampling error as in traditional multilevel approaches.

In the multilevel CFA framework, the total sample covariance matrix is decomposed into pooled within- and between-group covariance matrices, and the factor structure is evaluated at each of the levels. Accordingly, one can evaluate numerous alternatives regarding the factor structure at different levels. For instance, it is possible to model 1) the same number of factors at each level with equal factor loadings, 2) the same factor structure at each of the levels, but the loadings can be different, or 3) different factor structures at the different levels, to mention a few (Kim et al., 2016). This ensures that construct development and validation reflect both empirical evidence and theory.

Moreover, a multilevel CFA framework allows us to examine the substantive meaning of the constructs at different levels, especially when data are collected at the individual level (e.g. teacher data in this thesis) and analysed at the group level (e.g. schools in this thesis). Thus, against this background, I am interested in modelling group level behaviour and leadership practice, combining data collected at the individual teacher level with data collected at the school/principal level. Stapleton et al. (2016) distinguish between two different group level constructs- configural and shared. Configural constructs represent aggregates of individual constructs and are just a reflection of the construct at the individual level (e.g. averages at the group level), equivalent to what Marsh et al. (2012) label as contextual variables, whereas shared constructs represent a characteristic of the cluster itself, in which individuals within the cluster are expected to respond in a similar way, equivalent to what Marsh et al., (2012) label as climate variables.

3.3.3 Construct meaning at the group (school) level

I used multilevel modelling in this thesis with teacher data representing level 1 and school/principal data representing level 2 or group level. According to Marsh et al. (2012) constructs at the group level can be true group level constructs (e.g. number of computers in school) or aggregates of responses from the lower teacher level. The latter can be further subdivided into 1) aggregates that refer to the individual characteristics of the lower-level units (e.g. the average years of experience of teachers in the school) and aggregates that refer to some group characteristic (e.g. when teachers rate some aspects of the school, such as school climate). Thus, aggregates of responses from the lower level refer to contextual and climate variables, respectively. The key difference between climate and context variables is the *referent in the individual teacher level responses*.

Contextual variables refer to the individual teacher and the construct at the group level is just an aggregation of individual teacher characteristics. Contextual variables (configural constructs in Stapleton et al. terms) do not assume that the individual responses are interchangeable nor that individuals within the cluster respond in a similar way. Therefore, the factor structure is modelled at both levels, individual and group. Moreover, in configural cluster constructs, the cluster is not viewed as the source of variability of an individual construct and therefore metric cross-level measurement invariance is required.

On the contrary, climate variables (shared cluster constructs) refer to the school environment (rather than the teachers themselves) allowing each teacher within the group to rate the same group characteristics (e.g. leadership). The main characteristic of such responses is that they target qualities or characteristics that are shared among group members (e.g. this school has a culture of shared responsibility for school issues). Thus, the ratings of the teachers that belong to the same group should be interchangeable and their agreement should be strong. Stapleton et al. (2016) further expand the discussion by demonstrating how such responses collected at the individual level *only make sense* to be modelled and interpreted at the group level. To put it another way, because the measure is isomorphic across individuals and the responses within the school should be highly correlated, it makes no sense to model variability at the individual level. Accordingly, the individual (e.g. teacher) level, could be represented well by a so-called saturated model, where all items are allowed to freely correlate with each other without imposing a factor structure, while the factor structure is modelled only at the group level. Consequently, the saturated model is able to explain the remaining variability at

the individual level without making any assumptions about the factor structure. This strategy offers a practical approach to measure and interpret leadership when data are gathered from teachers because, as stated in the theory part, leadership is viewed as a group level characteristic.

Accordingly, the formulations of the questionnaire items are of crucial importance to differentiate between constructs as individual or shared. In this thesis, special care was given to the identification of items reported by teachers that could represent shared characteristics of the schools. In Article 1 teacher responses used to measure two dimensions of instructional leadership at the level of school, clearly refer to the school level. For instance, the item stem refers to the school (e.g. "In this school who uses the following methods to provide feedback to you", and "How strongly do you agree or disagree with the following statements about this school"). Similarly, in Article 2 teacher responses are used to measure school climate and the item stem clearly refers to the school (e.g. "How strongly do you agree or disagree with the following statements, as applied to this school).

3.4 Measurement invariance

Measurement invariance refers to the property that an instrument (e.g. questionnaire) should function equally across a range of conditions (e.g. language, culture) that are considered irrelevant to the construct of interest (e.g. school leadership) (Millsap, 2012). Establishing measurement invariance is a precondition for comparisons across groups (Byrne & Vijver, 2010; Millsap, 2012). In the framework of international large studies, the issues of comparability of constructs represent one of the biggest challenges, that is *if and at which level* data from different countries can be validly compared. In this thesis, in addition to this basic question, we have examined if data across teachers and principals can be meaningfully compared. Although the logic is the same, additional complexity is added by the fact that teachers are nested in principals. Or in other words, data come from different sources and different levels of analysis. Therefore, the standard approach to test for measurement invariance is not possible. Consequently, we used multilevel CFA to set the model and tested for invariance between teachers and principals at the school level. Since the whole idea was to explore possibilities to compare reports that reflect characteristics of the group, this choice was justified.

Theoretically, data from different groups are comparable at three levels: configural, metric, and scalar and we frequently refer to them as levels of measurement invariance (Van

de Vijver et al., 2019). Configural level of invariance means that items used to measure the construct adequately represent the construct in different groups. This further means that the construct is equally understood and has the same indicators across the groups, but the relation between indicators and the construct is not guaranteed (He et al., 2022). This, however, does not ensure the basis for any comparisons. Metric level of invariance means that indicators that measure the construct (items) are equally related to the construct across the groups, or in other words, that the factor loadings are equal across the groups. If the metric level of invariance is achieved, correlations and unstandardized factor loadings can be compared across groups. Finally, the scalar level of invariance means that the intercepts and factor loadings are the same across groups. Therefore, the scale (construct) has the same unit of measurement and a joint origin across groups which implies that valid comparisons of latent means can be made. However, this is rarely achieved in the context of ILSA due to the high number of groups being compared (e.g. in TALIS 2018, 47 countries) (He et al., 2022; Rutkowski & Svetina, 2014; Van de Vijver et al., 2019; Zieger et al., 2019). Sources of incomparability come from three types of cultural bias: a) construct bias, b) method bias, and c) item bias (differential item functioning) (He et al., 2022). Construct bias occurs when a construct has a different meaning across groups (e.g. what represents or defines effective school leadership may be very differently constructed across the world, but also across teachers and principals, due to different roles they have with respect to effective leadership). Method bias occurs when sampling, administration of the test (questionnaire) and instruments are different across groups. Instruments can be different across cultures, but also across teachers and principals, due to different response styles, item keying, or familiarity with the item stimuli. For example, principals may be more prone to give socially desirable answers because they feel more responsible for leadership in schools. Item bias occurs when a specific item has a different meaning across cultures (e.g. when teachers that are equally satisfied with their jobs, but coming from different countries, are not equally likely to endorse the same item).

As discussed above (Section 2.3.3.), the exact measurement invariance is not only hardly achievable, but also not reasonable to expect in the context of 47 educational systems. Therefore, several alternative approaches have been suggested to allow to compare the constructs across groups despite the lack of exact invariance. One of the approaches is to test for partial measurement invariance (Byrne et al., 1989; Steenkamp & Baumgartner, 1998), restricting the parameters of at least two indicators to be equal across groups. Both, full and partial invariance are considered as *exact* approaches to measurement invariance because the

parameters (all or some) are required to be *exactly equal* across groups (Cieciuch et al., 2019). The partial measurement invariance approach has been criticized for being insufficient (Steinmetz, 2018), however in a recent simulation study Pokropek et al., (2019) showed that this approach may effectively be used for the comparison of latent means. Another and conceptually different approach is the approximate invariance approach where the alignment method, Bayesian SEM, and mixture models as separate techniques are used to achieve approximate invariance (De Roover et al., 2022; Lek et al., 2018; Muthén & Asparouhov, 2018). In this thesis, the partial measurement invariance approach is used.

3.5 Ethical considerations

The data used in this thesis comes from the TALIS survey that is administered by OECD. In terms of data confidentiality, the data have already been anonymized and made freely and publicly available. In selecting schools and teachers for participation in TALIS, it is in theory possible to indirectly identify individuals by a limited number of researchers that work as a part of TALIS national teams. However, this information is only available for a short period of time and for the purpose of sampling and administration of the survey. After data are collected and coded, one cannot identify schools or teachers that participated in the survey. There are countries that sampled the whole population of schools, for example, Iceland, but the data availability for such populations is restricted. Therefore, data from Iceland were not used in this thesis. Consequently, formal approval from the Norwegian Centre for Research Data (NSD) was not required.

4. Summary of the results across the articles

This chapter presents separate summaries of the three articles included in the thesis. All three articles are written in collaboration with other researchers.

4.1 Article 1: Developing a shared cluster construct of instructional leadership in TALIS

In Article 1, the construct representation of the instructional leadership measure provided by TALIS 2013 is examined alongside the possibility to measure leadership using teacher collective reports. We hypothesized that it has been extremely unlikely that a complex construct such as instructional leadership can be captured by just three items, given that similar instruments that are used to measure leadership usually include much larger sets of items. As a starting point, we examined the association between the existing instructional leadership measure in TALIS and teacher job satisfaction. Contrary to well-established evidence in the educational literature, we found no significant associations between the two. Therefore, we proposed to use teacher data to capture two dimensions of instructional leadership at the school level. This approach allowed us to 1) keep the operationalization of instructional leadership close to the well-established instructional leadership model proposed by Hallinger and Murphy (Hallinger & Murphy, 1985), 2) model leadership as a shared group characteristic (Marsh et al., 2012; Stapleton et al., 2016), and 3) to deal with bias caused by principal self-reports (Podsakoff & Organ, 1986). To do so, we used data from TALIS 2013 from Norway, Sweden, Finland, and Denmark and we have answered the following research questions (RQ):

1. To what degree are principals' perceptions of instructional leadership, as measured by OECD in TALIS 2013, associated with teacher job satisfaction with the 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 instructional leadership as measured by TALIS?

The findings showed that in most countries, the two newly proposed measures of instructional leadership performed well. In particular, the measure reflecting the school learning climate had a superior fit. We found that the new measures had a moderately positive correlation with the old TALIS measure showing that both capture some overlapping characteristics of instructional leadership. Additionally, we were able to show that the newly proposed measures had a positive relationship with an outside criterion (teacher satisfaction with their current work environment) in contrast to the original measure. Even though the structural models did not meet the commonly accepted guidelines for evaluating fit, we found these results to be encouraging given that items that are used as indicators of leadership were originally included with the intention to measure other aspects of the school environment. The findings suggest that asking teachers about observable behaviours that are present in their environment is a promising approach to measuring leadership. Moreover, the questionnaire items should closely be related to the underlying theory of leadership and should target group characteristics. From a substantive standpoint, we discovered that stronger instructional leadership is positively related to teacher job satisfaction with the school environment, particularly in the school climate dimension of leadership. This finding was consistent across the four included countries.

4.2 Article 2: Teachers' and principals' perceptions of school climate: The role of leadership style in organizational quality

Article 2 was motivated by the fact that both teachers and principals are responsible for creating leadership in schools to support learning. The article establishes a theoretical framework which is used to demonstrate that only a small fragment of leadership is solely in the hands of the principal. Therefore, we have identified aspects of leadership where both teachers and principals are responsible and identify them as core elements subsumed under the umbrella concept of school climate. Moreover, we examined how such aspects of the school climate are perceived by teachers and principals. We did not discount that different perceptions of the same phenomena can coexist in schools where the climate is good, but to be able to understand this question further we first aimed to establish a comparable measure of school climate across teachers and principals. Moreover, we examined how different perceptions of school climate are associated with leadership style. The analyses were conducted across 37 countries that participated in TALIS 2018. The guiding research questions were:

1. What are the measurement properties of the proposed school climate indicators

based on teacher and principal reports?

2. Based on the newly proposed measures, to what extent do teachers' and principals' views of school climate differ?
3. To what extent is leadership style associated with school climate as perceived by principals?
4. To what extent is leadership style associated with school climate as perceived by teachers?
5. To what extent do the features of the national context associate with the teachers' and principals' perceptions of school climate?

Overall, the results indicate that teachers and principals consistently rate their school climate differently. In a majority of the countries, principals rated school climate as better than teachers. This is consistent with our argument in Article 1 that principals' self-reports may not be the most reliable source when researching leadership. However, we did find that teachers' and principals' ratings of school climate are moderately positively correlated in most countries. In other words, even if principals rate the school climate consistently as better than teachers, the reports from groups of teachers across schools are relatively in agreement with their principals. We further found that stronger instructional and distributed leadership are positively correlated with principals' and teachers' perceptions of school climate in the entire sample. However, within-country analyses revealed that distributed leadership is more strongly and consistently related to perceptions of school climate in schools and this relationship is more stable in the principals' sample. In as many as 30 countries principals' perceived school climate was positively moderately associated with distributed leadership, while this is true for 16 countries in the teacher sample. On the contrary, instructional leadership is only found to be significantly associated with teachers' perceptions of school climate in 4 countries. Additionally, around 10% of the variance in the principals' perceived school climate was accounted by country, as shown in the country fixed effects model. This finding points to the importance of cultural norms and their role in what principals considered as a good climate. Overall, the results indicate that teachers and principals differently perceive school climate, and that leadership style primarily predicts principals' perceptions. How teachers perceive school climate relates to distributed leadership style while instructional leadership seems to be of no importance.

4.3 Article 3: Exploring school leadership profiles across the world: a cluster analysis approach to TALIS 2018

By using cluster analysis this study intended to provide an exploratory and descriptive summary of the leadership measures from TALIS 2018 that were derived from both the teacher and principal questionnaires. By doing so, the article acknowledged that leadership for learning is a shared responsibility between teachers and principals and that some leadership functions, as represented in the LFL framework, can only be fully realized through collaboration. The study was guided by three research questions, which are listed as follows

1. To what extent can *countries* across the world be classified into groups based on leadership for learning practice as reported by teachers and principals?
2. To what extent can *schools* across the world be classified into groups based on leadership for learning practice as reported by teachers and principals?
3. How is group membership of schools associated with the demographic characteristics of schools and principals?

The unconditional three-level (teacher data), and two-level (principal data) models revealed that the variability at the county level was limited for most of the variables included in the analysis. At the school level, on the contrary, we found a significant and substantial amount of variability. Consequently, at the level of school we have identified five clusters. Cluster 1 was characterized by weak leadership for learning at all levels. Cluster 2 was distinguished mostly by neutral leadership for learning with great emphasis on school autonomy and teacher professional development. Cluster 3 represented balanced leadership for learning practice with all indicators moderately represented. Cluster 4 was characterized by strong leadership for learning at all levels. Finally, Cluster 5 was characterized by leadership for learning practice that is oriented towards the dimensions: instruction, curriculum, and assessment. The relative frequencies of clusters within countries did not reflect any easily identifiable patterns with respect to the country's geographical position, language, or culture. Moreover, most countries are dominated by many of these clusters rather than just one or two of them. Both these findings, the small variance component at the country level and the lack of country-unique patterns of clustering, are contrary to the literature suggesting that leadership is dependent on wider societal norms in a country or region. Moreover, we did not find that school or principal demographics can account for the identified clusters. Certainly, these findings need to be interpreted in line with data and method limitations. Overall, this study

illustrated a shortage of quantitative evidence regarding leadership that is influenced by culture. Instead, we only provided evidence in favour of the theory that leadership is a school level phenomenon.

5. Discussion and implications

In this last chapter, I engage in a reflective and analytical discussion by comparing, contrasting, and summarizing the findings presented in the three articles. Throughout this discussion, I endeavour to maintain a clear focus on the overarching aim of the thesis, which is to address five core challenges in measuring leadership internationally. This thesis accomplishes a delicate balance between the theory of leadership and complex methodology, which itself makes it unique. I could easily say that methodological choices form the cornerstone of school leadership research. But before that, it is important to present overall contributions and implications thoroughly examined and discussed through the lenses of the five core challenges. Therefore, I first demonstrate this synergy of leadership theory and methodology, by presenting some of the core problems and assumptions that people make when measuring leadership that hinder the search for substantive evidence in the overall study of leadership. Then, while maintaining the five-core challenge structure, I review the major conclusions and how they relate to the leadership theory and the TALIS research.

5.1 Theory and methods: Inseparable elements in the study of leadership

Through my years of reading, researching, and studying school leadership, I have come across two substantial research questions that stand out as particularly important. The first one is about *what* practices constitute effective or successful leadership and *if and how* those practices vary internationally. The second one is about *how* school leadership *influences* and/or associates with school effectiveness.

Most scholars, including myself, would agree that leadership plays an important role in ensuring the effectiveness of school, both in terms of common sense and practical reasoning. Some scholars even claim that leadership is "second only to classroom teaching as an influence on people learning" (p.28, Leithwood et al., 2008). This claim, although widely cited in the leadership literature, remains controversial due to the lack of empirical evidence to support it. In fact, the authors themselves revisited the claim years later to state that "School leadership has a significant effect on features of the school organization which positively influences the quality of teaching and learning. While moderate in size, this leadership effect is vital to the success of most school improvement efforts" (p.6, Leithwood et al., 2020). The revisited claim is likely closer to the truth, however it remains difficult to quantitatively capture and

demonstrate this role of leadership in the larger framework of school effectiveness research (Hallinger & Heck, 1998; Hendriks & Steen, 2012; Leithwood et al., 2004).

Consequently, we focus more on measurement and elaborate on how alternative approaches to measuring leadership may lead to new insights in the broader leadership field. Some of the well-known issues in measuring leadership effects include the fact that leadership impact on student learning is typically indirect (Bellibas & Liu, 2018; Burkhauser, 2017; Gumus et al., 2013; Liu et al., 2021; Sims, 2019). Another issue arises from the fact that school leadership is typically viewed through the perspective of principals (see for example, Bellibas & Liu, 2017; Blase & Blase, 1999; Gumus & Bellibas, 2016; Ham & Kim, 2015; Lambersky, 2016). Yet another difficulty arises from the lack of empirical support about how leadership is culturally embedded and from the lack of evidence that would support the comparison of leadership practice across the world (Gurr, 2014; Hallinger, 2018; Hallinger & Leithwood, 1998).

Therefore, the empirical studies presented in this thesis, by examining different substantive research questions, highlight the critical importance of measuring leadership effectively. Each of the articles faces and confronts one or more challenges of measurement in this context. It is easy to notice that those measurement challenges are strongly related to the theory of leadership. I have shown that the inability to measure something properly or effectively creates gaps in theoretical knowledge while theoretical unclarities create difficulties to soundly approach measurement. By focusing on challenges and potential solutions, this thesis demonstrates that measurement concerns and theoretical issues are almost inseparable. So, the advancements in either area or more likely, through collaborative efforts between the two, are likely needed to be able to go forward with studying how leadership fosters or hinders change in a desired direction.

The main conclusions are discussed through the lenses of empirical, methodological, and theoretical contributions in what follows in the next subsections.

5.2 The complexity of the leadership construct leaves it difficult to fully represent all its elements

Throughout I have demonstrated how the measures of school leadership that are available in TALIS represent a serious threat to validity with respect to construct representation. Both the instructional leadership scale and the distributed leadership scale suffer from the issue of

too few items, a weak theoretical foundation, and inconsistent use of indicators in relation to different latent constructs.

To address issues of narrow leadership operationalization in TALIS, this thesis explores a variety of alternative approaches to measure leadership with the same data. Consequently, in Article 1, we operationalize two dimensions of instructional leadership at the school level using teacher reports. In Article 2, we examine school leadership through the lenses of shared school climate, using reports from both teachers and principals and focusing on the aspects of the school environment for which they both are jointly responsible. Similarly, in Article 3, school leadership is explored and described using teacher and principal reported scales that are originally created to measure other school environmental factors, demonstrating that the overlap between TALIS and the leadership for learning framework is only partial. While this diversity of approaches demonstrates the flexibility of the TALIS data, it also shows the complexity and broadness of leadership constructs. Even though the articles in this thesis exhausted leadership indicators from TALIS, the construct most likely is not completely represented in any of them. These examples clearly demonstrate the need for richer indicators of school leadership to fully unleash the promise of TALIS to support both research and evidence-based evaluation of policy.

This thesis further shows how scales available in TALIS lack documentation on how and to what extent they are based on a theoretical foundation. While the 2008 study had a clear focus on instructional leadership with as many as 17 indicators, the latest cycles of TALIS made a shift towards a more distributed perspective of leadership by 1) reducing the instructional leadership indicators 2) giving more space to distributed leadership, and 3) including teachers' perspective. By utilizing what I refer to as the emerging leadership perspective in the introduction, these changes broadly align with the development of leadership theory. Some of these changes are indeed helpful and much needed, for example, the inclusion of teacher perspective on leadership. On the contrary, other changes are unclear and not adequately supported. For instance, TALIS 2013 dropped many indicators that reflected instructional leadership in the previous cycle but claimed that the study continues to focus on this perspective. It is fair to say that these changes probably reflect the limited capacity of TALIS in terms of the number of constructs included, primarily because the questionnaire must be shortened to make it possible for respondents to find time to participate. As a result, a trade-off is made by removing some of the instructional leadership indicators from previous cycles and adding some new indicators of distributed leadership.

Moreover, the instructional leadership scale was renamed to the school leadership scale in TALIS 2018. For those who use TALIS data to analyse leadership, such changes in labels pose serious challenges. It is a concern that different indicators are used to study the same construct and that the same indicators are used to measure different constructs. Wang & Ahn (2023) came to the same conclusion in a parallel study that has been published only recently.

While acknowledging the efforts made by international organizations that conduct large scale studies, I draw attention to the need of being precise and clear about what is being measured and that what we measure must have some basis in theory. Nevertheless, TALIS and other ILSA are transparent about their approach, which is pragmatic rather than hypothetical-deductive, to allow for a variety of perspectives and to balance the needs of diverse participating countries. Consequently, the data can be approached from a cautiously flexible perspective to identify how secondary analysis can be based on different theoretical positions. As previously mentioned, this comes at the cost of possibly unclear theoretical foundations and poor construct representations.

As an example, I have based this thesis on the grounds of leadership for learning framework (Hallinger, 2011; Kelley & Halverson, 2012; MacBeath, 2019). The strength of this model lies in the fact that it includes numerous actors of leadership. This theory also emphasizes that leadership is almost never exclusively instructional or distributed, but rather consists of a combination of these (and many more styles) depending on the context where it is executed and situational needs. This conceptualization further allows me to study leadership as a group characteristic or a feature of the school environment. Therefore, in Article 1 and Article 2, I present school leadership as a quality of the group and an attribute of the school environment. In Article 3, I jointly analyse data from teachers and principals acknowledging the role of both and their joint efforts for successful leadership. I find support for the claim that successful school leadership can only be achieved in the collaborative efforts of various school stakeholders as shown in Article 2 and in the previous literature (see for example, Harris, 2004; Leithwood et al., 2020; Spillane et al., 2004).

Overall, the thesis suggests that the operationalization of leadership is critical and that further studies, including TALIS, should be very careful and precise in describing how the measures operationalise specific theoretical concepts. Therefore, if the scales represent the construct only partially or from a more narrowly defined theoretical perspective, this should be clearly described and stated in the theoretical framework and various reports from the study.

Moreover, this thesis suggests that a trade-off between the quality and quantity of available indicators is one strategy to deal with this issue. In other words, the thesis suggests that TALIS might choose to decrease the number of constructs of interest, while enriching the scales that are of greater importance. Consequently, the time burden would not be critical, the indicators that are in the study would be more comprehensive and tightly connected to what matters the most, not only for secondary research with TALIS data but also for participating countries. I am aware that it is challenging to design questionnaires that include information perceived to be equally relevant for all countries, but I would strongly support the position that this is a case where the adage "less is more" is true. Alternatively, TALIS could make use of the well-known rotated questionnaire design similar to those implemented in other ILSA, e.g. in PISA. This would allow for increasing content coverage in terms of both the number of constructs and level of representation for each of the constructs, without putting additional time burden on teachers. Furthermore, attention should be drawn to the inconsistent use of terminology across cycles of TALIS, but also within one cycle of TALIS.

5.3 It matters whom we question about leadership in schools

The second challenge previously identified requires us to come up with arguments that explain why it is useful to gather data from different sources to measure leadership. At the same time, this challenge asks us to identify the limitations associated with either collecting evidence about leadership from one source only or from multiple sources (teachers and principals) who likely would report differently when asked about the same phenomenon. In Article 1, we demonstrate that leadership as measured from teacher reports and leadership as measured from principal reports, yield different results when examined in association with teacher job satisfaction. In Article 2, we closely investigate the issue of different perceptions of school climate between two groups and conclude that we should talk about the strength of the (dis)agreement rather than the presence or absence of it. In Article 3, we further confirm those perceptual differences between teachers and principals that relate to distributed leadership scale.

The thesis highlights the concern that relying strongly on principal self-reports when assessing leadership may bias our insights because principal self-reports often overestimate the positive aspects of the work environment (Brezicha et al., 2020; Park & Ham, 2016). It is important to note that this bias may not be intentional, but rather is a result of the role and position that principals hold that makes them feel primarily accountable for the overall

leadership in the school. Teacher collective reports, on the other hand, can provide more robust and stable insights about the school environment. However, as demonstrated in Article 2, there are aspects of school leadership that principals solely oversee. Therefore, the strong reliance on teacher reports only may also bias the evidence simply because they might be ignorant about the comprehensive perspective on leadership in schools.

To overcome such issues, this thesis examines the recommended approach to collect data from multiple sources (Podsakoff & Organ, 1986). Having collected data about leadership from different sources within the school (e.g. teachers and principals) provides insights into the complexity of leadership, and ultimately contributes to more robust and detailed information about its practice within the school. Most importantly, such data can serve to enrich the general study of leadership where both principals and teachers have a role to play, albeit with different levels of responsibilities.

Additionally, a "more people involved approach" would provide a more informative description of leadership as a joint, shared task, allowing for a detailed and comprehensive understanding of the topic (Halverson et al., 2014; Hunter et al., 2007). It would also allow for studying the dissonance (coherence) in how things are perceived between different school actors, that in some instances may be of critical importance (Braddy et al., 2014; Fleenor et al., 2010; Price, 2012). Such (dis)agreements are sometimes the core phenomena being investigated and represent an important aspect of the overall school environment.

Thus, similarly to pure leadership studies such as VEL-ED or CALL, this thesis suggests that assessing different perspectives when studying leadership is also crucial in TALIS. The context of the TALIS study is particularly interesting because it focuses on the experiences and opinions of teachers and principals. I do acknowledge the recent developments in TALIS that have included teacher and principal perspectives on distributed leadership allowing for investigation of school leadership with respect to teacher and principal (dis)agreement. However, the instructional leadership perspective is still assessed only by principals. In other words, the broader research seeks to find how different leadership practices influence teachers, instruction, and learning outcomes. Thus, in order to gain a comprehensive understanding of this influence, it is essential to also consider the perspective of leadership as experienced by the teachers.

5.4 It may not be feasible to compare leadership across TALIS countries

The thesis demonstrates that achieving scalar invariance for leadership measures in TALIS (needed for comparison of latent means) is highly unlikely, both across countries and among teachers and principals. This means that it is not reasonable or even meaningful to define a joint and universally applicable leadership construct. While some authors offered an alternative way to compare the construct by using the alignment method with the distributed leadership scale (Eryilmaz & Sandoval-Hernandez, 2023), the reality is that leadership practices and the role of the leader vary greatly across the world. The thesis provides evidence to support this argument, as demonstrated in Article 2, where achieving exact invariance across countries, and also across teachers and principals, is deemed highly challenging. This argument is even stronger in Article 1, where the example of the word "feedback" is used to illustrate how interpretations of concepts clearly differ, even among a more homogeneous group of Nordic countries. This thesis further demonstrates that TALIS data are most likely limited for studying leadership cross-culturally. First, the sample size does not allow for to study leadership accounting for all levels that matter, that is teacher, school, and system levels. Second, a limited number of indicators, or to be more precise, the complete lack of indicators at the system level, restricts the possibility to study this research question. Finally, as demonstrated in Article 3, it seems that the existing scales do not capture much of the variability at the county level. Consequently, our results in Article 3 could not support the claim that leadership is strongly culturally embedded.

It is also important to note that extensive effort has already been made by international organizations conducting ILSA to provide a basis for meaningful comparisons across nations. The efforts include high-quality translations (including vocabulary, grammar, language level, punctuation, modifiers, and qualifiers), national adaptations that follow a set of very strict rules, and a very strict quality control process, e.g. back-translation and verification by independent language experts (Dept et al., 2017; Koršňáková et al., 2020; Musu et al., 2020). While TALIS also considers these challenges and carefully examines the issues of measurement invariance by extensively testing for them, the problem of comparability of the measures remains. This is reflected in the fact that only three scales in the whole TALIS survey reached a scale level of invariance of which none measured leadership. At the same time, it is also vital to ensure that

the findings with the TALIS data accurately reflect the diversity of perspectives and experiences across countries and individuals.

Unfortunately, this thesis does not provide clear guidance with respect to multigroup testing and how it should be implemented when studying leadership. However, the thesis demonstrates that there is not a single or easy solution with respect to this issue. Theories of leadership applied in this thesis emphasize how leadership practices and perceptions are situated in culturally dependent norms and values, which suggests that developing measurement invariant scales across countries representing very different political, economic, and social realities would be challenging. However, we are not able to find such system level features of leadership across a larger set of indicators for leadership practices as suggested by Article 3. Rather the variability resided across schools within countries. Nevertheless, in Article 2 countries are represented as fixed effects in the model, and around 10% of the variation in the included principal reported measures could be associated with the country level - which is quite substantial relative to what is normally found with similar data.

Some alternative, interesting approaches to measurement invariance were proposed in the recent studies on instructional leadership and teacher job satisfaction scales from TALIS. The first study examined measurement invariance of instructional leadership scale from TALIS 2018 within eight sub-groups of countries: Nordic, Western European, Central and Eastern European, Southern European, Latin American, Anglo, Southern and East Asian, And Western and Central Asian countries (Eryilmaz & Sandoval Hernandez, 2021). The authors of this study found that within Anglo, Western and Central Asian, Southern European, and Latin American countries, meaningful comparisons of the means can be made. The authors concluded that it has been more probable to reach higher levels of invariance within more homogenous groups of countries (Erylmaz et al., 2020; Rutkowski & Rutkowski, 2018). The second study examined the measurement invariance of the teacher job satisfaction scale using multiple-pairwise comparisons with TALIS 2013 data (Zieger et al., 2019). The authors of this study compared different countries with one reference country (England) and concluded that fully trustworthy comparisons could be made between England and nine other countries of which three were other English-speaking countries and two were Scandinavian countries. Although this approach does not support the original intention of the studies to be able to compare across all countries, it does suggest a way forward to benchmark a single country of interest against other countries, which is frequently the main goal of participating countries (Zieger et al., 2019).

With that in mind, establishing measurement invariance across such a large number of countries remains a challenge in all ILSA, including TALIS, and including school leadership measures (Byrne & Vijver, 2010; Lubke & Muthén, 2004; Rutkowski & Svetina, 2014). Thus, the thesis suggests that it is of crucial importance to continue this line of research, and most importantly, not to ignore those kinds of questions.

5.5 Leadership is primarily a school level phenomenon

Challenge five relates to the difficulty of empirically identifying how leadership is culturally embedded and shaped. First, the thesis demonstrates the lack of measurement invariance supporting the comparison of leadership across different but also more homogeneous countries. This finding indicates that it is not very useful to operationalize leadership globally. Moreover, in Article 3, we are unable to group leadership practice on the basis of any cultural, linguistic, or geographical system features. Overall, our results suggest that leadership as represented in TALIS is clearly a school level phenomenon.

However, as shown in Article 3, the TALIS data might not be the best choice for those interested in studying leadership cross-culturally at the country level. The main reason is the inability to capture the heterogeneity of the leadership concepts in a valid and sound way at the country level. Moreover, the TALIS data might not represent a rich and detailed perspective of one specific leadership theory and some important variables are omitted. The inability to find enough variability at the level of countries, leads us to conclude that leadership is indeed a school level phenomenon. Thus, Article 1 demonstrates that instructional leadership is an important predictor of teacher job satisfaction with the school environment, whereas Article 2 shows that both instructional and distributed leadership are strongly associated with school climate.

The results taken together demonstrate that leadership is indeed important for a larger school environment, which can be represented by various school factors. This finding adds to the extensive literature showing that leadership effects are strongest at the school level and in association with teacher outcomes, as well as that these factors frequently serve as mediators through which leadership influences student learning (Devos et al., 2014; García Torres, 2019; Liu et al., 2021; Ross & Gray, 2006; Sims, 2019). It shows that the operationalization of leadership at different levels or from different perspectives heavily influences our quest for evidence that leadership matters for student learning outcomes. Thus, studies that do not find significant associations between measures of leadership and learning may have operationalized

leadership focusing on studying direct effects. Instead, we suggest that leadership should always be modelled in relation to phenomena that are more proximal to, and hence more likely directly affected by, leadership practices in the school. Then it may be possible to study how leadership is mediated through these phenomena eventually with an effect on students' outcomes.

6. Concluding remarks

This thesis presents a coherent argument that theory and measurement are inseparably linked in the study of leadership. Through the analysis of TALIS data, it becomes apparent that capturing leadership on a global scale represents a hard endeavour. However, by carefully investigating these challenges, the thesis demonstrates that sometimes simple shifts in approaches that are used can lead to notable improvements. For example, the inclusion of teacher reports as leadership indicators can help to address some of these issues. Moreover, the problem of construct underrepresentation of leadership measures in TALIS can be improved by enriching the study design. It would be naive to suggest that TALIS is unaware of these and similar solutions, but it is essential to emphasize the need for continuously improving the study design. This is especially important in light of findings that point to the importance of TALIS data in the study of leadership. Further research is necessary for challenges such as measurement invariance, where the optimal solution has yet to be reached. Thus, the thesis highlights the need for more research on the cross-cultural comparability of leadership phenomena. Finally, it is critical to understand that measurement challenges are substantial, and that leadership is a complex phenomenon. Therefore, we should not expect quick developments or quick fixes. Instead, incremental improvements over time will eventually result in more significant advancements in overall leadership study. I am pleased to see advancements in the field and a greater comprehension of the practice of leadership, with the hope that this thesis represents at least a tiny step in the right direction.

References

Aas, M., & Brandmo, C. (2016). Revisiting instructional and transformational leadership: The contemporary Norwegian context of school leadership. *Journal of Educational Administration*, *54*(1), 92–110. Scopus. https://doi.org/10.1108/JEA-08-2014-0105

Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). Forum: The rise of International Large-Scale Assessments and rationales for participation. *Compare: A Journal of Comparative and International Education*, *47*(3), 434–452. https://doi.org/10.1080/03057925.2017.1301399

Agasisti, T., Bowers, A. J., & Soncin, M. (2019). School principals' leadership types and student achievement in the Italian context: Empirical results from a three-step latent class analysis. *Educational Management Administration & Leadership*, *47*(6), 860–886. https://doi.org/10.1177/1741143218768577

Ahn, J., Bowers, A. J., & Welton, A. D. (2021). Leadership for learning as an organization-wide practice: Evidence on its multilevel structure and implications for educational leadership practice and research. *International Journal of Leadership in Education*, 1–52. https://doi.org/10.1080/13603124.2021.1972162

Ainley, J. and R. Carstens (2018), "Teaching and Learning International Survey (TALIS) 2018 conceptual framework", *OECD Education Working Papers*, No. 187, OECD Publishing, Paris, https://doi.org/10.1787/799337c2-en.

Aldridge, J. M., & Fraser, B. J. (2016). Teachers' views of their school climate and its relationship with teacher self-efficacy and job satisfaction. *Learning Environments Research*, *19*(2), 291–307. https://doi.org/10.1007/s10984-015-9198-x

Ashforth, B. E. (1985). Climate formation: Issues and extensions. *The Academy of Management Review*, *10*(4), 837–847. https://doi.org/10.2307/258051

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology*, *51*(3), 577–598. https://doi.org/10.1111/j.1744-6570.1998.tb00252.x

Avvisati, F., Le Donné, N., & Paccagnella, M. (2019). A meeting report: Cross-cultural comparability of questionnaire measures in large-scale international surveys.

Measurement Instruments for the Social Sciences, *1*(1), 8. https://doi.org/10.1186/s42409-019-0010-z

Begley, P. T. (2001). In pursuit of authentic school leadership practices. *International Journal of Leadership in Education*, *4*(4), 353–365. https://doi.org/10.1080/13603120110078043

Bellibas, M. S., Gümüs, S., & Liu, Y. (2021). Does school leadership matter for teachers' classroom practice? The influence of instructional leadership and distributed leadership on instructional quality. *School Effectiveness and School Improvement*, *32*(3), 387–412. http://dx.doi.org/10.1080/09243453.2020.1858119

Bellibas, M. S., & Liu, Y. (2017). Multilevel analysis of the relationship between principals' perceived practices of instructional leadership and teachers' self-efficacy perceptions. *Journal of Educational Administration*, *55*(1), 49–69. Scopus. https://doi.org/10.1108/JEA-12-2015-0116

Bellibas, M. S., & Liu, Y. (2018). The effects of principals' perceived instructional and distributed leadership practices on their perceptions of school climate. *International Journal of Leadership in Education*, *21*(2), 226–244. https://doi.org/10.1080/13603124.2016.1147608

Berkovich, I., & Bogler, R. (2020). The relationship between school leadership standards and school administration imperatives: An international perspective. *School Leadership and Management*, *40*(4), 321–334. Scopus. https://doi.org/10.1080/13632434.2019.1682538

Blase, J., & Blase, J. (1999). Principals' instructional leadership and teacher development: teachers' perspectives. *Educational Administration Quarterly*, *35*(3), 349–378. https://doi.org/10.1177/0013161X99353003

Blase, J., & Blase, J. (2000). Effective instructional leadership. *Journal of Educational Administration*, *38*, 130–141. https://doi.org/10.1108/09578230010320082

Blase, J., & Blase, J. (2004). The dark side of school leadership: implications for administrator preparation. *Leadership and Policy in Schools*, *3*(4), 245–273. https://doi.org/10.1080/15700760490503733

Blömeke, S., Nilsen, T., Olsen, R. V., & Gustafsson, J.-E. (2022). Conceptual and methodological accomplishments of ILSAs, remaining criticism and limitations. In T.

Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–54). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_22-1

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Bossert, S. T., Dwyer, D. C., Rowan, B., & Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64. https://doi.org/10.1177/0013161X82018003004

Bowers, A. J. (2020). Examining a congruency-typology model of leadership for learning using two-level latent class analysis with TALIS 2018. *OECD Education Working Papers No. 219*. https://doi.org/10.1787/c963073b-en

Boyce, J., & Bowers, A. J. (2018a). Toward an evolving conceptualization of instructional leadership as leadership for learning: Meta-narrative review of 109 quantitative studies across 25 years. *Journal of Educational Administration*, *56*(2), 161–182. https://doi.org/10.1108/JEA-06-2016-0064

Boyce, J., & Bowers, A. J. (2018b). Different levels of leadership for learning: Investigating differences between teachers individually and collectively using multilevel factor analysis of the 2011-2012 Schools and Staffing Survey. *International Journal of Leadership in Education*, *21*(2), 197–225. https://doi.org/10.1080/13603124.2016.1139187

Braddy, P. W., Gooty, J., Fleenor, J. W., & Yammarino, F. J. (2014). Leader behaviors and career derailment potential: A multi-analytic method examination of rating source and self–other agreement. *The Leadership Quarterly*, *25*(2), 373–390. https://doi.org/10.1016/j.leaqua.2013.10.001

Brauckmann, S., & Schwarz, A. (2014). Autonomous leadership and a centralised school system: An odd couple? Empirical insights from Cyprus. *International Journal of Educational Management*, *28*(7), 823–841. https://doi.org/10.1108/IJEM-08-2013-0124

Brezicha, K. F., Ikoma, S., Park, H., & LeTendre, G. K. (2020). The ownership perception gap: Exploring teacher job satisfaction and its relationship to teachers' and principals' perception of decision-making opportunities. *International Journal of Leadership in Education*, *23*(4), 428–456. https://doi.org/10.1080/13603124.2018.1562098

Brooks, M. C., & Mutohar, A. (2018). Islamic school leadership: A conceptual framework. *Journal of Educational Administration and History*, *50*(2), 54–68. https://doi.org/10.1080/00220620.2018.1426558

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. (2016). School leadership and management in England: The paradox of simultaneous centralisation and decentralisation. *Research in Educational Administration & Leadership*, *1*(1), 1–23. http://dx.doi.org/10.30828/real/2016.1.1

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M., & Vijver, F. J. R. van de. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. https://doi.org/10.1080/15305051003637306

Cieciuch, J., Davidov, E., Schmidt, P., & Algesheimer, R. (2019). How to obtain comparable measures for cross-national comparisons. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie*, *71*(1), 157–186. https://doi.org/10.1007/s11577-019-00598-7

Clarke, S., & O'Donoghue, T. (2016). Educational leadership and context: A rendering of an inseparable relationship. *British Journal of Educational Studies*, *65*(2), 167–182. https://doi.org/DOI: 10.1080/00071005.2016.1199772

Çoban, Ö., & Atasoy, R. (2020). Relationship between distributed leadership, teacher collaboration and organizational innovativeness. *International Journal of Evaluation*

and Research in Education, *9*(4), 903–911. Scopus. https://doi.org/10.11591/ijere.v9i4.20679

Collie, R. J. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, *104*(4), 1189. https://doi.org/10.1037/a0029356

Condon, C., & Clifford, M. (2012). *Measuring principal performance. How rigorous are commonly used principal performance assessment instruments?* [Evaluative Report]. AIR. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED565757.pdf

Daniëls, E., Hondeghem, A., & Dochy, F. (2019). A review on leadership and leadership development in educational settings. *Educational Research Review*, *27*, 110–125. https://doi.org/10.1016/j.edurev.2019.02.003

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

Day, D. V., & Harrison, M. M. (2007). A multilevel, identity-based approach to leadership development. *Human Resource Management Review*, *17*(4), 360–373. https://doi.org/10.1016/j.hrmr.2007.08.007

De Roover, K., Vermunt, J. K., & Ceulemans, E. (2022). Mixture multigroup factor analysis for unraveling factor loading noninvariance across many groups. *Psychological Methods*, *27*, 281–306. https://doi.org/10.1037/met0000355

Deal, T. E., & Kennedy, A. A. (1983). Culture and school performance. *Educational Leadership*, *40*(5), 14–15.

Deal, T. E., & Peterson, K. D. (1999). Shaping school culture: The heart of leadership. In *Undefined* (p. 160). Jossey-Bass Inc., Publishers.

Dept, S., Ferrari, A., & Halleux, B. (2017). Translation and cultural appropriateness of survey material in large-scale assessments. In *Implementation of Large-Scale Education Assessments* (pp. 168–192). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118762462.ch6

Devos, G., Tuytens, M., & Hulpia, H. (2014). Teachers' organizational commitment: examining the mediating effects of distributed leadership. *American Journal of Education*, *120*(2), 205–231. https://doi.org/10.1086/674370

Dickhäuser, O., Janke, S., Daumiller, M., & Dresel, M. (2021). Motivational school climate and teachers' achievement goal orientations: A hierarchical approach. *British Journal of Educational Psychology*, *91*(1), e12370. https://doi.org/10.1111/bjep.12370

Dumay, X., Boonen, T., & Van Damme, J. (2013). Principal leadership long-term indirect effects on learning growth in mathematics. *The Elementary School Journal*, *114*(2), 225–251. https://doi.org/10.1086/673198

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*(1), 149–167. https://doi.org/10.1016/j.leaqua.2004.09.009

Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, *5*(1), 23–35. https://doi.org/10.1207/s15327574ijt0501_3

Eryilmaz, N., & Sandoval Hernandez, A. (2021). Improving cross-cultural comparability: Does school leadership mean the same in different countries? *Educational Studies*, *0*(0), 1–22. https://doi.org/10.1080/03055698.2021.2013777

Eryilmaz, N., & Sandoval-Hernandez, A. (2023). Is distributed leadership universal? A cross-cultural, comparative approach across 40 countries: An alignment optimisation approach. *Education Sciences*, *13*(2), Article 2. https://doi.org/10.3390/educsci13020218

Erylmaz, N., Rivera Gutiérrez, M., & Sandoval Hernández, A. (2020). Should different countries participating in PISA interpret socioeconomic background in the same way? : A measurement invariance approach. *Revista Iberoamericana de Educación*, *84*(1), 109–133. https://doi.org/10.35362/rie8413981

Favero, N., & Bullock, J. B. (2015). How (not) to solve the problem: an evaluation of scholarly responses to common source bias. *Journal of Public Administration Research and Theory*, *25*(1), 285–308. https://doi.org/10.1093/jopart/muu020

Firestone, W. A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, *16*(7), 16–21. https://doi.org/10.3102/0013189X016007016

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing*, *17*(2), 105–120. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2<105::AID-MAR3>3.0.CO;2-9

Flath, B. (1989). The principal as instructional leader. *ATA Magazines*, *69*(3), 19–22, 47–49.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self–other rating agreement in leadership: A review. *The Leadership Quarterly*, *21*(6), 1005–1034. https://doi.org/10.1016/j.leaqua.2010.10.006

Flessa, J., Bramwell, D., Fernandez, M., & Weinstein, J. (2018). School leadership in Latin America 2000–2016. *Educational Management Administration & Leadership*, *46*(2), 182–206. https://doi.org/10.1177/1741143217717277

Frase, L., & Sorenson, L. (2016). Teacher motivation and satisfaction: Impact on participatory management. *NASSP Bulletin*, *76*(540). https://doi.org/10.1177/01926365920765400

Frost, D., & Swaffield, S. (2008). Researching LfL through an international collaborative project. In *Leadership for Learning. International perspectives* (pp. 101–122). Brill. https://brill.com/display/book/edcoll/9789087903909/BP000007.xml

García Torres, D. (2019). Distributed leadership, professional collaboration, and teachers' job satisfaction in U.S. schools. *Teaching and Teacher Education*, *79*, 111–123. Scopus. https://doi.org/10.1016/j.tate.2018.12.001

Glassow, L. N., Rolfe, V., & Hansen, K. Y. (2021). Assessing the comparability of teacher-related constructs in TIMSS 2015 across 46 education systems: An alignment optimization approach. *Educational Assessment, Evaluation and Accountability*, *33*(1), 105–137. Social Science Premium Collection. https://doi.org/10.1007/s11092-020-09348-2

Goldring, E., Cravens, X., Porter, A., Murphy, J., & Elliott, S. (2015). The convergent and divergent validity of the Vanderbilt Assessment of Leadership in Education (VAL-ED): Instructional leadership and emotional intelligence. *Journal of Educational Administration*, *53*(2), 177–196. https://doi.org/10.1108/JEA-06-2013-0067

Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: What influences principal practice? *Journal of Educational Administration*, *46*(3), 332–352. https://doi.org/10.1108/09578230810869275

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

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, *24*(1), 23–37. https://doi.org/10.1080/02680930802412669

Griffith, J. (1999). The school leadership/school climate relation: identification of school configurations associated with change in principals. *Educational Administration Quarterly*, *35*(2), 267–291. https://doi.org/10.1177/00131619921968545

Gronn, P. (2000). Distributed properties: A new architecture for leadership. *Educational Management & Administration*, *28*(3), 317–338. https://doi.org/10.1177/0263211X000283006

Gronn, P. (2002). Distributed leadership as a unit of analysis. *The Leadership Quarterly*, *13*(4), 423–451. https://doi.org/10.1016/S1048-9843(02)00120-0

Gronn, P. (2003). Leadership: Who needs it? *School Leadership & Management*, *23*(3). https://doi.org/10.1080/1363243032000112784

Gronn, P. (2009). Leadership configurations. *Leadership*, *5*(3), 381–394. https://doi.org/10.1177/1742715009337770

Gumus, E., & Bellibas, M. S. (2016). The effects of professional development activities on principals' perceived instructional leadership practices: Multi-country data analysis using TALIS 2013. *Educational Studies*, *42*(3), 287–301. https://doi.org/10.1080/03055698.2016.1172958

Gumus, S. (2013). The effects of teacher- and school-level factors on teachers' participation in professional development activities: the role of principal leadership. *Journal of International Education Research*, *9*(4), 371. http://dx.doi.org/10.19030/jier.v9i4.8089

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

Gumus, S., Bulut, O., & Bellibas, M. S. (2013). The relationship between principal leadership and teacher collaboration in Turkish primary schools: A multilevel analysis. *Education Research and Perspectives*, *40*(1), 1–29.

Gurr, D. (2014). Successful school leadership across contexts and cultures. *Leading and Managing*, *20*(2), 75–88. https://doi.org/10.3316/informit.036841514593192

Gustafsson, J. E., & Nilsen, T. (2016). The impact of school climate and teacher quality on mathematics achievement: A difference-in-differences approach. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 81–95). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8_4

Hairon, S., & Goh, J. W. (2015). Pursuing the elusive construct of distributed leadership: Is the search over? *Educational Management Administration & Leadership*, *43*(5), 693–718. https://doi.org/10.1177/1741143214535745

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. (2009, September). *Leadership for the 21st century schools: From instructional leadership to leadership for learning* [Public Lecture Series of the Hong Kong Institute of Education]. https://repository.eduhk.hk/en/publications/leadership-for-the-21st-century-schools-from-instructional-leader-3

Hallinger, P. (2010). Developing instructional leadership. In B. Davies & M. Brundrett (Eds.), *Developing Successful Leadership* (pp. 61–76). Springer Netherlands. 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. (2014). Reviewing reviews of research in educational leadership: An empirical assessment. *Educational Administration Quarterly*, *50*(4), 539–576. https://doi.org/10.1177/0013161X13506594

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 *Principles of Educational Leadership anf Management* (Third edition, p. 408). 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., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management*, *30*(2), 95–110. https://doi.org/10.1080/13632431003663214

Hallinger, P., & Heck, R. H. (2011). Conceptual and methodological issues in studying school leadership effects as a reciprocal process. *School Effectiveness and School Improvement*, *22*(2), 149–173. https://doi.org/10.1080/09243453.2011.565777

Hallinger, P., & Huber, S. (2012). School leadership that makes a difference: International perspectives. *School Effectiveness and School Improvement*, *23*(4), 359–367. https://doi.org/10.1080/09243453.2012.681508

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

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

Halverson, R., Kelley, C., & Shaw, J. (2014). A call for improved school leadership. *Phi Delta Kappan*, *95*(6), 57–60. https://doi.org/10.1177/003172171409500612

Ham, S., & Kim, R. Y. (2015). The influence of principals' instructional leadership on teachers' use of autonomy-supportive instruction: An analysis of three Asia-Pacific countries.

The Asia - Pacific Education Researcher, *24*(1), 57–65. Social Science Premium Collection. https://doi.org/10.1007/s40299-013-0158-x

Harris, A. (2004). Distributed leadership and school improvement: leading or misleading? *Educational Management Administration & Leadership*, *32*(1), 11–24. https://doi.org/10.1177/1741143204039297

Harris, A. (Ed.). (2009a). *Distributed leadership: Different perspectives*. Springer.

Harris, A. (2009b). Distributed leadership: What we know. In A. Harris (Ed.), *Distributed Leadership: Different Perspectives* (pp. 11–21). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9737-9_2

Harris, A. (2013). *Distributed School Leadership: Developing Tomorrow's Leaders*. Routledge. https://doi.org/10.4324/9780203607909

Harris, A. (2020). Leading school and system improvement: Why context matters. *European Journal of Education*, *55*(2), 143–145. https://doi.org/10.1111/ejed.12393

Harris, A., Leithwood, K., Day, C., Sammons, P., & Hopkins, D. (2007). Distributed leadership and organizational change: Reviewing the evidence. *Journal of Educational Change*, *8*(4), 337–347. https://doi.org/10.1007/s10833-007-9048-4

Hastedt, D., & Sibberns, H. (2020). Future directions, recommendations, and potential developments of ILSA. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_24-1

He, J., Barrera-Pedemonte, F., & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice*, *26*(4), 369–385. https://doi.org/10.1080/0969594X.2018.1469467

He, J., Buchholz, J., & Fischer, J. (2022). Cross-cultural comparability of latent constructs in ILSAs. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–26). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_58-1

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

Hendriks, M. A., & Scheerens, J. (2013). School leadership effects revisited: A review of empirical studies guided by indirect-effect models. *School Leadership & Management*, *33*(4), 373–394. https://doi.org/10.1080/13632434.2013.813458

Hendriks, M., & Steen, R. (2012). Results from school leadership effectiveness studies (2005–2010). In J. Scheerens (Ed.), *School Leadership Effects Revisited: Review and Meta-Analysis of Empirical Studies* (pp. 65–129). Springer Netherlands. https://doi.org/10.1007/978-94-007-2768-7_4

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and* (2nd ed.). SAGE Publications.

Holzberger, D., & Prestele, E. (2021). Teacher self-efficacy and self-reported cognitive activation and classroom management: A multilevel perspective on the role of school characteristics. *Learning and Instruction*, *76*. Scopus. https://doi.org/10.1016/j.learninstruc.2021.101513

Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2017). *Multilevel Analysis: Techniques and Applications, Third Edition*. Routledge.

Hoy, A. W., & Hoy, W. K. (2006). *Instructional leadership: A research-based guide to learning in schools, 2nd ed* (2nd ed.). Pearson Education New Zealand.

Hoy, W. K. (1990). Organizational climate and culture: A conceptual analysis of the school workplace. *Journal of Educational and Psychological Consultation*, *1*(2), 149–168. https://doi.org/10.1207/s1532768xjepc0102_4

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425–446. https://doi.org/10.3102/00028312043003425

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). Springer Netherlands. https://doi.org/10.1007/978-90-481-3501-1_4

Hulpia, H., & Devos, G. (2009). Exploring the link between distributed leadership and job satisfaction of school leaders. *Educational Studies*, *35*(2), 153–171. https://doi.org/10.1080/03055690802648739

Hulpia, H., Devos, G., & Rosseel, Y. (2009). Development and validation of scores on the distributed leadership inventory. *Educational and Psychological Measurement*, *69*(6), 1013–1034. https://doi.org/10.1177/0013164409344490

Hulpia, H., Devos, G., Rosseel, Y., & Vlerick, P. (2012). Dimensions of distributed leadership and the impact on teachers' organizational commitment: A study in secondary education. *Journal of Applied Social Psychology*, *42*(7), 1745–1784. https://doi.org/10.1111/j.1559-1816.2012.00917.x

Hunter, S. T., Bedell-Avers, K. E., & Mumford, M. D. (2007). The typical leadership study: Assumptions, implications, and potential remedies. *The Leadership Quarterly*, *18*(5), 435–446. https://doi.org/10.1016/j.leaqua.2007.07.001

Jackson, K. M. (2019). A review of research methods trends in educational leadership journals. *Education Leadership Review*, *20*(1), 119–132.

Jakubowski, M., & Gajderowicz, T. (2022). Application of multilevel models to International Large-Scale Student Assessment data. In M. S. Khine (Ed.), *Methodology for Multilevel Modelling in Educational Research: Concepts and Applications* (pp. 185–201). Springer. https://doi.org/10.1007/978-981-16-9142-3_10

Johansson, S. (2016). International large-scale assessments: What uses, what consequences? *Educational Research*, *58*(2), 139–148. https://doi.org/10.1080/00131881.2016.1165559

Johnson, L., Møller, J., Jacobson, S. L., & Wong, K. C. (2008). Cross-national comparisons in the International Successful School Principalship Project (ISSPP): The USA, Norway and China. *Scandinavian Journal of Educational Research*, *52*(4), 407–422. https://doi.org/10.1080/00313830802184582

Karadag, E. (2020). The effect of educational leadership on students' achievement: A cross-cultural meta-analysis research on studies between 2008 and 2018. *Asia Pacific Education Review*, *21*(1), 49–64. https://doi.org/10.1007/s12564-019-09612-1

Katsantonis, I. G. (2020). Investigation of the impact of school climate and teachers' self-efficacy on job satisfaction: A cross-cultural approach. *European Journal of*

Investigation in Health, Psychology and Education, *10*(1), Article 1. https://doi.org/10.3390/ejihpe10010011

Kelley, Carolyn and Halverson, Richard (2012). The comprehensive assessment of leadership for learning: A next generation formative evaluation and feedback system. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *3*(2).

Kelley, R. C. (2005). Relationships between measures of leadership and school climate. *Education*, *Vol.126*(1), 17–25.

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

Kılınç, A. Ç., Polatcan, M., Turan, S., & Özdemir, N. (2022). Principal job satisfaction, distributed leadership, teacher-student relationships, and student achievement in Turkey: A multilevel mediated-effect model. *Irish Educational Studies*. Scopus. https://doi.org/10.1080/03323315.2022.2061567

Klar, H. W., Huggins, K. S., Hammonds, H. L., & Buskey, F. C. (2016). Fostering the capacity for distributed leadership: A post-heroic approach to leading school improvement. *International Journal of Leadership in Education*, *19*(2), 111–137. https://doi.org/10.1080/13603124.2015.1005028

Klenke, K. (2008). *Qualitative Research in the Study of Leadership*. Emerald Group Publishing.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modelling, Fourth Edition*. Guilford Publications.

Koršňáková, P., Dept, S., & Ebbs, D. (2020). Translation: The preparation of national language versions of assessment instruments. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment: Understanding IEA's Comparative Studies of Student Achievement* (pp. 85–111). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_6

Kozlowski, S. W., & Doherty, M. L. (1989). Integration of climate and leadership: Examination of a neglected issue. *Journal of Applied Psychology*, *74*(4), 546–553. https://doi.org/10.1037/0021-9010.74.4.546

Krug, S. E. (1992). Instructional leadership: A constructivist perspective. *Educational Administration Quarterly*, *28*(3), 430–443. https://doi.org/10.1177/0013161X92028003012

Krüger, M. L., 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?* [Data set]. National Center for Analysis of Longitudinal Data in Education Research. https://doi.org/10.1037/e722072011-001

Lambersky, J. (2016). Understanding the human side of school leadership: Principals' impact on teachers' morale, self-efficacy, stress, and commitment. *Leadership and Policy in Schools*, *15*(4), 379–405. https://doi.org/10.1080/15700763.2016.1181188

Lambert, L. (2002). A framework for shared leadership. *Educational Leadership*, *8*(59), 37–40.

Leeuw, J. de, & Meijer, E. (2008). Introduction to multilevel analysis. In J. de Leeuw & E. Meijer (Eds.), *Handbook of Multilevel Analysis* (pp. 1–75). Springer. https://doi.org/10.1007/978-0-387-73186-5_1

Leithwood, K. (2001). School leadership in the context of accountability policies. *International Journal of Leadership in Education*, *4*(3), 217–235. https://doi.org/10.1080/13603120110057082

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., Harris, A., & Hopkins, D. (2020). Seven strong claims about successful school leadership revisited. *School Leadership & Management*, *40*(1), 5–22. https://doi.org/10.1080/13632434.2019.1596077

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning.* Executive summary. Retrieved from https://www.wallacefoundation.org/knowledge-center/Documents/How-Leadership-Influences-Student-Learning.pdf

Lek, K., Oberski, D., Davidov, E., Cieciuch, J., Seddig, D., & Schmidt, P. (2018). Approximate measurement invariance. In *Advances in Comparative Survey Methods* (pp. 911–929). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118884997.ch41

Liu, J., Wang, K., Chen, Z., & Pan, Z. (2023). Exploring the contributions of job resources, job demands, and job self-efficacy to STEM teachers' job satisfaction: A commonality analysis. *Psychology in the Schools*, *60*(1), 122–142.

Liu, S., Keeley, J. W., Sui, Y., & Sang, L. (2021). Impact of distributed leadership on teacher job satisfaction in China: The mediating roles of teacher autonomy and teacher collaboration. *Studies in Educational Evaluation*, *71*. Scopus. https://doi.org/10.1016/j.stueduc.2021.101099

Liu, Y. (2020). Focusing on the practice of distributed leadership: The international evidence from the 2013 TALIS. *Educational Administration Quarterly*, *56*(5), 779–818. Scopus. https://doi.org/10.1177/0013161X20907128

Liu, Y., Bellibas, M. S., & Gümüs, S. (2021). 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*, *49*(3), 430–453.

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

Loeb, S., & Byun, E. (2022). Educational accountability and the role of international large-scale assessments. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education:*

Perspectives, Methods and Findings (pp. 75–95). Springer International Publishing. https://doi.org/10.1007/978-3-030-88178-8_6

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

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modelling: A Multidisciplinary Journal*, *11*(4), 514–534. https://doi.org/10.1207/s15328007sem1104_2

Lumby, J. (2016). Distributed leadership as fashion or fad. *Management in Education*, *30*(4), 161–167. https://doi.org/10.1177/0892020616665065

Lunenburg, F. C. (2010). The principal and the school: What do principals do? *National Forum Of Educational Administration And Supervision Journal*, *27*(4).

Lyons-Thomas, J., Ercikan, K., Gonzalez, E., & Kirsch, I. (2022). Implementing ILSAs. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_28-1

MacBeath, J. (2019). Leadership for learning. In T. Townsend (Ed.), *Instructional Leadership and Leadership for Learning in Schools: Understanding Theories of Leading* (pp. 49–75). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-23736-3

MacBeath, J. (2020). Leadership is for learning—A critique of current misconceptions around leadership for learning. *Zeitschrift Für Erziehungswissenschaft*, *23*(5), 903–923. https://doi.org/10.1007/s11618-020-00967-5

MacBeath, J., & Dempster, N. (2008). *Connecting Leadership and Learning: Principles for Practice*. Routledge.

MacBeath, J., Frost, D., & Swaffield, S. (2005). Researching leadership for learning in seven countries (The Carpe Vitam project). *Educational Research and Perspectives*, *23*(2), 19.

MacBeath, J., & Townsend, T. (2011). Thinking and acting both locally and globally: What do we know now and how do we continue to improve? In T. Townsend & J. MacBeath

(Eds.), *International Handbook of Leadership for Learning* (pp. 1237–1254). Springer Netherlands. https://doi.org/10.1007/978-94-007-1350-5_66

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education*, *12*(1), 73–84. https://doi.org/10.1080/13603120701576241

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., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Meinck, S., & Vandenplas, C. (2021). Sampling design in ILSA. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–25). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_25-1

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741. https://doi.org/10.1037/0003-066X.50.9.741

Miller, P. W. (2018). The nature of school leadership. In P. W. Miller (Ed.), *The Nature of School Leadership: Global Practice Perspectives* (pp. 165–185). Springer International Publishing. https://doi.org/10.1007/978-3-319-70105-9_9

Millsap, R. E. (2012). *Statistical Approaches to Measurement Invariance*. Routledge.

Mincu, M. (2022). Why is school leadership key to transforming education? Structural and cultural assumptions for quality education in diverse contexts. *PROSPECTS*. https://doi.org/10.1007/s11125-022-09625-6

Mitchell, C., & Sackney, L. (2011). Building and leading within learning ecologies. In T. Townsend & J. MacBeath (Eds.), *International Handbook of Leadership for Learning* (pp. 975–990). Springer Netherlands. https://doi.org/10.1007/978-94-007-1350-5_53

Møller, J., & Schratz, M. (2009). Leadership development in Europe. In *International Handbook on the Preparation and Development of School Leaders*. Routledge.

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *The Journal of Classroom Interaction*, *37*(2), 3–15.

Murphy, J. (1988). Methodological, measurement, and conceptual problems in the study of instructional leadership. *Educational Evaluation and Policy Analysis*, *4*(4), 290–310.

Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership & Management*, *27*(2), 179–201. https://doi.org/10.1080/13632430701237420

Musu, L., Dohr, S., & Netten, A. (2020). Quality control during data collection: Refining for rigor. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment: Understanding IEA's Comparative Studies of Student Achievement* (pp. 131–150). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_8

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research*, *47*(4), 637–664. https://doi.org/10.1177/0049124117701488

Nettles, S. M., & Herrington, C. (2007). Revisiting the importance of the direct effects of school leadership on student achievement: The implications for school improvement policy. *Peabody Journal of Education*, *82*(4), 724–736. https://doi.org/10.1080/01619560701603239

Oc, B. (2018). Contextual leadership: A systematic review of how contextual factors shape leadership and its outcomes. *The Leadership Quarterly*, *29*(1), 218–235. https://doi.org/10.1016/j.leaqua.2017.12.004

OECD. (2010). *TALIS 2008 Technical Report*. OECD Publishing.

OECD. (2014a). *Evaluating Measurement Invariance of TALIS 2013 Complex Scales | READ online*. OECD ILibrary. https://read.oecd-ilibrary.org/education/evaluating-measurement-invariance-of-talis-2013-complex-scales_5jz2kbbvlb7k-en

OECD. (2014b). *TALIS 2013 Technical report*. OECD Publishing.

OECD. (2016). *School Leadership for Learning: Insights from TALIS 2013, TALIS*. OECD Publishing. https://www.oecd.org/education/school-leadership-for-learning-9789264258341-en.htm

OECD. (2019a). *TALIS 2018 Technical Report.* OECD Publishing.

OECD. (2019b). *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners*. OECD. https://doi.org/10.1787/1d0bc92a-en

OECD. (2022). *TALIS FAQ*. OECD. https://www.oecd.org/education/talis/talisfaq/

Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*(2), 224–243. https://doi.org/10.1177/0013161X95031002004

Oplatka, I., & Arar, K. (2017). Context and implications document for: The research on educational leadership and management in the Arab world since the 1990s: A systematic review. *Review of Education*, *5*(3), 308–310. https://doi.org/10.1002/rev3.3096

Özdemir, N., Gümüş, S., Kılınç, A. Ç., & Bellibaş, M. Ş. (2022). A systematic review of research on the relationship between school leadership and student achievement: An updated framework and future direction. *Educational Management Administration & Leadership*, 17411432221118662. https://doi.org/10.1177/17411432221118662

Park, J.-H., & Ham, S.-H. (2016). Whose perception of principal instructional leadership? Principal-teacher perceptual (dis)agreement and its influence on teacher collaboration. *Asia Pacific Journal of Education*, *36*(3), 450–469. Scopus. https://doi.org/10.1080/02188791.2014.961895

Pitt, A. (2010). On having one's chance: Autonomy as education's limit. *Educational Theory*, *60*(1), 1–18. https://doi.org/10.1111/j.1741-5446.2009.00342.x

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544. https://doi.org/10.1177/014920638601200408

Pokropek, A., Borgonovi, F., & McCormick, C. (2017). On the cross-country comparability of indicators of socioeconomic resources in PISA. *Applied Measurement in Education*, *30*(4), 243–258. https://doi.org/10.1080/08957347.2017.1353985

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modelling: A Multidisciplinary Journal*, *26*(5), 724–744. https://doi.org/10.1080/10705511.2018.1561293

Pont, B., Nusche, D., & Moorman, H. (2008). *Improving School Leadership, Volume 1 | READ online*. OECD ILibrary. https://read.oecd-ilibrary.org/education/improving-school-leadership_9789264044715-en

Porter, A. C., Murphy, J., Goldring, E. B., Elliott, S. N., Polikoff, M. S., & May, H. (2008). *Vanderbilt Assessment of Leadership in Education*. https://doi.org/10.1037/t67398-000

Price, H., & Carstens, R. (2020). Teaching and Learning International Survey (TALIS) 2018 Analysis Plan. *OECD Education Working Papers, No. 220*. OECD Publishing.

Price, H. E. (2012). Principal–teacher interactions: How affective relationships shape principal and teacher attitudes. *Educational Administration Quarterly*, *48*(1), 39–85. https://doi.org/10.1177/0013161X11417126

Printy, S., & Liu, Y. (2021). Distributed leadership globally: The interactive nature of principal and teacher leadership in 32 countries. *Educational Administration Quarterly*, *57*(2), 290–325. https://doi.org/10.1177/0013161X20926548

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

Ross, J. A., & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *29*(3), 798–822. https://doi.org/10.2307/20054196

Rutkowski, D., Rutkowski, L., Bélanger, J., Knoll, S., Weatherby, K., & Prusinski, E. (2013). Teaching and Learning International Survey TALIS 2013: Conceptual framework. Final. *OECD Publishing*.

Rutkowski, L., & Rutkowski, D. (2010). Getting it 'better': The importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies*, *42*(3), 411–430. https://doi.org/10.1080/00220272.2010.487546

Rutkowski, L., & Rutkowski, D. (2018). Improving the comparability and local usefulness of international assessments: A look back and a way forward. *Scandinavian Journal of Educational Research*, *62*(3), 354–367. https://doi.org/10.1080/00313831.2016.1261044

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Pergamon. https://research.utwente.nl/en/publications/the-foundations-of-educational-effectiveness

Schein, E. H. (2010). *Organizational Culture and Leadership*. John Wiley & Sons.

Scherer, R., Jansen, M., Nilsen, T., Areepattamannil, S., & Marsh, H. (2016). The quest for comparability: studying the invariance of the Teachers' Sense of Self-Efficacy (TSES) Measure across Countries. *PLOS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0150829

Scherer, R., & Nilsen, T. (2016). The relations among school climate, instructional quality, and achievement motivation in mathematics. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 51–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8_3

Schleicher, A., Ikeda, M., Thorn, W., & Tremblay, K. (2020). OECD studies and the case of PISA, PIAAC, and TALIS. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International Handbook of Comparative Large-Scale Studies in Education: Perspectives, Methods and Findings* (pp. 1–42). Springer International Publishing. https://doi.org/10.1007/978-3-030-38298-8_17-1

Schoen, L. T., & Teddlie, C. (2008). A new model of school culture: A response to a call for conceptual clarity. *School Effectiveness and School Improvement*, *19*(2), 129–153. https://doi.org/10.1080/09243450802095278

Ševkušić, S., Malinić, D., & Teodorović, J. (2021). Leadership in education. Initiatives and trends in selected European countries. *CEPS Journal : Center for Educational Policy Studies Journal*, *11*(1), 139–144. https://doi.org/10.26529/cepsj.1165

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

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. Sage.

Southworth, G. (1993). School leadership and school development: Reflections from research. *School Organisation*, *13*(1), 73–87. https://doi.org/10.1080/0260136930130107

Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership & Management*, *22*(1), 73–91. https://doi.org/10.1080/13632430220143042

Spillane, J. P., Camburn, E. M., & Stitziel Pareja, A. (2007). Taking a distributed perspective to the school principal's workday. *Leadership and Policy in Schools*, *6*(1), 103–125. https://doi.org/10.1080/15700760601091200

Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher*, *30*(3), 23–28. https://doi.org/10.3102/0013189X030003023

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, *36*(1), 3–34. https://doi.org/10.1080/0022027032000106726

Spillane, J., & Zuberi, A. (2009). Designing and piloting a leadership daily practice log: Using logs to study the practice of leadership. *Educational Administration Quarterly*, *45*(3), 375–423. https://doi.org/10.1177/0013161X083292

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

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90. https://doi.org/10.1086/209528

Steinmetz, H. (2018). Estimation and comparison of latent means across cultures. In *Cross-Cultural Analysis: Methods and Applications* (Second edition). Routledge.

Tan, C. Y., Gao, L., & Shi, M. (2020). Second-order meta-analysis synthesizing the evidence on associations between school leadership and different school outcomes. *Educational Management Administration & Leadership*, 1741143220935456. https://doi.org/10.1177/1741143220935456

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Thomson, P. (2017). A little more madness in our methods? A snapshot of how the educational leadership, management and administration field conducts research. *Journal of Educational Administration and History*, *49*(3), 215–230. https://doi.org/10.1080/00220620.2017.1315381

Thoonen, E. E. J., Sleegers, P. J. C., Oort, F. J., & Peetsma, T. T. D. (2012). Building school-wide capacity for improvement: The role of leadership, school organizational conditions, and teacher factors. *School Effectiveness and School Improvement*, *23*(4), 441–460. https://doi.org/10.1080/09243453.2012.678867

Tian, M., Risku, M., & Collin, K. (2016). A meta-analysis of distributed leadership from 2002 to 2013: Theory development, empirical evidence and future research focus. *Educational Management Administration & Leadership*, *44*(1), 146–164. https://doi.org/10.1177/1741143214558576

Tulowitzki, P., Pietsch, M., & Spillane, J. (2021). Leadership for learning in Germany and the US: Commonalties and differences. In A. Wilmers & S. Jornitz (Eds.), *International perspectives on school settings, education policy and digital strategies: A transatlantic discourse in education research* (pp. 62–79). Verlag Barbara Budrich.

Türker, Y., & Kahraman, Ü. (2021). School climate and self-efficacy as predictor of job satisfaction. *Kuramsal Eğitimbilim*, *14*(4), 548–569. https://doi.org/10.30831/akukeg.901457

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*, *18*(3), 249–269. https://doi.org/10.1080/15700763.2017.1384499

Van de Vijver, F. J. R., Avvisati, F., Davidov, E., Eid, M., Fox, J.-P., Le Donné, N., Lek, K., Meuleman, B., Paccagnella, M., & van de Schoot, R. (2019). Invariance analyses in large-scale studies. OECD education working papers, No. 201. *OECD Publishing*.

Van Houtte, M. (2005). Climate or culture? A plea for conceptual clarity in school effectiveness research. *School Effectiveness and School Improvement*, *16*(1), 71–89. https://doi.org/10.1080/09243450500113977

Vedøy, G., & Moller, J. (2007). Successful school leadership for diversity? Examining two contrasting examples of working for democracy in Norway. *International Studies in Educational Administration (Commonwealth Council for Educational Administration & Management (CCEAM))*, *35*(3), 58–66.

Veletić, J., & Olsen, R. V. (2021). Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018. *International Journal of Leadership in Education*. Scopus. https://doi.org/10.1080/13603124.2021.1953612

Wagemaker, H. (2020). Study design and evolution, and the imperatives of reliability and validity. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment: Understanding IEA's Comparative Studies of Student Achievement* (pp. 7–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-53081-5_2

Walker, A., & Dimmock, C. (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, & B. Mulford (Eds.), *Second International Handbook of Educational Leadership and Administration* (pp. 167–202). Springer Netherlands. https://doi.org/10.1007/978-94-010-0375-9_7

Wang, K., Li, Y., Luo, W., & Zhang, S. (2019). Selected factors contributing to teacher job satisfaction: A quantitative investigation using 2013 TALIS data. *Leadership and Policy in Schools*, *0*(0), 1–21. https://doi.org/10.1080/15700763.2019.1586963

Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, Y., & Ahn, J. (2023). The more the merrier? A network analysis of construct content validity in school leadership literature. *Educational Management Administration & Leadership*, 17411432231155730. https://doi.org/10.1177/17411432231155730

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

Xia, J., & O'Shea, C. (2022). To what extent does distributed leadership support principal instructional leadership? Evidence from TALIS 2013 data. *Leadership and Policy in Schools*. Scopus. https://doi.org/10.1080/15700763.2022.2056059

Yammarino, F. J., Dionne, S. D., Uk Chun, J., & Dansereau, F. (2005). Leadership and levels of analysis: A state-of-the-science review. *The Leadership Quarterly*, *16*(6), 879–919. https://doi.org/10.1016/j.leaqua.2005.09.002

Zieger, L., Sims, S., & Jerrim, J. (2019). Comparing teachers' job satisfaction across countries: A multiple-pairwise measurement invariance approach. *Educational Measurement: Issues and Practice*, *38*(3), 75–85. https://doi.org/10.1111/emip.12254

Appendix

Table 1

List of articles included in the systematic literature review of peer-reviewed journal articles in which TALIS data were used to study school leadership

ID	Authors	Title	Journal	Year
1	Birasnav M; Gantasala S B; Gantasala V P; Singh A;	Total quality leadership and organizational innovativeness: the role of social capital development in American schools	Benchmarking	2022
2	Kılınç A Ç; Polatcan M; Turan S; Özdemir N;	Principal job satisfaction, distributed leadership, teacher-student relationships, and student achievement in Turkey: a multilevel mediated-effect model	Irish Educational Studies	2022
3	Ólafsson R F; Hansen B;	Characteristics of the Authority Basis of Icelandic Compulsory School Principals in Comparison to Other TALIS Countries	Education Sciences	2022
4	Xia J; O'Shea C;	To What Extent Does Distributed Leadership Support Principal Instructional Leadership? Evidence from TALIS 2013 Data	Leadership and Policy in Schools	2022
5	Bellibaş M Ş; Gümüş S; Liu Y;	Does school leadership matter for teachers' classroom practice? The influence of instructional leadership and distributed leadership on instructional quality	School Effectiveness and School Improvement	2021
6	Eryilmaz N; Sandoval Hernandez A;	Improving cross-cultural comparability: does school leadership mean the same in different countries?	Educational Studies	2021
7	Fackler S; Malmberg L E; Sammons P;	An international perspective on teacher self-efficacy: Personal, structural and environmental factors	Teaching and Teacher Education	2021

ID	Authors	Title	Journal	Year
8	Holzberger D; Prestele E;	Teacher self-efficacy and self-reported cognitive activation and classroom management: A multilevel perspective on the role of school characteristics	Learning and Instruction	2021
9	Liu S; Keeley J W; Sui Y; Sang L;	Impact of distributed leadership on teacher job satisfaction in China: The mediating roles of teacher autonomy and teacher collaboration	Studies in Educational Evaluation	2021
10	Liu Y; Bellibaş M Ş; Gümüş S;	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 and Leadership	2021
11	Liu Yan;	Contextual influence on formal and informal teacher leadership	International Journal of Educational Research Open	2021
12	Ning B;	Principals' time allocation in the context of Shanghai school administration	Asia Pacific Journal of Education	2021
13	O'Shea Cai;	Distributed leadership and innovative teaching practices	International Journal of Educational Research Open	2021
14	Printy S; Liu Y;	Distributed Leadership Globally: The Interactive Nature of Principal and Teacher Leadership in 32 Countries	Educational Administration Quarterly	2021
15	Veletić J; Olsen R V;	Developing a shared cluster construct of instructional leadership in TALIS	Studies in Educational Evaluation	2021
16	Veletić J; Olsen R V;	Exploring school leadership profiles across the world: a cluster analysis approach to TALIS 2018	International Journal of Leadership in Education	2021
17	Yang Y; Qin L; Ning L;	School Violence and Teacher Professional Engagement: A Cross-National Study	Frontiers in Psychology	2021

ID	Authors	Title	Journal	Year
18	Berkovich I; Bogler R;	The relationship between school leadership standards and school administration imperatives: an international perspective	School Leadership and Management	2020
19	Brezicha K F; Ikoma S; Park H; LeTendre G K;	The ownership perception gap: exploring teacher job satisfaction and its relationship to teachers' and principals' perception of decision-making opportunities	International Journal of Leadership in Education	2020
20	Ceylan E; Özdogan Özbal E;	The Effects of Extrinsic and Intrinsic Factors on Teachers' Job Satisfaction in TALIS 2018	International Online Journal of Primary Education	2020
21	Çoban Ö; Atasoy R;	Relationship between distributed leadership, teacher collaboration and organizational innovativeness	International Journal of Evaluation and Research in Education	2020
22	Fackler S; Sammons P; Malmberg L E;	A comparative analysis of predictors of teacher self-efficacy in student engagement, instruction and classroom management in Nordic, Anglo-Saxon and East and South-East Asian countries	Review of Education	2020
23	Kim T; Lee Y;	Principal instructional leadership for teacher participation in professional development: evidence from Japan, Singapore, and South Korea	Asia Pacific Education Review	2020
24	Liu Y;	Focusing on the Practice of Distributed Leadership: The International Evidence From the 2013 TALIS	Educational Administration Quarterly	2020
25	Lopes J; Oliveira C;	Teacher and school determinants of teacher job satisfaction: a multilevel analysis	School Effectiveness and School Improvement	2020
26	Wang K; Li Y; Luo W; Zhang S;	Selected Factors Contributing to Teacher Job Satisfaction: A Quantitative Investigation Using 2013 TALIS Data	Leadership and Policy in Schools	2020
27	Zoller K; Bacskai K;	Teacher Work and Job Satisfaction among Romanian Lower Secondary Teachers	Central European Journal of Educational Research	2020

ID	Authors	Title	Journal	Year
28	Cooc N;	Teaching students with special needs: International trends in school capacity and the need for teacher professional development	Teaching and Teacher Education	2019
29	García Torres D;	Distributed leadership, professional collaboration, and teachers' job satisfaction in U.S. schools	Teaching and Teacher Education	2019
30	Huang J; Tang Y; He W; Li Q;	Singapore's School Excellence Model and student learning: evidence from PISA 2012 and TALIS 2013	Asia Pacific Journal of Education	2019
31	Liu Y; Werblow J;	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	2019
32	Madero C;	Secondary teacher's dissatisfaction with the teaching profession in Latin America: the case of Brazil, Chile, and Mexico	Teachers and Teaching: Theory and Practice	2019
33	Ortega-Rodríguez P J;	School Autonomy in France According to TALIS 2013: The Importance of Educational Leadership	Turkish Online Journal of Educational Technology - TOJET	2019
34	Bellibas M S; Liu Y;	The effects of principals' perceived instructional and distributed leadership practices on their perceptions of school climate	International Journal of Leadership in Education	2018
35	García Torres D;	Distributed leadership and teacher job satisfaction in Singapore	Journal of Educational Administration	2018
36	Han I; Byun S Y; Shin W S;	A comparative study of factors associated with technology-enabled learning between the United States and South Korea	Educational Technology Research and Development	2018
37	Liu Y; Bellibas M S; Printy S;	How school context and educator characteristics predict distributed leadership: A hierarchical structural equation model with 2013 TALIS data	Educational Management Administration and Leadership	2018

ID	Authors	Title	Journal	Year
38	Sun A; Xia J;	Teacher-perceived distributed leadership, teacher self-efficacy and job satisfaction: A multilevel SEM approach using the 2013 TALIS data	International Journal of Educational Research	2018
39	Bellibas M S; Liu Y;	Multilevel analysis of the relationship between principals' perceived practices of instructional leadership and teachers' self-efficacy perceptions	Journal of Educational Administration	2017
40	Gil-Flores J;	The Role of Personal Characteristics and School Characteristics in Explaining Teacher Job Satisfaction	Revista de Psicodidactica	2017
41	Urick A; Bowers A J;	Assessing International Teacher and Principal Perceptions of Instructional Leadership: A Multilevel Factor Analysis of TALIS 2008	Leadership and Policy in Schools	2017
42	Angnakoon P ; Allen J M;	Exploring secondary school teachers' constructivist beliefs using TALIS 2013	Turkish Online Journal of Educational Technology	2016
43	Fackler S; Malmberg L E;	Teachers' self-efficacy in 14 OECD countries: Teacher, student group, school and leadership effects	Teaching and Teacher Education	2016
44	Gumus E ; Bellibas M S;	The effects of professional development activities on principals' perceived instructional leadership practices: multi-country data analysis using TALIS 2013	Educational Studies	2016
45	Lenskaya E; Brun I;	Are principals of russian schools ready for transformational leadership?	Voprosy Obrazovaniya / Educational Studies Moscow	2016
46	Park J H; Ham S H;	Whose perception of principal instructional leadership? Principal-teacher perceptual (dis)agreement and its influence on teacher collaboration	Asia Pacific Journal of Education	2016

ID	Authors	Title	Journal	Year
47	Sans-Martín A; Guàrdia Olmos; J; Triadó-Ivern X M;	Educational leadership in Europe: A transcultural approach	Revista de Educacion	2016
48	Ham S H; Duyar I; Gumus S;	Agreement of self-other perceptions matters: Analyzing the effectiveness of principal leadership through multi-source assessment	Australian Journal of Education	2015
49	Ham S-H; Kim R Y;	The Influence of Principals' Instructional Leadership on Teachers' Use of Autonomy-Supportive Instruction: An Analysis of Three Asia-Pacific Countries	The Asia - Pacific Education Researcher	2015
50	Zhou Y;	The Relationship Between School Organizational Characteristics and Reliance on Out-of-Field Teachers in Mathematics and Science: Cross-National Evidence from TALIS 2008	Asia-Pacific Education Researcher	2014
51	Duyar I; Gumus S; Bellibas M S;	Multilevel analysis of teacher work attitudes: The influence of principal leadership and teacher collaboration	International Journal of Educational Management	2013
52	Gumus S; Bulut O; Bellibas M S;	The Relationship between Principal Leadership and Teacher Collaboration in Turkish Primary Schools: A Multilevel Analysis	Education Research and Perspectives	2013
53	Gumus S;	The Effects of Teacher- and School-Level Factors on Teachers' Participation in Professional Development Activities: The Role of Principal Leadership	Journal of International Education Research	2013
54	Schleicher A;	Lessons from the World on Effective Teaching and Learning Environments	Journal of Teacher Education	2011
55	Piwowarski R;	Preconditions for effective teaching (in the light of data from the TALIS 2008 project - polish perspective)	New Educational Review	2010

Part II

Articles

I

Article 1

Veletić, J., & Olsen, R. V. (2021). Developing a shared cluster construct of instructional leadership in TALIS. *Studies in Educational Evaluation*, 68, 100942. https://doi.org/10.1016/j.stueduc.2020.100942.

Status: Published

Studies in Educational Evaluation 68 (2021) 100942

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

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 underrepresented construct. Furthermore, instructional leadership can be perceived differently by those who are led and those who are leaders (Urick & 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 (J. Veletić).

https://doi.org/10.1016/j.stueduc.2020.100942
Received 18 May 2020; Received in revised form 28 September 2020; Accepted 23 October 2020
Available online 10 November 2020

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)[1] .

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., 2007Krüger, Witziers, & Sleegers, 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.

[1] 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 (Urick & 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, Yalçin, & 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 trough 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[2] 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 multilevel 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 there 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 form 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, 1998-2017Muthén & 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

[2] International Standard Classification of Education (ISCED, 1997).

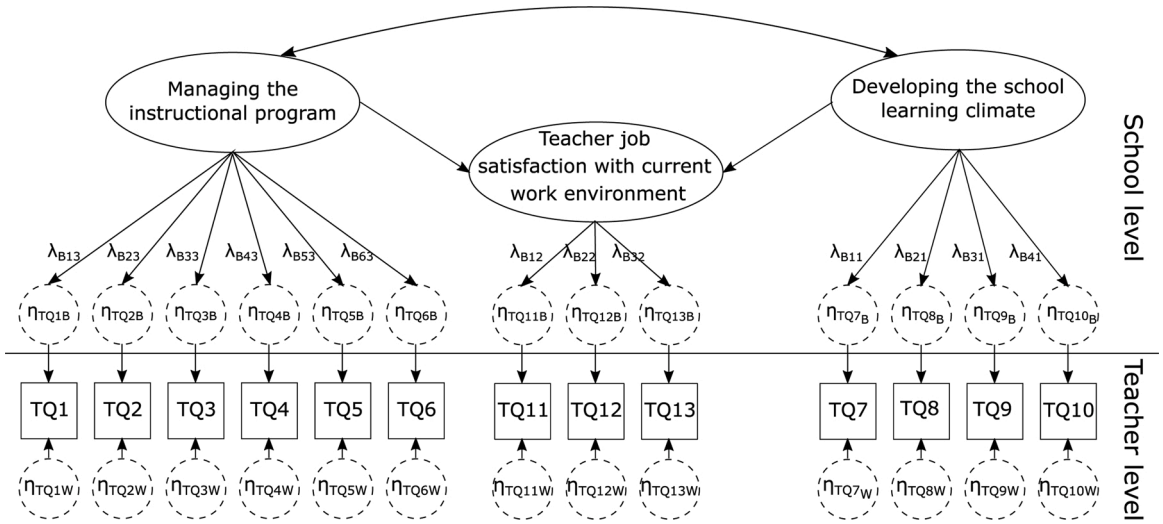


Fig. 1. Hypothesized model of instructional leadership dimensions—*managing the instructional program* and *developing the school learning climate*—and their association with school-level *teacher job satisfaction with current work environment.*
Note 1. At the teacher level, all items correlate. For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

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

Item	Item wording	Original TALIS code
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.]		
TQ1	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
How strongly do you agree or disagree with the following statements about this school?		
TQ7	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
[Finally] We would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?		
TQ11	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

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 (χ^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 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 ($SRMR_b$). The $SRMR_b$ 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[3] 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 (SRMRb = .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^2 not statistically different from zero. This result is not consistent with most 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.

[3] For more, see Appendix A.

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
χ2	4.774	9.084	8.248	6.325
df	8	8	8	8
SRMRb	0.027	0.044	0.027	0.029
β (S.E.)	−0.04 (0.12)	−0.03 (0.11)	−0.10 (0.09)	−0.00 (0.11)
R^2	0.002	0.001	0.010	0.000

***p < .001; **p < .01; *p < .05.
Note. The regression coefficients are standardized.

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 $SRMR_b$ of .080 and .046, respectively (see Table 5). The model fits in Denmark and Finland are somewhat higher but still having a $SRMR_b$ 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
χ2	18.188*	10.567	33.048***	18.883*
df	9	9	9	9
SRMRb	0.094	0.096	0.080	0.046
Managing the instructional program				
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)
Residuals				
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

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 $SRMR_b$ 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 $SRMR_b$ values obtained are .000 in Denmark, .004 in Finland, .021 in Sweden, and .025 in Norway.[4] 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 a particular school) are important indicators of teachers' shared perception of being satisfied with working environment

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

	Denmark	Finland	Sweden	Norway
χ2	2.612	2.627	3.733	2.739
df	3	2	2	2
SRMRb	0.046	0.023	0.027	0.018
Developing the school learning climate by				
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)
Residual variances				
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

[4] See Appendix B.

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
χ2	42.155*	16.072	45.699**	81.036***
df	26	26	26	27
SRMRb	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.

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

	Denmark	Finland	Sweden	Norway
χ2	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
χ2	34.900	26.01	64.349***	33.685
df	28	27	29	27
SRMRb	0.107	0.097	0.103	0.055
β (S.E.)	0.181 (0.152)	0.474*** (0.133)	0.167 (0.111)	0.415*** (0.096)
R^2 (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
χ2	26.738*	18.817	19.234	22.661
df	15	15	15	15
SRMRb	0.093	0.044	0.045	0.057
β (S.E.)	0.553*** (0.128)	0.517*** (0.084)	0.505*** (0.095)	0.609*** (0.072)
R^2 (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.[5] 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 asses, 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*"—(β1)—and Between "*Developing the School Learning Climate*" and "*Teacher Job Satisfaction With Current Work Environment*"—(β2).

	Denmark	Finland	Sweden	Norway
χ2	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)
β1 (S.E.)	0.214* (0.083)	0.242*** (0.064)	0.199*** (0.054)	0.255*** (0.046)
β2 (S.E.)	0.109* (0.047)	0.335*** (0.064)	0.162** (0.056)	0.256*** (0.046)
R^2 (S.E.)	0.103 (0.078)	0.285** (0.099)	0.121 (0.065)	0.250** (0.080)

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

[5] 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.

Acknowledgments

Jelena Veletić and Rolf Vegar Olsen are part of the European training network OCCAM. This project has received funding from the European Union's Horizon 2020 and innovation program under the Marie-Sklodowska-Curie grant agreement number 765400.

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
Managing the instructional program								
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" rd Errors

	Denmark	Finland	Sweden	Norway
χ2	0.000	0.036	8.912**	7.276**
df	1	1	1	1
SRMRb	0.000	0.004	0.021	0.025
Teacher Job Satisfaction with current work environment				
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)
Residuals				
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
χ2	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, 106*. 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., Yalçin, 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.

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.

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 anf 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.

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.

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. *NASSP 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 ans assessment in education*. Norway. Paris: 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/oecdreviewonevaluationandassessmentframeworksforimprovingschooloutcomes.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'as, 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., Spybrook, 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.
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.
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.

Article 2

Veletić, J., Price, E.H., & Olsen, R. V. (under review). Teachers' and principals' perception of school climate: The ole of principals' leadership style in organizational quality. *Educational Assessment, Evaluation and Accountability.*

Status: Re-submitted

II

Teachers' and Principals' Perceptions of School Climate: The Role of Principals' Leadership Style in Organizational Quality

Jelena Veletić

Centre for Educational Measurement, University of Oslo, Norway,

jelena.veletic@cemo.uio.no

https://orcid.org/0000-0002-3240-9674

Heather E. Price

Loyola University Chicago, USA,

hprice2@luc.edu

https://orcid.org/0000-0001-5359-3231

Rolf Vegar Olsen

Centre for Educational Measurement, University of Oslo, Norway,

r.v.olsen@cemo.uio.no

https://orcid.org/0000-0002-9621-4083

Acknowledgments

Funding: Jelena Veletić and Rolf Vegar Olsen were part of the European training network OCCAM. This project has received funding from the European Union's Horizon 2020 and innovation program under the Marie-Sklodowska-Curie grant agreement number 765400.

Abstract

In this paper we report on teachers and principals' shared perceptions regarding beliefs, rules, trust, and encouragement of new initiatives. Collectively, these are aspects of leadership for learning (LFL) describing an overall shared climate in schools. We demonstrate how these perceptions on school climate differ across teachers and principals within and across countries. Moreover, we report how different perceptions of school climate are associated with leadership style. We analyze data from 37 countries that participated in the last cycle of the Teaching and Learning International Survey (TALIS) in 2018. To build the measurement model, we employ multigroup multilevel confirmatory factor analysis, whereas multivariate linear regression is used to inspect associations. Overall, principals and teachers differ in their views of school climate. In the majority of the countries, principals report stronger school climate than teachers. We further confirm these perceptual differences between teachers and principals by separately studying the relationships between teacher perceived school climate and principal perceived school climate with relevant leadership variables. In the entire sample, we find that principals perceptions of school climate are more strongly and consistently associated with leadership in schools. This relationship is particularly stable for distributed leadership. In the entire sample, leadership styles are weakly positively correlated with teacher perceptions of school climate too, however this association is less pronounced and less stable within individual countries. The analyses conducted within countries revealed that the distributed leadership rather than instructional leadership shapes teachers' perceptions of school climate. More discussion is presented on the need for alignment between different perceptions of school climate and leadership styles in the overall organizational quality.

Keywords: *Leadership for Learning, School Climate, teachers' and principals' perceptions, TALIS 2018, Multigroup Multilevel Confirmatory Factor Analysis (MG-MCFA)*

Introduction

Educational research emphasizes a tight connection between school leadership and school climate (Griffith, 1999; Kelley, 2005; Kozlowski & Doherty, 1989). While there is currently limited empirical evidence about the nature of this association, it seems intuitive to suggest that a favorable climate can facilitate effective leadership and vice versa. School climate defined as *a shared perception* of behaviors, work environment, and organizational life (Ashforth, 1985; Hoy, 1990; Peterson & Spencer, 1990) constitutes a crucial factor in fostering teaching and instruction, supporting teachers' and students' development, and promoting healthy relationships, which are essential for successful learning (Cohen et al., 2009; Grazia & Molinari, 2020; Thapa et al., 2013). Assessing the perspectives of teachers and principals in relation to these shared aspects of school climate is a key measure of effective leadership (Brezicha et al., 2020; Park & Ham, 2016). The reason is that the school climate acts as a bridge between leadership and learning in schools. Building this bridge occurs by indirectly fostering working conditions, caring about teachers' well-being, and supporting instructional practices (Burkhauser, 2017; Ladd, 2009; Sims, 2019).

Accordingly, Ogawa and Bossert (1995) conceptualize leadership as an organizational quality that travels through the networks of actors and roles that constitute an organization. Furthermore, Otero (2019) describes Leadership for learning (LFL) as a system of relationships between principals, teachers, students, families, and communities. Although certain aspects of LFL can be achieved individually, for example, by principals or teachers, many of these aspects are only achievable jointly through the network of interactions between school stakeholders (MacBeath & Dempster, 2008; Pietsch et al., 2019). Such a system requires constant communication about learning that further fosters an environment of collaboration, trust, and dialog. Despite constant communication, common goals, and joint activities, the perceptions likely differ between teachers and principals due to different roles and hierarchical positions (Bandura, 1988; Ramsey et al., 2016). Only a few articles investigate how larger perceptual differences regarding the aspects of school leadership are associated with poor teacher collaboration (Park & Ham, 2016) and lower teacher job satisfaction (Brezicha et al., 2020).

Døjbak Haakonsson et al., (2008) argue that leadership and climate should be in

harmony in order to promote the organizational environment effectively. To better understand how the combined characteristics of leadership and school environment impact organizational quality, we analyze the degree to which school climate as measured from teachers' and principals' perspectives differ. Moreover, we also examine the association between different perspectives on school climate and leadership styles.

We do not discount that different perceptions of school climate can coexist in healthy learning environments too, nor are they necessarily destructive. For example, Ramsey et al. (2016) found that respondents give lower ratings to school climate dimensions that are closely related to their own behaviors because of either greater awareness or a more critical perspective. Moreover, the organizational literature in general assumes that leaders have tendency to overestimate their performance (Bass & Yammarino, 2008), whereas followers' ratings are more likely to be influenced by their personal experiences with leaders (D. J. Brown & Keeping, 2005). By developing a comparable measure of school climate from both teacher and principal perspectives at the level of school, we investigate these differences in perceptions of school climate.

Our findings add to the research about the conceptual linkage between climate and leadership in schools. Importantly, we establish a comparable measure of school climate between teachers and principals at the level of school. By using these measures, we demonstrate how principals and teachers differ in their perception of school climate. Lastly, we examine the association between both the teacher and principal reported school climate and school leadership as reported by principals across 37 countries. Overall, the results show a tight connection between climate and leadership in schools and their joint contribution in shaping the overall organizational quality.

Theoretical Background

School Climate

School climate refers to shared perceptions of the work environment and behaviors (Ashforth, 1985; Hoy, 1990). In the organizational literature, climate represents an internal distinguishing characteristic of an organization that influences the behaviors of its members (Woodman & King, 1978). The same line of research emphasizes that

“climate is external to the individual, yet cognitively the climate is internal to the extent that it is affected by individual perceptions” (Woodman & King, 1978, p. 818). The “commonality of perceptions” and homogeneity within organizations represent a critical attribute that differentiates climate from other organizational variables (Drexler, 1977; Woodman & King, 1978).

In the education literature, students, school personnel, and parents’ experiences of school life socially, emotionally, civically, ethically, and academically represent the school climate (Thapa et al., 2013). Similarly, Grazia and Molinari (2020) describe the moral, relational, and institutional aspects of school life as school climate dimensions. Therefore, school climate represents a broadly scoped quality and character of school life. It stands as a group phenomenon that includes norms, values, and expectations that support people (Cohen et al., 2009).The commonality of perceptions (Van Vianen et al., 2011; Woodman & King, 1978) and the teacher–principal relationship (Barnett & McCormick, 2004; Price, 2012; Van Maele & Van Houtte, 2015) represent an important attribute of organizational climate. Moreover, a positive school climate is determined by the presence of trustworthy relationships between school stakeholders which is often cultivated by the principal (Kutsyuruba et al., 2016). Thus, by establishing and maintaining positive school climate and healthy working environment, the school leadership shapes teacher and student outcomes (Dutta & Sahney, 2016; Özdemir et al., 2022; Sebastian & Allensworth, 2012). As such, positive climate also represents an indicator of leadership effectiveness.

A good school climate has multiple benefits, influencing students’ affective and cognitive outcomes, such as learning and well-being (Gustafsson & Nilsen, 2016; Hoy et al., 2006; Kutsyuruba et al., 2015; Scherer & Nilsen, 2016) and also teachers’ outcomes, such as beliefs, commitment, and engagement (Collie, 2012; Collie et al., 2011; Dickhäuser et al., 2021; Muijs & Reynolds, 2002). Higher self-efficacy and job satisfaction of teachers are associated with a better school climate (Aldridge & Fraser, 2016; Collie, 2012; Katsantonis, 2020). Furthermore, school climate enhances students’ self-concept (Coelho et al., 2020), cognitive engagement (Yang et al., 2018), and life satisfaction (Suldo et al., 2013; Zullig et al., 2011). It is also an inevitable factor for successful learning (Cohen, 2013; Cohen et al., 2009; Sherblom et al., 2006).

From a measurement perspective, researchers recognize the multidimensionality of the school climate construct across multiple studies (Grazia & Molinari, 2020; Lenz et al., 2021; Wang & Degol, 2016; Zullig et al., 2010). In their systematic review of the literature on school climate measures, Lenz et al. (2021) identified nine studies conceptualizing school climate as a multidimensional construct. Within these nine studies, 27 subscales relate to interpersonal relationships between school stakeholders emphasizing the social character of school climate (Lenz et al., 2021). In Wang and Degol (2016), which seems to be the most popular conceptualization, school climate is distinguished into four domains (academic, community, safety, and institutional environment) that are further subdivided into 13 dimensions. The academic, community, safety, and institutional environment domains refer to the 1) academic atmosphere, leadership, professional development, and instruction, 2) interpersonal relationships between school members, 3) physical and emotional safety and order and discipline, and 4) the physical and structural organization of the school and resource availability associated with teaching and learning, respectively (Wang & Degol, 2016).

In TALIS, school climate is represented by several measures derived from sets of questions in the school questionnaire (academic pressure, parent–community involvement, student delinquency scale, lack of resources and personnel), the teacher questionnaire (classroom disciplinary climate and student–teacher relations), or both (participation of stakeholder measure) (Ainley & Carstens, 2018). In addition, both questionnaires in TALIS 2018 contain numerous identical stand-alone items (teacher–teacher trust, common teaching beliefs, climate of shared rules, and teacher initiative). Therefore, TALIS does not provide a comprehensive measure of overall school climate. Instead, TALIS includes various scales that rather partially represent specific aspects of the broader school climate construct. Thus, by utilizing stand-alone items we seek to provide an overall climate measure that captures the shared aspects of school environment (shared beliefs, shared rules, shared trust, shared initiatives). In addition, because the items were included in both teacher and principal questionnaires in TALIS, we analyzed the extent to which perceptions of these shared characteristics differ between teachers and principals. Such insights provide important knowledge about the theoretical aspects of the tight connection between school climate and leadership for learning as an organizational quality (Ahn et al., 2021)

Teachers' and Principals' Perceptions of School Climate

The majority of school climate research relies on a single perspective, that is, principal, teacher, or student (Ramsey et al., 2016). Although multiple perspectives can provide a more accurate and comprehensive account of the school environment (Park & Ham, 2016; Thapa et al., 2013; Veletić & Olsen, 2021b) those are not frequently reported. For instance, students, teachers, and parents rate differently the aspects of school climate related to connectedness, safety, academic emphasis (Price, 2016; Ramsey et al., 2016), bullying (Stockdale et al., 2002), leadership (Park & Ham, 2016), and overall climate (Mitchell et al., 2010). Different perceptions of the same phenomena are due to numerous factors, including individuals' organizational position, experience, knowledge, and self-awareness, or methodological aspects, such as whether the respondents are asked to rate themselves or others (Atwater et al., 1998; Braddy et al., 2014; Fisher & Katz, 2000).

As such, the perceptions of teachers and principals within the same school are being recognized as important, but empirical evidence about their coexistence is scarce (Moye et al., 2005; Park & Ham, 2016; Price, 2012). According to some authors, a total congruence between principals and teachers perceptions is an ideal, but hardly (if ever) achievable in practice (Braddy et al., 2014). Hence, we represent this (in) congruence through reporting the climate as perceived by teachers and principals. Recognizing such differences may be vital to understand behaviors within an organization and gain insights into organizational quality and teacher–principal dynamics. Moreover, understanding the differences in perception between principals and teachers regarding the school environment can offer a more precise representation of the effectiveness of school leadership and, ultimately, the quality of the organization (Park & Ham, 2016).

For instance, Park and Ham (2016) utilized TALIS 2008 data and found that the gap in perception of instructional leadership between teachers and principals negatively associated with teacher engagement in collaborative activities and collegial interactions in Australia, Malaysia, Korea, and Turkey. Moreover, using the same sample Ham et al. (2015) established a negative association between the principal-teacher gap regarding the instructional leadership and teacher self-efficacy. Brezicha et al., (2019) examined the teacher and principal perceptions of teachers' involvement in decision making and teachers job satisfaction. Using TALIS 2013 data across 29 countries, the authors

demonstrated large differences between teacher and principal reports. The association between these gaps in reporting and teacher job satisfaction in the US sample was negative and significant.

Gaps are not necessarily counter-productive. For instance, Brezicha et al., (2019) found that even in the presence of the gaps, the opportunity to collaborate improved teacher job satisfaction, adding to the argument about the importance of constant communication and good relationships between teachers and principals. Ahn et al., (2021) using TALIS 2018, demonstrated that collective teacher perceptions and principal perceptions of leadership tasks, were not correlated globally which was interpreted as concerning given that leadership for learning advocates that collective efforts of school members are crucial for effective leadership and ultimately school improvement. Similarly, Price (2012) suggests that cultivating positive relationships between school members, particularly teachers and principals, can enhance the school climate and ultimately align their perceptions of the environment. Finally, Bellibas et al. (2017) showed that principals perceived distributed and instructional leadership are significant predictors of mutual respect in schools (one aspect of school climate). However, they did not find a correlation between leadership style and school delinquency and violence (another aspect of school climate). These findings suggest that, indeed, school leadership appears to have a greater impact on teacher-related outcomes such as efficacy and job satisfaction (García Torres, 2019; S. Liu et al., 2021; Sun & Xia, 2018; Veletić & Olsen, 2021b) whereas the association with school climate might be less stable and dependent on the specific aspect of school climate being investigated. Thus, this study seeks to establish a comparable measure of school climate that relate to shared beliefs, rules, trust, and encouragement of new initiatives between teachers and principals which collectively embody what is considered effective leadership for learning.

Leadership for Learning

The roles, practices, and actions of principals and teachers in schools bridge leadership and learning (Hallinger & Heck, 2010; Leithwood & Mascall, 2008; Lovett & Andrews, 2011; Sims, 2019). Principals are responsible for setting the ground for teachers to achieve their full working potential. Principals are also fundamental in developing the school learning climate, managing instructional programs, and communicating high-

order goals through the school mission and vision (Hallinger, 2009, 2011). Leadership theory that emerged in the United Sates in the 1950s focused on principals' roles in shaping and nurturing high-quality instruction in schools. Such theories are commonly known as instructional leadership (Hallinger, 2015). However, over the years, perceptions and practices of leadership functions dispersed among other school members, allowing for a distributed and shared leadership practice (Day et al., 2016; Marks & Printy, 2003; Spillane et al., 2004). Although little is known about the shortcomings and inadequacies of distributed leadership practice (Harris, 2009), this approach to leadership was embraced by many and it became an advocated approach of leading schools. It allowed for more people in leadership roles, emphasizing the complex process of mutual influences and the importance of the context. Moreover, attention shifted from *instruction* to *learning*, which is particularly detectable in the LFL model that unites previously established models of leadership, mainly instructional and distributed approach (Bowers, 2020). Thus, leadership becomes more responsive to students as actors, connected to the broader community outside of the school, and less hierarchical (Dempster, 2019; Imig et al., 2019).

Data from TALIS have been extensively used to study leadership because it provides a comprehensive source across as many as 47 countries from both teacher and principal perspectives. Apart from being used to study teacher-principal agreement, TALIS data are extensively used to study distributed leadership (Çoban & Atasoy, 2020; García Torres, 2019; Kılınç et al., 2022; Liu, 2020; Liu et al., 2018), instructional leadership (Bellibas & Liu, 2017; Eryilmaz & Sandoval Hernandez, 2021; Gumus & Bellibas, 2016) or both conceptualizations simultaneously (Bellibas & Liu, 2018; Xia & O'Shea, 2022). There are several attempts in the literature where TALIS data are used to map the leadership for learning framework (Ahn et al., 2021; Bowers, 2020; Veletić & Olsen, 2021a).

Scholars proposed several LFL models, of which four are widely used in the literature: (1) the comprehensive assessment of LFL (CALL) study in the USA (Kelley & Halverson, 2012), (2) Murphy et al.'s (2007) research-based model and taxonomy of behaviors, (3) Hallinger's (2011) synthesis of literature, and (4) Boyce and Bowers' (2018) multilevel factor analysis. These models share the same fundamental concepts but

broadly capture LFL practice differently. The CALL study captures leadership practice and school cultures across five domains: focus on learning, monitoring teaching and learning, building nested learning communities, acquiring and allocating resources, and maintaining a safe and effective learning environment. Murphy et al.'s LFL model suggests eight dimensions of LFL: vision for learning, instructional program, curricular program, assessment program, communities of learning, resource acquisition and use, organizational culture, and social advocacy. Hallinger, in contrast, proposes four dimensions of the model of LFL: values leadership, leadership focus (vision and goals, academic structures and processes, and people), the leadership context, and leadership sharing. Lastly, Boyce and Bowers describe six factors at the teacher level (classroom control, teacher commitment, school influence, collegial climate, student attendance, and neighborhood context) and three at the school level (instructional leadership, management, and social environment).

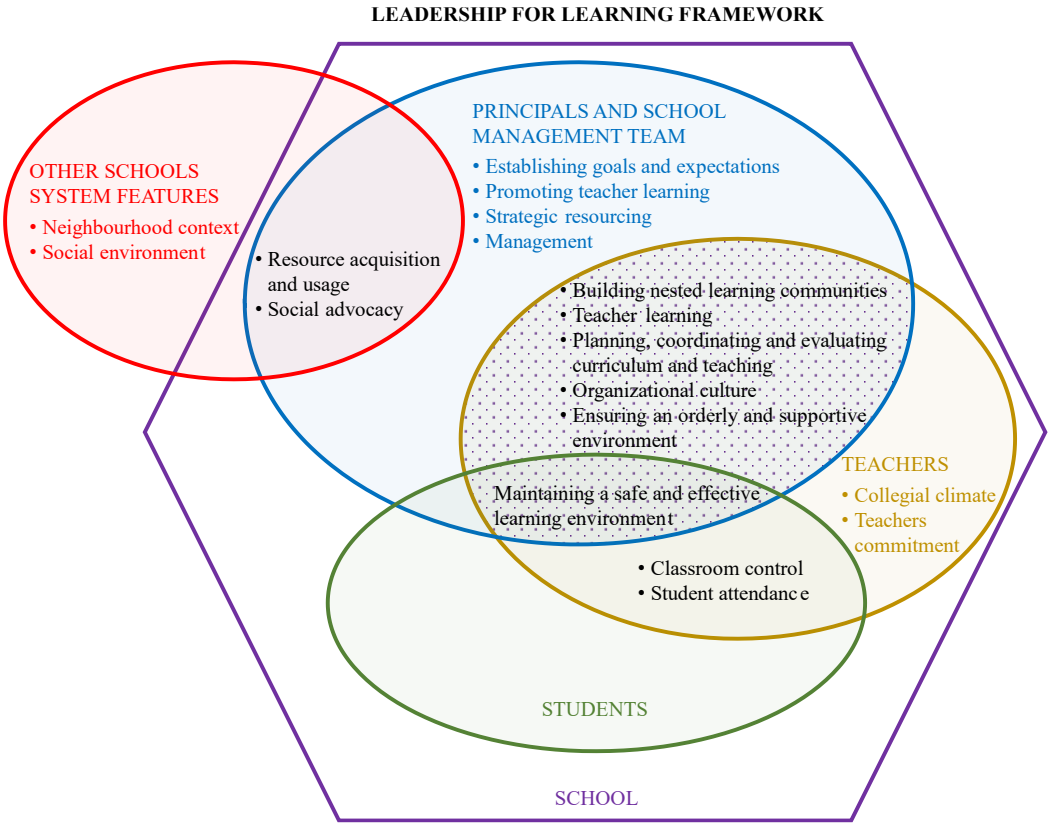
Significant overlaps exist between these LFL models. In Figure 1, we synthesize the LFL domains by combining the elements of the four above-mentioned models. Our framework (Figure 1) represents four main actors of LFL (represented in ovals): principals and school management team, teachers, students, and the system.[1] The purple hexagon divides actions inside and outside of the school. The figure further shows that certain LFL domains are achieved by one actor (e.g., principal or teachers only), whereas others (the intersecting parts) are achieved jointly, either by principals and teachers, principals, teachers, and students, or principals and stakeholders outside of the school. Figure 1 shows that joint efforts and shared perceptions are crucial for successful leadership and enhanced school climate. Therefore, in this article, we focus on school environment aspects that are achieved jointly by teachers and principals (dotted area of Figure 1). The overall framework of leadership for learning as presented in Figure 1 encompasses both instructional leadership, distributed leadership, and shared aspects of school climate as important indicators of quality of organization. The framework further clarifies how school leadership may be considered as a part of school climate, while also

[1] The figure does not show parents although they are important for certain LFL domains, e.g., student attendance.

emphasizing how school climate may be considered as integral part in school leadership.

Figure 1

LFL Framework Synthesizing Domains Proposed by the Four Most Common LFL Conceptualizations



Control Variables

School-level factors, such as school size, location, and composition, shape the school environment directly or indirectly (DiPietro et al., 2015; Goldkind & Farmer, 2013; Koth et al., 2008; McCoy et al., 2013; Sulak, 2018). Analyzing the data from the Schools and Staffing Survey (SASS) in the United States, Shakeel and DeAngelis (2018) showed that private schools may have an advantage over public schools in the United States in the form of fewer restrictions on school climate and safety and more comfortable and trustworthy environment for students.

Teacher-level factors are also important, among which the association between teachers' years of experience and school climate is particularly intriguing. Students internationally report that schools with experienced teachers tend to have a good school

climate in the PISA study. The average number of years of experience among teachers had a significant, positive association with classroom disciplinary climate in several countries (Avvisati, 2018). Furthermore, Kalis (1980) showed that experienced teachers (more than 6 years of experience in the same school) perceive a less favorable school climate. These findings suggest an inconsistent or nonlinear association between teachers' years of experience and their perception of school climate.

Moreover, the average socioeconomic status (SES) for schools influences several variables reflecting school climate. However, findings are ambiguous, and consistent evidence of the importance of school SES does not exist (Armor et al., 2018; Marks, 2015). Lastly, school facilities and resources are found to be consistently significant (Akomolafe & Adesua, 2016; Greenwald et al., 1996; Uline & Tschannen-Moran, 2008). Taken together, these results indicate that models investigating school climate should consider school and teacher characteristics.

Present Study

In the present study, we examined the broader framework of leadership for learning as "an organization-wide practice" that goes beyond that of principal (p.1, Ahn et al., 2021). This framework not only emphasizes learning, but also encompasses other sources of leadership, "and paths and means by which leadership contributes to overall improvement including school climate" (p.8, Ahn et al., 2021). Therefore, first, we established and investigated a new measure of school climate by combining a set of parallel items included in both the principal and teacher questionnaires of the TALIS 2018 survey implemented in 37 countries. This measure represents an overall measure of school climate and has an advantage over the existing sub-dimensions of school climate in the TALIS dataset as it allows for comparisons across principals and teachers. We use this new measure to examine the differences in perception of school climate across teachers and principals in the overall sample and within countries included in the final analyses. Moreover, we investigate the association between school climate as perceived by principals/teachers and leadership styles. Thus, we aim to answer the following research questions (RQ):

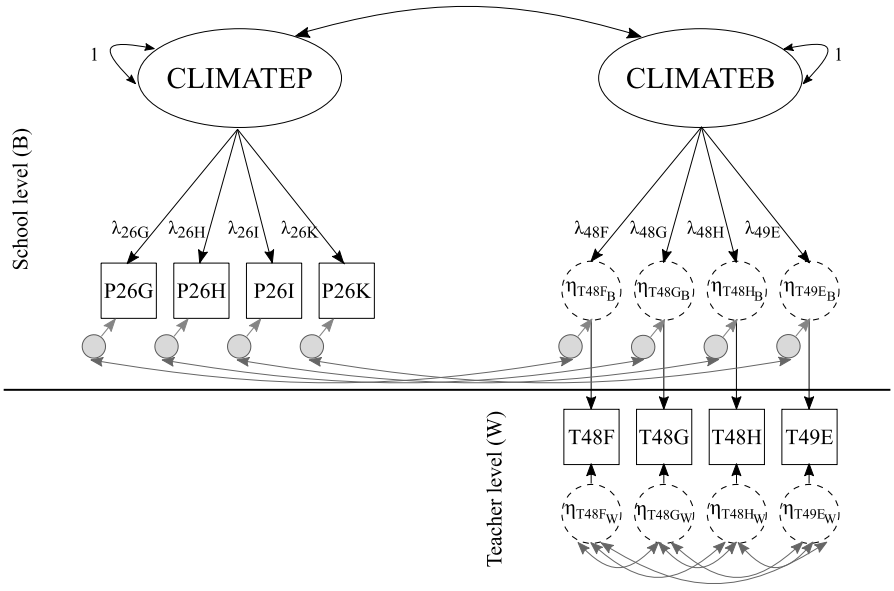
1. What are the measurement properties of the proposed school climate indicators based on teacher and principal reports?

2. Based on the newly proposed measures, to what extent do teachers' and principals' views of school climate differ?
3. To what extent is leadership style associated with school climate as perceived by principals?
4. To what extent is leadership style associated with school climate as perceived by teachers?
5. To what extent do features of the national context associate with the teacher and principal perceptions of school climate?

Figure 2 below displays the measurement model applied in this study. We modeled teacher responses in a multilevel setting with a saturated structure and factor structure at levels 1 and 2, respectively. Principal data are modeled at level 2, with correlated residuals among the same worded items from principal and teacher questionnaires (P26G-T48F . . . P26K-T49E).

Figure 2

Configural Model of Principal and Teacher Responses for School Climate



Note: The ovals represent latent constructs of school climate reported by principals (CLIMATEP) and teachers (CLIMATEB). The rectangles (P26G-P26K…T48F-T49E) represent observed variables, whereas the curves with arrowheads on both sides represent correlations. The shaded cycles represent correlated residuals. The dashed line cycles (η_{T48Fb} - η_{T49Eb} . . . η_{T48Fw} - η_{T49Ew}) represent latent variables at the between (b) and within (w) levels.

Methods

Data and Sample

The data for this study come from the third and most recent cycle of the TALIS study administered in 2018. TALIS is an international large-scale survey concerned with teaching and learning conditions, learning environments, and school leadership among others (Ainley & Carstens, 2018). In TALIS 2018, 48 countries or provinces participated in the core survey including teachers and principals from lower secondary education (ISCED level 2). TALIS 2018 set the minimum sample size at 20 teachers within each participating school and required a minimum sample of 200 schools from the national population. This two-level complex survey design implies that schools and teachers had unequal probability to be included in the final sample and creates a cluster structure in the dataset. In the analyses, we accounted for these deviations from simple sampling. For additional details about the sampling design in TALIS, we refer to the TALIS technical report (OECD, 2019).

In this study, we analyzed data only for countries, excluding provinces or cities, such as Alberta, Canada, and Ciudad Autónoma de Buenos Aires, Argentina. In addition, we excluded five countries (Italy, Singapore, Romania, Israel, and the Netherlands) due to systematically missing data on key items. According to the TALIS technical report, data from Australia for ISCED level 2 did not meet the standards for inclusion. Consequently, we also excluded this country. As a result, the final sample included 125,520 teachers clustered in 7384 schools from 37 countries. The average cluster size is 16.65 teachers per school within the country. Appendix 1 provides an overview of the final sample sizes per country, and Appendix 2 shows the basic descriptive statistics for the entire sample and each country separately.

Measures

Outcomes

Teacher Perception of School Climate (CLIMATEB). The school climate measure based on teacher responses (*CLIMATEB*) was assessed by teacher ratings of four statements as shown in Table 1. We modeled teacher responses using multilevel confirmatory factor analysis (MCFA) to obtain factor scores at the school level. The

modeling included factor structure at the school level and a fully saturated model at the teacher level, commonly referred to as a shared cluster construct (Stapleton et al., 2016). The reliability omega coefficients ranged from 0.849 in France (FRA)[2] to 0.972 in Kazakhstan (KAZ). Appendix 3 shows detailed information about model fit and reliability coefficients.

Principal Perception of School Climate (CLIMATEP). The school climate measure based on the reports of principals (*CLIMATEP*) was assessed by their ratings on the same four statements (see Table 1). Using confirmatory factor analysis (CFA), we modeled principal responses at the school level and extracted factor scores. The scale reliabilities were decent in most countries when the model worked, with the omega coefficient ranging from 0.625 in Japan (JPN) to 0.830 in United Arab Emirates (UAE). Appendix 3 provides details about model fit and reliability coefficients.

Table 1

Parallel Items Measuring School Climate

Item wording	**TQ**	**PQ**
How strongly do you agree or disagree with these statements as applied to this school? (1-strongly disagree; 2-disagree; 3-agree; 4-strongly agree)		
The school staff share a common set of beliefs in teaching and learning.	TQ48F	PQ26G
The school staff enforce rules for student behavior consistently throughout the school.	TQ48G	PQ26H
This school encourages staff to lead new initiatives.	TQ48H	PQ26I
Teachers can rely on one another.	TQ49E	PQ26K

Note: TQ: teacher questionnaire; PQ: principal questionnaire

[2] For a complete list of country codes, see Appendix 1.

Predictors

Instructional Leadership (T3PLEADS). The scale for instructional leadership was available directly from the TALIS dataset. The scale combines principal ratings on three items where principals indicated (on a 4-point Likert scale) how frequently they engaged with the following activities in the last 12 months: (1) "supporting co-operation among teachers to develop new teaching practices," (2) "ensuring that teachers take responsibility for improving their teaching skills," and (3) "ensuring that teachers feel responsible for their students' learning outcomes" (OECD, 2019). A higher score indicates stronger instructional leadership practice. As reported in the 2018 TALIS technical report (OECD, 2019), the scale achieved a metric level of invariance across countries and the omega reliability coefficient was high for all populations (excluding Hungary), ranging from 0.702 in Kazakhstan (KAZ) to 0.962 in Australia (AUS) (OECD, 2019).

Distributed Leadership (T3PLEADP). The scale for distributed leadership combines five items in the 2018 TALIS study. TALIS refers to this scale as participation among stakeholders (OECD, 2019). The measure combines principal ratings on a 4-point Likert scale indicating how much they (dis)agreed with the following: (1)–(3) "This school provides [staff], [parents], [students] with opportunities to actively participate in school decisions," (4) "This school has a culture of shared responsibility for school issues," and (5) "There is a collaborative school culture which is characterized by mutual support" (OECD, 2019). A higher score represents stronger distributed leadership in the school, that is, decision-making involves several people, and a strong culture of shared responsibilities and mutual respect can be observed. The scale is metrically invariant across countries with acceptable scale reliabilities in most countries, ranging from 0.599 in Japan (JPN) to 0.927 in the Russian Federation (RUS) (OECD, 2019).

Control Variables

In addition to the main independent variables, the final model controlled for several principal and school characteristics relevant to school climate. We carefully selected these variables to limit data loss due to systematically not administered questions about school and principal characteristics in certain countries. For example, several countries skipped questions about school location and level of formal teacher education

(e.g., New Zealand and Spain). Therefore, we did not include these two aspects as control variables, though they may be relevant to school climate in certain countries. In other cases, countries did not administer questions about principals' years of experience and private and public schools (e.g., Italy, Singapore, and Israel). Nevertheless, we argue that such factors influence the final model. Consequently, we excluded these countries from the analysis. Table 2 below shows the final list of control variables at the school level.

Table 2

List of Control Variables

Original variable name TALIS	**Variable name**	**Label**	**Coding**
NENRSTUD	School size	Number of enrolled students	Ordinal treated as continuous
TC3G12	Public	Publicly or privately managed school	0 = private 1 = public
TC3G17A	Language of students	% students [first language] diff. from instr. language	Ordinal treated as continuous
T3PLACPE	LACK_PESRON	Lack of pedagogical personnel	0 = not a problem 1 = a problem
T3PLACRE	LACK_RESOUR	Lack of resources	0 = not a problem 1 = a problem
TT3G11A	Teachers' experience	Teachers' average experience in this school	Continuous
TC3G01	Female	Gender of the principal	0 = male, 1 = female
TC3G04A	Principal experience	Principals' years of experience in this school	Continuous

Statistical Analysis

We estimated the main measurement and regression models using Mplus Version 8.4 (Muthén & Muthén, 2017) through the Rstudio package "MplusAutomation" (Hallquist & Wiley, 2018). To account for the possible non-normality of the data, we used the robust maximum likelihood (MLR) estimator. The MLR estimator is also used to

handle missing data. No variables had more than 5% missing values. We incorporated the final school weight for the analysis at the school level and the teacher and school weight for the multilevel analysis to account for unequal selection probabilities (Rutkowski et al., 2010). Due to the high complexity, we performed analyses in four steps as follows:

Step 1: We identified parallel items in the teacher and principal questionnaires regarding school characteristics closely related to school climate and modeled these items in separate CFAs for the two groups. As illustrated in Figure 2, the principal data are modeled at the school level, whereas the teacher measure is based on a multilevel model of a shared cluster construct (T. A. Brown, 2015; Kim et al., 2018; Stapleton et al., 2016). We used standard fit indices to evaluate the model fit: the chi-square (χ^2) with corresponding degrees of freedom (df), the root mean square error of approximation (RMSEA) close to 0.06 or below, the comparative fit index (CFI) close to 0.95, the Tucker–Lewis index (TLI) close to 0.95, and standardized root mean square residual at within and between level (SRMRw and SRMRb) close to 0.08 (Hu & Bentler, 1999). We allow for certain deviations from these criteria due to model complexity (Asparouhov & Muthen, 2018).

Step 2: We tested measurement invariance (MI) across respondents (teachers and principals), which itself consists of numerous steps. Establishing MI is a precondition for comparison across groups (Chen, 2008; Millsap, 2012; Rutkowski & Svetina, 2014). For a meaningful comparison of cluster means, the scalar level of invariance is necessary (Millsap, 2012). Because exact invariance is rarely achieved in practice (Byrne & Vijver, 2010; Rutkowski & Svetina, 2014; Zieger et al., 2019), certain authors suggested that constraining at least two fixed parameters across groups while freely estimating the remaining items is sufficient to compare latent means (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). To add to the complexity of model estimation in this article, the standard procedures for testing MI were not possible. The reason is that teachers and principals were at different hierarchical levels of the model, with teachers clustered in principals (schools). Therefore, we followed Kim et al.'s (2018) recommendations and used MCFA to test the invariance between teachers and principals at the school level. The focus of this article on the cluster (school) level supports our choice. We performed analyses on a pooled sample and for each country separately and evaluated models based on common guidelines for model fit evaluation and invariance testing (e.g., CFI $\geq$ 0.95,

RMSEA ≤ 0.08, SRMR ≤ 0.06, ΔCFI ≤ −0.010, ΔRMSEA ≤ 0.015, ΔSRMR ≤ 0.030) (Chen, 2008). Again, we allowed for deviations from common guidelines due to the complexity of the sample and models (Byrne et al., 1989; Marcoulides & Yuan, 2020; Marsh et al., 2004; OECD, 2019).

Step 3: We ran a school-level multivariate regression analysis to assess the association between school leadership and the climate as reported by teachers and principals while controlling for other school and principal characteristics, separately.

Step 4: We ran the final model on the pooled dataset with fixed effects for countries. This analysis provides us with an estimate of systematic variation in the climate measures across countries, thus informing us about the extent to which the climate measure as reported by teachers versus the climate measure as reported by principals relate to the system features of the countries.

Results

Appropriateness of the Multilevel Approach

To ensure that the items included in the model have substantial variability at the cluster level needed for multilevel modeling (Snijders & Bosker, 1999), we inspected the intraclass correlation coefficient 1 (ICC1) as a measure of agreement, and intraclass correlation coefficient 2 (ICC2) as a measure of clustering for teacher ratings of school climate for each country separately (see Appendix 3). The coefficients for all items in all countries were acceptable according to common guidelines (Geldhof et al., 2014; Stapleton et al., 2016), with ranges (ICC1) 0.062 (KAZ)–0.265 (NZL), 0.041 (KAZ)–0.345 (NZL), 0.065 (MLT)–0.236 (NZL), and 0.076 (USA)–0.248 (MEX) for items TT3G48F, TT3G48G, TT3G48H, and TT3G49E, respectively. The majority of the teachers in New Zealand were consistent in their ratings of school climate, with high ICC1 (> 0.20) across all items, followed by Swedish and Norwegian teachers. On the contrary, teachers in Kazakhstan did not agree with each other consistently, followed by teachers in Saudi Arabia, Latvia, Lithuania, Portugal, and Cyprus. Across all countries, the teachers showed the most agreement when responding to item TT3G49E (“Teachers can rely on one another”), with the highest ICC1 on average.

Evaluating the Measurement Models and Testing the Measurement Invariance of Teacher and Principal Ratings for School Climate

We tested the measurement properties of teacher and principal ratings of the newly established school climate scale to answer RQ1. According to standard fit indices, the MCFA model of teacher ratings of school climate (CLIMATEB) with the saturated structure at level 1 exhibited an excellent fit to the data for the entire sample ($\chi2$=14.986, df=2, CFI=0.999, TLI=0.991, RMSEA =0.007, SRMRw=0.001, SRMRb =0.029), and within each of national samples, according to standard fit indices. The CFA model of principal ratings of school climate (CLIMATEP) exhibited an excellent fit to the data for the entire sample ($\chi2$=13.227, df=2, CFI=0.976, TLI=0.929, RMSEA =0.028, SRMRw=0.027). When tested separately for each country, excellent model fit was exhibited in 22 countries (CFI > 0.095, TLI > 0.095, RMSEA < 0.08, SRMRw < 0.06). In eight countries, the model fit was acceptable with CFI, TLI, and SRMR within the recommended cut-offs and RMSEA above the recommended cut-off, though still below 0.1. In five countries, the model did not fit the data well (see Appendix 3).

With the school climate scale reported by principals and teachers now established at the school level, we proceeded to the MI testing across teachers and principals to provide evidence about the comparability of these two measures at the school level. First, we tested the MI on a pooled dataset where the configural and metric models across teachers and principals yield an excellent fit. The scalar model with constrained intercepts across two groups was also acceptable. However, the fit for this model was significantly lower than for the metric model, particularly regarding SRMRb (ΔCFI ≤ -0.010, ΔRMSEA ≤ 0.004, ΔSRMRw ≤ 0.000, ΔSRMRb ≤ 0.083). In the second step of MI testing, we performed analyses for each country separately. The configural model showed an excellent model fit in all countries. The metric model with constrained factor loadings across respondents also showed an acceptable fit in the majority of the countries. However, when we constrained intercepts to be equal across teachers and principals to establish scalar invariance, the model fit deteriorated significantly in most countries, with SRMRb > 0.10 (see Appendix 5 for the complete reports by country). We were unable to establish full scalar invariance across teachers and principals. Thus, we established the minimum requirements for partial invariance as recommended by certain authors (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). According to these authors, in addition

to the marker item loading fixed to 1 and intercept fixed to 0, at least one indicator must have invariant loadings and intercepts across the groups. Table 3 shows the final model fit of the partial scalar model across respondents, with a saturated structure at level 1.

Table 3

Partial Scalar Measurement Invariance Model across Teacher and Principal Responses at the School Level

Model fit estimates by country								**Correlation between latent factors at the school level**	
Model	χ^2	df	CFI	TLI	RMSEA	SRMRw	SRMRb	Estimate	*SE*
ARE	61.37	17	.995	.991	.017	.001	.08	.326***	.058
AUT	34.08	17	.995	.989	.015	.001	.082	.545***	.102
BEL	63.34	17	.983	.966	.023	.002	.093	.598***	.075
BGR	14.14	17	1	1	0	.001	.052	.290**	.109
BRA	22.69	17	.997	.994	.012	.001	.048	.263**	.101
CHL	14.26	17	1	1	0	.001	.039	.322**	.102
COL	24.95	17	.994	.988	.014	.001	.072	.455**	.132
CYP[a]	31.50	17	.99	.98	.023	.003	.137	.506**	.165
CZE	26.28	17	.997	.993	.013	.001	.068	.473***	.097
DNK	26.14	17	.992	.983	.016	.001	.058	.757***	.078
ESP	38.20	17	.993	.985	.013	.001	.062	.537***	.108
EST	24.23	17	.998	.996	.012	.001	.059	.488***	.102
FIN	24.94	17	.997	.993	.013	.001	.073	.391***	.100
FRA	60.71	17	.974	.948	.029	.002	.09	.340**	.118
GEO	20.37	17	.999	.998	.008	.001	.054	.310*	.132
HRV	29.44	17	.995	.99	.015	.001	.082	.518***	.141
HUN[a]	55.22	17	.989	.978	.026	.002	.106	.395***	.098
JPN	29.72	17	.996	.992	.015	.001	.074	.607***	.097
KAZ	44.01	17	.992	.983	.016	.002	.096	.346**	.105
KOR	28.43	17	.996	.992	.015	.002	.058	.253	.128
LTU[a]	43.01	17	.991	.981	.02	.002	.135	.294*	.134
LVA[a]	46.40	17	.986	.972	.027	.002	.1	.484***	.125
MEX	24.64	17	.996	.993	.012	.002	.071	.432***	.107
MLT[a]	31.95	17	.989	.979	.023	.001	.112	.371**	.129
NOR	25.70	17	.996	.993	.011	.001	.08	.600**	.090
NZL[a]	33.08	17	.983	.965	.021	.005	.143	.573***	.110
PRT	36.38	17	.992	.985	.018	.001	.085	.275*	.124
RUS[a]	39.70	17	.991	.981	.018	.001	.105	.613***	.101

SAU[a]	47.22	17	.986	.971	.026	.004	.113	.108	.171
SVK	21.17	17	.998	.996	.009	.001	.058	.551***	.109
SVN	30.92	17	.993	.987	.02	.002	.069	.567***	.122
SWE	21.59	17	.997	.994	.01	.002	.058	.760***	.065
TUR	29.11	17	.996	.992	.013	.002	.075	.229	.134
USA[a]	25.88	17	.99	.98	.014	.002	.113	.227	.157
VNM	36.54	17	.99	.98	.017	.001	.099	.028	.153
ZAF	25.59	17	.996	.991	.016	.003	.084	.423**	.134

Note: [a]The data did not fit the model (SRMRb > 0.1), and the results should be interpreted with caution

*, **, and *** denote significance level at 0.05, 0.01, and 0.001, respectively.

School Climate Reported by Principals and Teachers

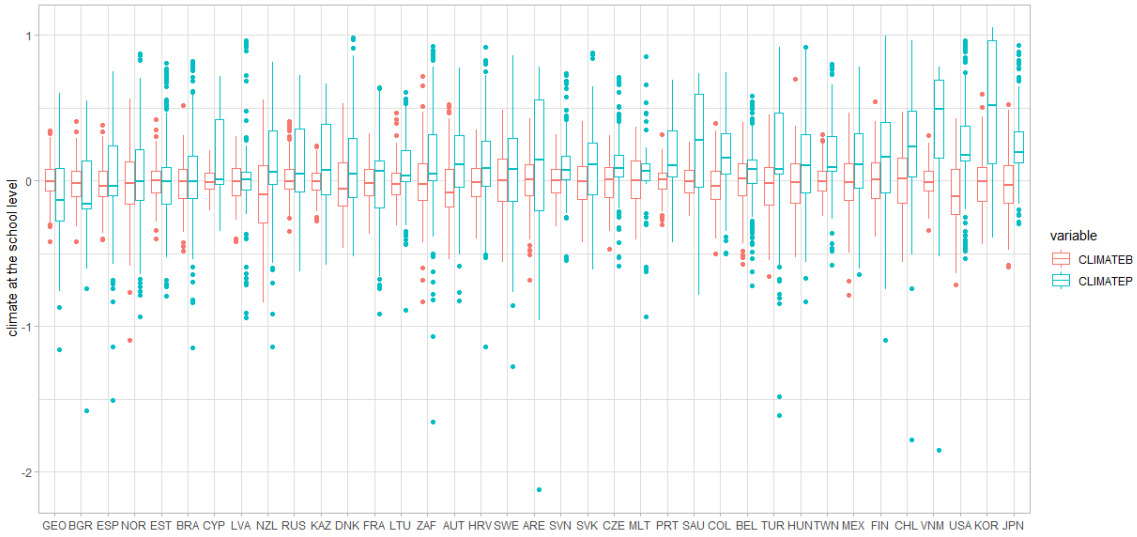
After establishing the partial invariance model, we extracted factor scores at the school level for the climate measure reported by teachers and principals addressing RQ2 (see Figure 3). The descriptive statistics show that principals across countries consistently reported a better climate than teachers did, except for Georgia (GEO) and Bulgaria (BGR), where we find the opposite. The differences in perception of school climate between principals and teachers were, on average, the widest in Korea (KOR), Viet Nam (VNM), and the United States (USA), whereas the narrowest average distance, close to zero, occurred in Bulgaria (BGR), France (FRA), Latvia (LVA), Estonia (EST), Malta (MLT), and Norway (NOR). Within countries, standard deviations for principal reports range from 0.2 in the Czech Republic (CZE) to 0.4 in Turkey (TUR). Given that the factor scores (at the school level) for teachers reflect an average measure across several teachers, the associated dispersions are, as expected, smaller, with standard deviation ranging from 0.084 in Kazakhstan (KAZ) to 0.029 in New Zealand (NZL) (for details, see Appendix 2, Table 12, and Table 13). We also find similar results in the entire sample as displayed in Figure 4. We will return to this issue as a potential limitation of the study.

The positive correlation between the climate reports by the principals and teachers is another interesting element showing partial congruence between the two groups across countries (see Table 3). The correlations were the highest in the Scandinavian countries, namely, Sweden (SWE), Denmark (DNK), and Norway (NOR), and in Japan (JPN) (0.76, 0.76, 0.60, 0.61, respectively). We find that in countries with no significant correlations, such as Viet Nam (VNM), Saudi Arabia (SAU), the United States (USA), and Turkey (TUR), the differences between teachers and principals average perception of school

climate were also the largest (see Figure 3). However, in countries where the correlation was high, the agreement in terms of simple averages was not necessarily among the highest (e.g., Denmark (DNK), Japan (JPN) and New Zealand (NZL)). This indicates that teachers and principals in the same schools, indeed, responded in the same direction, however the strength or magnitude of the climate as perceived by teachers and principals differed.

Figure 3

The Averages of School Climate Reported by Principals (CLIMATEP) and Teachers (CLIMATEB) at the School Level

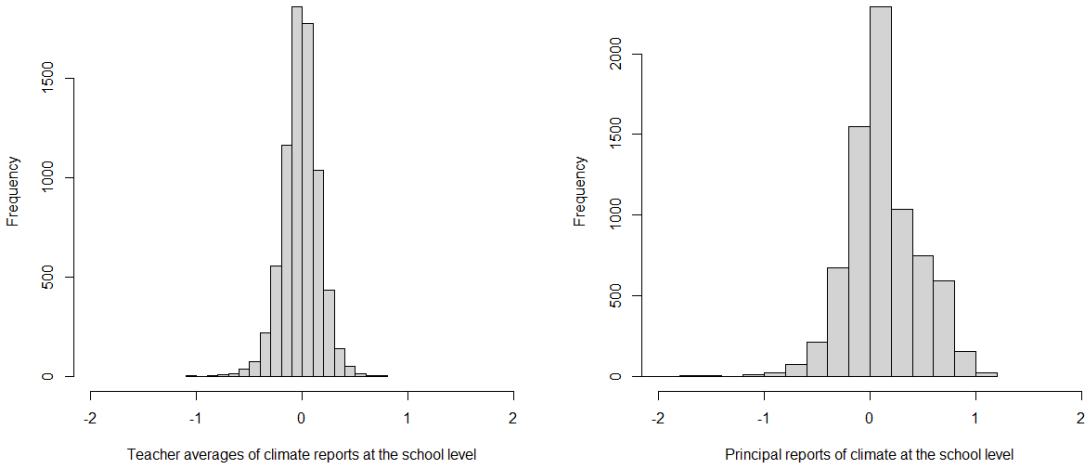


Note. The mean and sd of the climate measures should not be compared across countries as we have limited evidence about cross-country comparability

Note. The box in the boxplot represents the middle 50% of scores for each of the groups whereas the line that divides the box into two parts represents median.

Figure 4

Histograms Representing the Distribution of the Factor Score of Climate Measure Reported by Teachers and Principals in the Entire Sample



Association between Leadership Style and School Climate Reported by Principals and Teachers

We addressed RQ3, RQ4, and RQ5 by conducting a set of multivariate regression models in the entire sample to assess how the school climate perception as reported by principals (see Table 4) and school climate as reported by teachers (see Table 5) associate to different leadership styles (models 1–3). Model 1 is the reference model and includes only the main variables. Model 2 controls for the school and principal background factors, and Model 3 includes a country dummy variable.

Table 4

Standardized Regression Coefficients Showing the Association between Leadership Style (T3PLEADS and T3PLEADP) and Climate Reported by Principals (CLIMATEP), in the Entire Sample

	Model 1			**Model 2**			**Model 3**		
	beta	*SE*	*p*	beta	*SE*	*p*	beta	*SE*	*p*
Climate reported by principals									
Instructional leadership (*T3PLEADS)*	0.15	0.02	0.00	0.14	0.02	0.000	0.13	0.02	0.00
Distributed leadership (*T3PLEADP)*	0.31	0.02	0.00	0.32	0.02	0.000	0.32	0.02	0.00
Control									
School size				-0.04	0.01	0.01	-0.06	0.01	0.00
Language of students				-0.06	0.02	0.00	-0.01	0.02	0.48
Lack of personnel				-0.06	0.02	0.00	-0.08	0.02	0.00
Lack of resources				0.01	0.02	0.63	0.00	0.02	0.90
Principal experience				0.07	0.02	0.00	0.10	0.01	0.00
Teachers experience				-0.12	0.02	0.00	0.00	0.02	0.97
Female				-0.03	0.02	0.07	0.04	0.02	0.01
Public				-0.09	0.02	0.00	-0.12	0.02	0.00

Note: Model 1: $R^2 = 0.14$ (0.01); model 2: $R^2 = 0.17$ (0.01); model 3: $R^2 = 0.26$ (0.01)

Table 5

Standardized Regression Coefficients Showing the Association between Leadership Style (T3PLEADS and T3PLEADP) and Climate Reported by Teachers (CLIMATEB), in the Entire Sample

	Model 1			**Model 2**			**Model 3**		
	beta	*SE*	*p*	beta	*SE*	*p*	beta	*SE*	*p*
Climate reported by principals									
Instructional leadership (*T3PLEADS)*	0.05	0.02	0.01	0.04	0.02	0.02	0.04	0.02	0.04
Distributed leadership (*T3PLEADP)*	0.16	0.02	0.00	0.17	0.02	0.00	0.17	0.02	0.00
Control									
School size				-0.15	0.02	0.00	-0.16	0.02	0.00
Language of students				-0.03	0.02	0.14	-0.03	0.02	0.15
Lack of personnel				-0.03	0.02	0.07	-0.04	0.02	0.03

Lack of resources	-0.00	0.02	0.86	-0.00	0.02	0.79
Principal experience	0.08	0.02	0.00	0.09	0.02	0.00
Teachers experience	-0.02	0.02	0.31	-0.00	0.03	0.74
Female	0.01	0.02	0.47	0.02	0.02	0.20
Public	-0.13	0.02	0.00	-0.14	0.02	0.00

Note: Model 1: R2 = 0.03 (0.01); model 2: R2 = 0.08 (0.01); model 3: R2 = 0.08 (0.01)

The analysis of the pooled sample revealed small positive association and moderate positive association between leadership styles and the school climate as perceived by teachers and principals, respectively (see Table 4 and Table 5). A stronger instructional leadership in school associates with stronger school climate as perceived by principals in the entire sample ($\beta_{T3PLEADS}$ = 0.15*** [0.02]). Moreover, teacher perceived school climate positively associates with instructional leadership in schools, however, this association is very small ($\beta_{T3PLEADS}$ = 0.05** [0.02]). On the other hand, distributed leadership in schools associates with stronger school climate as perceived by both teachers and principals ($\beta_{T3PLEADP}$ = 0.16*** [0.02], $\beta_{T3PLEADP}$ = 0.31*** [0.02], respectively) however this association is much stronger in the sample of principals. After controlling for school and principal characteristics in Model 2, the effects of leadership styles only slightly change in the model that predicted teacher perceived school climate ($\beta_{T3PLEADS}$ = 0.04* [0.02]; $\beta_{T3PLEADp}$ = 0.17*** [0.02]), similarly to the model that predicted principals perceived school climate ($\beta_{T3PLEADS}$ = 0.14*** [0.02]; $\beta_{T3PLEADp}$ = 0.32*** [0.02]). The change in explained variance from Model 1 to Model 2 was approximately 2% in both instances, indicating that the control variables did not greatly contribute to the analyses.

To address RQ5, we also included a set of dummy variables in Model 3, identifying the countries to estimate country fixed effects. A similar approach was used in other leadership studies with the same sample to control for unobserved country characteristics and their effects on the outcome variable (Bellibas & Liu, 2018; Gumus & Bellibas, 2016). After including the country dummy, the effect of leadership styles and climate only slightly changed. However, R^2 almost doubled (R^2 = 0.26*** [0.01]) in the model that included principal perceptions of school climate, indicating that, after controlling for between- country variance, we could explain approximately 26% of the variance in the climate as perceived by principals. On the contrary, the Table 5 shows that

between-country variance did not substantially matter for the teachers' results.

We expand RQ3 and RQ4 by isolating the country context using a within-country analytical approach (see Appendix 4). This approach provides a robustness check to the reference model in Table 4 and Table 5. The within country analysis showed that both leadership styles together can explain on average 16% of the variation in principals' perceived school climate, ranging from 37% in Korea (KOR) to only 2% in France (FRA). Both leadership styles can on average explain 5% of the variation in teacher perceived school climate, ranging from 15% in Croatia (HRV) to close to zero values in Bulgaria (BGR) and Estonia (EST).

Following the analysis of the pooled international sample, principals perceive stronger school climate in schools where they also report stronger instructional and distributed leadership approaches. Compared to instructional leadership, distributed leadership has a stronger and more consistent relationship with the principals' perception of school climate. The regression coefficient for distributed leadership is substantial and statistically significant in the majority of countries (n= 30), whereas that of instructional leadership is more moderate and statistically significant in less than half of the included countries (n= 16). For the rest of the countries this relationship appears insignificant. Moreover, the results do not reveal a pattern among countries with geographical proximity or linguistic similarities.

In comparison to principals, teachers perceived school climate cannot be explained with instructional leadership in the international sample nor within countries. This is only partially true for Viet Nam (VNM), Portugal (PRT), Mexico (MEX), and Brazil (BRA) where stronger instructional leadership as reported by principals was positively associated with teachers perceived school climate. The results point instead to the predominance of the distributed leadership, as reported by principals, positively relating to teacher perceived school climate in many countries (HRV, CHL, NZL, DNK, ARE, BEL, COL, BRA, ZAF, SWE, AUT, SAU, GEO, FIN, SVN, SVK).

Discussion and Conclusion

Over the three cycles of TALIS the principal questionnaires consistently included items on school leadership and school climate. With each new cycle, the teacher perspective received increasing attention, allowing us to now study the features of these organizations comprehensively (OECD, 2019; Veletić & Olsen, 2021a). In this study, we utilized parallel items in the teacher and principal questionnaires from TALIS 2018 to capture certain core aspects of school climate jointly achieved by teachers and principals (dotted parts in Figure 1). Figure 1 further emphasized the importance of a strong shared climate for strong LFL. Comprehending the connection between leadership, how climate is perceived, school environment, and teacher–principal actions and roles provides additional insights into overall organizational quality in schools. The first step toward such an understanding was to examine how perceptions of school climate differ between teachers and principals.

Altogether, we found that teachers and principals consistently rate their environment in the same direction, albeit to differing magnitudes. In the majority of countries principals rate school climate as better than the teacher average in the same schools. This finding is consistent with previous research investigating the gap between teachers' and principals' perceptions of other school-level factors, such as leadership and decision-making (Braddy et al., 2014; Brezicha et al., 2020; Park & Ham, 2016). A notable exception is the teachers from BGR and GEO who on average reported a better school climate than their principals. Only in Spain (ESP), Norway (NOR), Estonia (EST), Brazil (BRA), Cyprus (CYP), and Latvia (LVA) is the difference in magnitudes of perceptions of school climate negligible.

This distance between teachers and principals about the strength of school climate does not necessarily indicate a weakened school climate. For several countries both groups reported about a good school climate, but since principal reports were higher, we still perceived differences. However, if everyone agrees that the climate is negative, there were not degrees of perceptual difference between two groups. Our analysis shows that it is, therefore, the average direction of the climate as positive or negative, rather than the magnitude of the climate, more informative for the overall study of school climate (Van Vianen et al., 2011). So, indeed, it is possible to have a strong school climate even when

there are some disagreements in magnitudes of the perceptions, as long as these perceptions are positive.

The LFL framework presented in Figure 1 highlights that responsibilities and opportunities for teachers to participate in various school decisions create a strong LFL. Indeed, both leadership measures in TALIS deal with (1) the extent to which staff, parents, and students are given opportunity to participate in school decisions and 2) the extent to which teachers take responsibility to develop new teaching practices and improve teaching skills and student learning (OECD, 2019). Thus, schools seeking to implement LFL are characterized by activities where principals interact with other school stakeholders around specific tasks related to decision-making and instruction.

Our finding regarding the consistent positive association between principals' reported school leaderships and their perception of school climate is not surprising. In most countries, a principal who reports that leadership in their school is strongly distributed also tends to report about a good school climate. In a lower number of countries, the same tendency is observed for the relationship between instructional leadership and school climate.

Overall, the similar associations are weaker between teachers' perceived climate and their principal's reported level of instructional and distributed leadership. In particular there is no substantial association between principals' level of instructional leadership and teachers' perception of school climate. However, we find that stronger distributed leadership predicts the school climate as perceived by teachers in almost half of the countries. This finding is partially in line with previous research that shows that distributed rather than instructional leadership associates positively with teacher outcomes (Çoban & Atasoy, 2020; García Torres, 2019; Kılınç et al., 2022). We believe that other factors not accounted for, such as teacher collaboration or decision making, are essential in the countries where we did not find significant associations (Brezicha et al., 2020; Çoban & Atasoy, 2020; Hariri et al., 2016; Sarafidou & Chatziioannidis, 2013). Moreover, because effective leadership assumes that climate and leadership are aligned (Døjbak Haakonsson et al., 2008), finding such no association between the two, might also indicate that leadership in these countries is poor.

Another interesting finding is that when representing countries as fixed effects in the model, the increase in explained variance for the model of teachers' reported school climate was close to zero, while for model for the principals' we observed a 10% increase in the explained variance, approximately a doubling of the explained variance. This finding indicates the need to consider how cultural norms and assumptions on educational expectations influence normative views on successful leadership and how high-quality school climate differs across educational systems. Compared with several other measures included in international comparative studies, this represents a large between-country variability. Thus, further studies are needed to explore and understand how specific country characteristics or stable features of educational systems could account for this variability across countries (e.g., features reflecting educational policy, governance structures, and shared norms, values, or beliefs).

Knowing that the perceptual differences between school stakeholders is one of the indicators of effective leadership in schools calls for more attention, especially in school leaders' professional development. Principals can be more effective with their teachers if they work with teachers to understand where the school climate could be improved. This is crucial in circumstances where, for instance, principals believe that there is a common climate of shared beliefs about teaching and learning but teachers think they are excluded or left on their own (Brezicha et al., 2020). Such a situation can create a disruption in the process of teaching and learning further influencing student outcomes. Therefore, identifying such discrepancies can raise awareness of and stimulate efforts to improve communication and collaboration, and ultimately lead to enhanced organizational quality. Consequently, this reciprocal interaction between teachers and principals becomes crucial to improving school climate. Particularly, the principal has an important role and thus must be approachable, socially oriented toward their teachers, supportive, and trustful; these attitudes will create a school environment where teachers can thrive (Price, 2012, 2015). Principals are expected to perceive themselves as directly responsible for establishing conducive school leadership and climate. Accordingly, social desirability, self-awareness, personal characteristics, and culture are likely to be involved in the principals' self-report of such phenomena (Daniëls et al., 2020; Devos et al., 2013; Fleenor et al., 2010).

Future Directions and Limitations

This study applies organizational quality theoretical concepts in an LFL framework to communicate the tight connection between school leadership and climate, particularly addressing the tight connection between the two core actors within schools, principals and teachers. The proposed LFL framework in Figure 1 illustrates the need to deepen the communication and relationship between teachers and principals. As Figure 1 shows, only a small fragment of leadership is solely in the hands of principal. However, principals still feel most pressed and responsible for creating and maintaining organizational quality. The dotted area in Figure 1 emphasizes organizational factors directing teachers and principals. This part explains the existence of different perceptions of school climate as reported by principals and teachers.

Once we established a comparable measure of school climate across teachers and principals our original intent was to represent the dissonance as a simple gap measure (difference in school climate score) for the two actors within schools to enable further and more detailed examination of this phenomena. However, closer inspection of this absolute measure of the dissonance clearly indicated that such a gap score is largely decided by principals' reports of school climate, simply because teachers' average reports have much less variability across schools. For further studies investigating phenomena from different perspectives and levels of analysis we generally warn against using simple differences since the measure from individual reports (either teachers or principals) will largely influence the final measure of dissonance.

In this study, we focused on teacher–principal relationships, though other actors are also important. Students, the broader community, and parents have important functions to realize LFL and school climate. However, TALIS does not include students and parents as respondents, thus limiting the investigation for LFL with the available data.

Although the present study used advanced statistical including MI tests between principals and teachers at the school level, certain methodological limitations should be noted. The complexity of the models and computational challenges did not allow us to test cross-country, cross-level, and cross-respondent MI in one comprehensive model. Consequently, direct comparisons across countries are not advisable. A two-level model

with countries at the higher level was not possible with the TALIS dataset because it does not provide any country-level variables for analysis and the sample size is limited. However, the countries as fixed-effects model demonstrates a large variability in how principals perceived school climate across countries.

The main strengths of this study are threefold. First, it brings several different LFL models into one comprehensive framework, thus exhausting leadership functions and actors. Second, it examines school climate from both teachers and principal perceptions by providing a comparable measure at the school level. Third, it applies organizational quality ideas to educational research, expanding the opportunities to understand and describe complex networks and relationships between school stakeholders and their association with leadership style. Together, our framework establishes a better understanding of how leadership and climate perceptions affect school organizational quality.

Data Availability Statement

Data for this study come from Teaching and Learning International Survey TALIS 2018 and are publicly available here: https://www.oecd.org/education/talis/talis-2018-data.htm

References

Ahn, J., Bowers, A. J., & Welton, A. D. (2021). Leadership for learning as an organization-wide practice: Evidence on its multilevel structure and implications for educational leadership practice and research. *International Journal of Leadership in Education*, *0*(0), 1–52. https://doi.org/10.1080/13603124.2021.1972162

Ainley, J., & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 Conceptual Framework*. https://doi.org/10.1787/799337c2-en

Akomolafe, C. O., & Adesua, V. O. (2016). The impact of physical facilities on students' level of motivation and academic performance in senior secondary schools in South West Nigeria. *Journal of Education and Practice*, *7*(4), 38–42.

Aldridge, J. M., & Fraser, B. J. (2016). Teachers' views of their school climate and its relationship with teacher self-efficacy and job satisfaction. *Learning Environments Research*, *19*(2), 291–307. https://doi.org/10.1007/s10984-015-9198-x

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018). The impact of school SES on student achievement: evidence from U.S. statewide achievement data. *Educational Evaluation and Policy Analysis*, *40*(4), 613–630. https://doi.org/10.3102/0162373718787917

Ashforth, B. E. (1985). Climate formation: Issues and extensions. *The Academy of Management Review*, *10*(4), 837–847. https://doi.org/10.2307/258051

Asparouhov, T., & Muthen, B. (2018). *SRMR in Mplus*. http://www.statmodel.com/download/SRMR2.pdf

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: does it really matter? *Personnel Psychology*, *51*(3), 577–598. https://doi.org/10.1111/j.1744-6570.1998.tb00252.x

Avvisati, F. (2018). *How are school performance and school climate related to teachers' experience?* https://doi.org/10.1787/af992283-en

Bandura, A. (1988). Organisational applications of social cognitive theory. *Australian Journal of Management*, *13*(2), 275–302. https://doi.org/10.1177/031289628801300210

Barnett, K., & McCormick, J. (2004). Leadership and individual principal-teacher relationships in schools. *Educational Administration Quarterly*, *40*(3), 406–434. https://doi.org/10.1177/0013161X03261742

Bellibas, M. S., & Liu, Y. (2017). Multilevel analysis of the relationship between principals' perceived practices of instructional leadership and teachers' self-efficacy perceptions. *Journal of Educational Administration*, *55*(1), 49–69. Scopus. https://doi.org/10.1108/JEA-12-2015-0116

Bellibas, M. S., & Liu, Y. (2018). The effects of principals' perceived instructional and distributed leadership practices on their perceptions of school climate. *International Journal of Leadership in Education*, *21*(2), 226–244. https://doi.org/10.1080/13603124.2016.1147608

Bowers, A. J. (2020). Examining a congruency-typology model of leadership for learning using two-level latent class analysis with TALIS 2018. *OECD Education Working Papers No. 219*. https://doi.org/10.1787/c963073b-en

Boyce, J., & Bowers, A. J. (2018). Toward an evolving conceptualization of instructional leadership as leadership for learning: Meta-narrative review of 109 quantitative studies across 25 years. *Journal of Educational Administration*, *56*(2), 161–182. https://doi.org/10.1108/JEA-06-2016-0064

Braddy, P. W., Gooty, J., Fleenor, J. W., & Yammarino, F. J. (2014). Leader behaviors and career derailment potential: A multi-analytic method examination of rating source and self–other agreement. *The Leadership Quarterly*, *25*(2), 373–390. https://doi.org/10.1016/j.leaqua.2013.10.001

Brezicha, K. F., Ikoma, S., Park, H., & LeTendre, G. K. (2020). The ownership perception gap: Exploring teacher job satisfaction and its relationship to teachers' and principals' perception of decision-making opportunities. *International Journal of Leadership in Education*, *23*(4), 428–456. Scopus. https://doi.org/10.1080/13603124.2018.1562098

Brown, D. J., & Keeping, L. M. (2005). Elaborating the construct of transformational leadership: The role of affect. *The Leadership Quarterly*, *16*(2), 245–272. https://doi.org/10.1016/j.leaqua.2005.01.003

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (Second edition). The Guilford Press; Second edition (January 8, 2015).

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M., & Vijver, F. J. R. van de. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005. https://doi.org/10.1037/a0013193

Çoban, Ö., & Atasoy, R. (2020). Relationship between distributed leadership, teacher collaboration and organizational innovativeness. *International Journal of Evaluation and Research in Education*, *9*(4), 903–911. Scopus. https://doi.org/10.11591/ijere.v9i4.20679

Coelho, V. A., Bear, G. G., & Brás, P. (2020). A multilevel analysis of the importance of school climate for the trajectories of students' self-concept and self-esteem throughout the middle school transition. *Journal of Youth and Adolescence*, *49*(9), 1793–1804. https://doi.org/10.1007/s10964-020-01245-7

Cohen, J. (2013). Creating a positive school climate: A foundation for resilience. In S. Goldstein & R. B. Brooks (Eds.), *Handbook of Resilience in Children* (pp. 411–423). Springer US. https://doi.org/10.1007/978-1-4614-3661-4_24

Cohen, J., McCabe, Michelli, N., & Pickeral, N. M. (2009). School climate: Research, policy, teacher education and practice. *Teachers College Record*, *Vol 111*, 180–213.

Collie, R. J. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, *104*(4), 1189. https://doi.org/10.1037/a0029356

Collie, R. J., Shapka, J. D., & Perry, N. E. (2011). Predicting teacher commitment: The impact of school climate and social–emotional learning. *Psychology in the Schools*, *48*(10), 1034–1048. https://doi.org/10.1002/pits.20611

Daniëls, E., Hondeghem, A., & Heystek, J. (2020). School leaders' and teachers' leadership perceptions: Differences and similarities. *Journal of Educational Administration*, *58*(6), 645–660. https://doi.org/10.1108/JEA-11-2019-0199

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

Dempster, N. (2019). Leadership for learning: Embracing purpose, people, pedagogy and place. In T. Townsend (Ed.), *Instructional Leadership and Leadership for Learning in Schools: Understanding Theories of Leading* (pp. 403–421). Springer International Publishing. https://doi.org/10.1007/978-3-030-23736-3_16

Devos, G., Hulpia, H., Tuytens, M., & Sinnaeve, I. (2013). Self-other agreement as an alternative perspective of school leadership analysis: An exploratory study. *School Effectiveness and School Improvement*, *24*(3), 296–315. https://doi.org/10.1080/09243453.2012.693103

Dickhäuser, O., Janke, S., Daumiller, M., & Dresel, M. (2021). Motivational school climate and teachers' achievement goal orientations: A hierarchical approach. *British Journal of Educational Psychology*, *91*(1), e12370. https://doi.org/10.1111/bjep.12370

DiPietro, S. M., Slocum, L. A., & Esbensen, F.-A. (2015). School climate and violence: Does immigrant status matter? *Youth Violence and Juvenile Justice*, *13*(4), 299–322. https://doi.org/10.1177/1541204014547589

Døjbak Haakonsson, D., Burton, R. M., Obel, B., & Lauridsen, J. (2008). How failure to align organizational climate and leadership style affects performance. *Management Decision*, *46*(3), 406–432. https://doi.org/10.1108/00251740810863861

Drexler, J. A. (1977). Organizational climate: Its homogeneity within organizations. *Journal of Applied Psychology*, *62*(1), 38–42. https://doi.org/10.1037/0021-9010.62.1.38

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

Eryilmaz, N., & Sandoval Hernandez, A. (2021). Improving cross-cultural comparability: Does school leadership mean the same in different countries? *Educational Studies*. Scopus. https://doi.org/10.1080/03055698.2021.2013777

Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing*, *17*(2), 105–120. https://doi.org/10.1002/(SICI)1520-6793(200002)17:2<105::AID-MAR3>3.0.CO;2-9

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self–other rating agreement in leadership: A review. *The Leadership Quarterly*, *21*(6), 1005–1034. https://doi.org/10.1016/j.leaqua.2010.10.006

García Torres, D. (2019). Distributed leadership, professional collaboration, and teachers' job satisfaction in U.S. schools. *Teaching and Teacher Education*, *79*, 111–123. Scopus. https://doi.org/10.1016/j.tate.2018.12.001

Goldkind, L., & Farmer, G. L. (2013). The enduring influence of school size and school climate on parents' engagement in the school community. *School Community Journal*, *23*(1), 223–244.

Grazia, V., & Molinari, L. (2020). School climate multidimensionality and measurement: A systematic literature review. *Research Papers in Education*, *0*(0), 1–27. https://doi.org/10.1080/02671522.2019.1697735

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361–396. https://doi.org/10.3102/00346543066003361

Griffith, J. (1999). The school leadership/school climate relation: Identification of school configurations associated with change in principals. *Educational Administration Quarterly*, *35*(2), 267–291. https://doi.org/10.1177/00131619921968545

Gumus, E., & Bellibas, M. S. (2016). The effects of professional development activities on principals' perceived instructional leadership practices: Multi-country data analysis using TALIS 2013. *Educational Studies*, *42*(3), 287–301. https://doi.org/10.1080/03055698.2016.1172958

Gustafsson, J. E., & Nilsen, T. (2016). The impact of school climate and teacher quality on mathematics achievement: A difference-in-differences approach. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 81–95). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8_4

Hallinger, P. (2009, September). *Leadership for the 21st century schools: From instructional leadership to leadership for learning* [Public Lecture Series of the Hong Kong Institute of Education]. https://repository.eduhk.hk/en/publications/leadership-for-the-21st-century-schools-from-instructional-leader-3

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. (2015). The evolution of instructional leadership. In P. Hallinger & W.-C. Wang (Eds.), *Assessing Instructional Leadership with the Principal Instructional Management Rating Scale* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-15533-3_1

Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management*, *30*(2), 95–110. https://doi.org/10.1080/13632431003663214

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(25), 621–638.

Ham, S.-H., Duyar, I., & Gumus, S. (2015). Agreement of self-other perceptions matters: Analyzing the effectiveness of principal leadership through multi-source assessment. *Australian Journal of Education*, *59*(3), 225–246.

Hariri, H., Monypenny, R., & Prideaux, M. (2016). Teacher-perceived principal leadership styles, decision-making styles and job satisfaction: How congruent are data from Indonesia with the Anglophile and Western literature? *School Leadership & Management*, *36*(1), 41–62. https://doi.org/10.1080/13632434.2016.1160210

Harris, A. (2009). Distributed leadership: What we know. In A. Harris (Ed.), *Distributed Leadership: Different Perspectives* (pp. 11–21). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9737-9_2

Hoy, W. K. (1990). Organizational climate and culture: A conceptual analysis of the school workplace. *Journal of Educational and Psychological Consultation*, *1*(2), 149–168. https://doi.org/10.1207/s1532768xjepc0102_4

Hoy, W. K., Tarter, C. J., & Hoy, A. W. (2006). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, *43*(3), 425–446. https://doi.org/10.3102/00028312043003425

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

Imig, D., Holden, S., & Placek, D. (2019). Leadership for learning in the US. In T. Townsend (Ed.), *Instructional Leadership and Leadership for Learning in*

Schools: Understanding Theories of Leading (pp. 105–131). Springer International Publishing. https://doi.org/10.1007/978-3-030-23736-3_5

Kalis, M. C. (1980). Teaching experience: Its effect on school climate, teacher morale. *NASSP Bulletin*, *64*(435), 89–102. https://doi.org/10.1177/019263658006443512

Katsantonis, I. G. (2020). Investigation of the impact of school climate and teachers' self-efficacy on job satisfaction: A cross-cultural approach. *European Journal of Investigation in Health, Psychology and Education*, *10*(1), Article 1. https://doi.org/10.3390/ejihpe10010011

Kelley, C., & Halverson, R. (2012). *The Comprehensive Assessment of Leadership for Learning: A Next Generation Formative Evaluation and Feedback System*. 24.

Kelley, R. C. (2005). Relationships between measures of leadership and school climate. *Education*, *Vol.126*(1), 17–25.

Kim, E. S., Wang, Y., & Kiefer, S. M. (2018). Cross-level group measurement invariance when groups are at different levels of multilevel data. *Educational and Psychological Measurement*, *78*(6), 973–997. https://doi.org/10.1177/0013164417739062

Kılınç, A. Ç., Polatcan, M., Turan, S., & Özdemir, N. (2022). Principal job satisfaction, distributed leadership, teacher-student relationships, and student achievement in Turkey: A multilevel mediated-effect model. *Irish Educational Studies*. Scopus. https://doi.org/10.1080/03323315.2022.2061567

Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2008). A multilevel study of predictors of student perceptions of school climate: The effect of classroom-level factors. *Journal of Educational Psychology*, *100*(1), 96–104. https://doi.org/10.1037/0022-0663.100.1.96

Kozlowski, S. W., & Doherty, M. L. (1989). Integration of climate and leadership: Examination of a neglected issue. *Journal of Applied Psychology*, *74*(4), 546–553. https://doi.org/10.1037/0021-9010.74.4.546

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

Kutsyuruba, B., Walker, K., & Noonan, B. (2016). The trust imperative in the school principalship: The Canadian perspective. *Leadership and Policy in Schools*, *15*(3), 343–372. https://doi.org/10.1080/15700763.2016.1164866

Ladd, H. F. (2009). *Teachers' Perceptions of Their Working Conditions: How Predictive of Policy-Relevant Outcomes?* [Data set]. National Center for Analysis of Longitudinal Data in Education Research. https://doi.org/10.1037/e722072011-001

Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly*, *44*(4), 529–561. https://doi.org/10.1177/0013161X08321221

Lenz, A. S., Rocha, L., & Aras, Y. (2021). Measuring school climate: A systematic review of initial development and validation studies. *International Journal for the Advancement of Counselling*, *43*(1), 48–62. https://doi.org/10.1007/s10447-020-09415-9

Liu, S., Keeley, J., Sui, Y., & Sang, L. (2021). Impact of distributed leadership on teacher job satisfaction in China: The mediating roles of teacher autonomy and teacher collaboration. *Studies in educational evaluation*, *71*. https://doi.org/10.1016/j.stueduc.2021.101099

Liu, Y. (2020). Focusing on the practice of distributed leadership: The international evidence from the 2013 TALIS. *Educational Administration Quarterly*, 0013161X20907128. https://doi.org/10.1177/0013161X20907128

Liu, Y., Bellibas, M., & Printy, S. (2018). How school context and educator characteristics predict distributed leadership: A hierarchical structural equation model with 2013 TALIS data. *Educational management administration and leadership*. *46*(3), 401–423. https://doi.org/10.1177/1741143216665839

Lovett, S., & Andrews, D. (2011). Leadership for learning: What it means for teachers. In T. Townsend & J. MacBeath (Eds.), *International Handbook of Leadership for Learning* (Vol. 25). Springer International Handbooks of Education. https://www.nzcer.org.nz/research/publications/leadership-learning-what-it-means-teachers

MacBeath, J., & Dempster, N. (2008). *Connecting Leadership and Learning: Principles for Practice*. Routledge.

Marcoulides, K. M., & Yuan, K.-H. (2020). Using equivalence testing to evaluate goodness of fit in multilevel structural equation models. *International Journal of Research & Method in Education*, *43*(4), 431–443. https://doi.org/10.1080/1743727X.2020.1795113

Marks, G. N. (2015). Are school-SES effects statistical artefacts? Evidence from longitudinal population data. *Oxford Review of Education*, *41*(1), 122–144. https://doi.org/10.1080/03054985.2015.1006613

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., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McCoy, D. C., Roy, A. L., & Sirkman, G. M. (2013). Neighborhood crime and school climate as predictors of elementary school academic quality: A cross-lagged panel analysis. *American Journal of Community Psychology*, *52*(1–2), 128–140. https://doi.org/10.1007/s10464-013-9583-5

Millsap, R. E. (2012). *Statistical Approaches to Measurement Invariance*. Routledge.

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of school climate: A multilevel exploration of patterns of discrepancy. *Journal of School Health*, *80*(6), 271–279. https://doi.org/10.1111/j.1746-1561.2010.00501.x

Moye, M. J., Henkin, A. B., & Egley, R. J. (2005). Teacher-principal relationships: Exploring linkages between empowerment and interpersonal trust. *Journal of*

Educational Administration, *43*(3), 260–277. https://doi.org/10.1108/09578230510594796

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What really matters? *The Journal of Classroom Interaction*, *37*(2), 3–15.

Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership & Management*, *27*(2), 179–201. https://doi.org/10.1080/13632430701237420

Muthén, L. K., & Muthén, B. O. (2017). *Mplus User's Guide*. https://www.statmodel.com/html_ug.shtml

OECD. (2019). *TALIS 2018 Technical Report*.

Ogawa, R. T., & Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly*, *31*(2), 224–243. https://doi.org/10.1177/0013161X95031002004

Otero, G. (2019). Creating and leading powerful learning relationships through a whole school community approach. In T. Townsend (Ed.), *Instructional Leadership and Leadership for Learning in Schools: Understanding Theories of Leading* (pp. 317–346). Springer International Publishing. https://doi.org/10.1007/978-3-030-23736-3_13

Özdemir, N., Gümüş, S., Kılınç, A. Ç., & Bellibaş, M. Ş. (2022). A systematic review of research on the relationship between school leadership and student achievement: An updated framework and future direction. *Educational Management Administration & Leadership*, 17411432221118662. https://doi.org/10.1177/17411432221118662

Park, J.-H., & Ham, S.-H. (2016). Whose perception of principal instructional leadership? Principal-teacher perceptual (dis)agreement and its influence on teacher collaboration. *Asia Pacific Journal of Education*, *36*(3), 450–469. Scopus. https://doi.org/10.1080/02188791.2014.961895

Peterson, M. W., & Spencer, M. G. (1990). Understanding academic culture and climate. *New Directions for Institutional Research*, *1990*(68), 3–18. https://doi.org/10.1002/ir.37019906803

Pietsch, M., Tulowitzki, P., & Koch, T. (2019). On the differential and shared effects of leadership for learning on teachers' organizational commitment and job satisfaction: A multilevel perspective. *Educational Administration Quarterly*, *55*(5), 705–741. https://doi.org/10.1177/0013161X18806346

Price, H. E. (2012). Principal–teacher interactions: How affective relationships shape principal and teacher attitudes. *Educational Administration Quarterly*, *48*(1), 39–85. https://doi.org/10.1177/0013161X11417126

Price, H. E. (2015). Principals' social interactions with teachers: How principal-teacher social relations correlate with teachers' perceptions of student engagement. *Journal of Educational Administration*, *53*(1), 116–139. https://doi.org/10.1108/JEA-02-2014-0023

Price, H. E. (2016). Assessing U.S. public school quality: The advantages of combining internal "consumer ratings" with external NCLB Ratings. *Educational Policy*, *30*(3), 403–433. https://doi.org/10.1177/0895904814551273

Ramsey, C. M., Spira, A. P., Parisi, J. M., & Rebok, G. W. (2016). School climate: Perceptual differences between students, parents, and school staff. *School Effectiveness and School Improvement*, *27*(4), 629–641. https://doi.org/10.1080/09243453.2016.1199436

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

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257

Sarafidou, J., & Chatziioannidis, G. (2013). Teacher participation in decision making and its impact on school and teachers. *International Journal of Educational Management*, *27*(2), 170–183. https://doi.org/10.1108/09513541311297586

Scherer, R., & Nilsen, T. (2016). The relations among school climate, instructional quality, and achievement motivation in mathematics. In T. Nilsen & J.-E.

Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 51–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-41252-8_3

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

Shakeel, M. D., & DeAngelis, C. A. (2018). Can private schools improve school climate? Evidence from a nationally representative sample. *Journal of School Choice*, *12*(3), 426–445. https://doi.org/10.1080/15582159.2018.1490383

Sherblom, S. A., Marshall, J. C., & Sherblom, J. C. (2006). The relationship between school climate and math and reading achievement. *Journal of Character Education*, *4*(1/2), 19–31.

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

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, *36*(1), 3–34. https://doi.org/10.1080/0022027032000106726

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

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90. https://doi.org/10.1086/209528

Stockdale, M. S., Hangaduambo, S., Duys, D., Larson, K., & Sarvela, P. D. (2002). Rural elementary students', parents', and teachers' perceptions of bullying. *American*

Journal of Health Behavior, *26*(4), 266–277. https://doi.org/10.5993/AJHB.26.4.3

Sulak, T. N. (2018). School climate: The controllable and the uncontrollable. *Educational Studies*, *44*(3), 279–294. https://doi.org/10.1080/03055698.2017.1373630

Suldo, S. M., Thalji-Raitano, A., Hasemeyer, M., Gelley, C. D., & Hoy, B. (2013). Understanding middle school students life satisfaction: Does school climate matter? *Applied Research in Quality of Life*, *8*(2), 169–182. https://doi.org/10.1007/s11482-012-9185-7

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

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, *83*(3), 357–385. https://doi.org/10.3102/0034654313483907

Uline, C., & Tschannen-Moran, M. (2008). The walls speak: The interplay of quality facilities, school climate, and student achievement. *Journal of Educational Administration*, *46*(1), 55–73. https://doi.org/10.1108/09578230810849817

Van Maele, D., & Van Houtte, M. (2015). Trust in school: A pathway to inhibit teacher burnout? *JOURNAL OF EDUCATIONAL ADMINISTRATION*, *53*(1), Article 1. https://doi.org/10.1108/JEA-02-2014-0018

Van Vianen, A. E. M., De Pater, I. E., Bechtoldt, M. N., & Evers, A. (2011). The strength and quality of climate perceptions. *Journal of Managerial Psychology*, *26*(1), 77–92. https://doi.org/10.1108/02683941111099637

Veletić, J., & Olsen, R. V. (2021a). Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018. *International Journal of Leadership in Education*. Scopus. https://doi.org/10.1080/13603124.2021.1953612

Veletić, J., & Olsen, R. V. (2021b). Developing a shared cluster construct of instructional leadership in TALIS. *Studies in Educational Evaluation*, *68*, 100942. https://doi.org/10.1016/j.stueduc.2020.100942

Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Woodman, R. W., & King, D. C. (1978). Organizational climate: Science or folklore? *Academy of Management Review*, *3*(4), 816–826. https://doi.org/10.5465/amr.1978.4289290

Xia, J., & O'Shea, C. (2022). To what extent does distributed leadership support principal instructional leadership? Evidence from TALIS 2013 data. *Leadership and Policy in Schools*. Scopus. https://doi.org/10.1080/15700763.2022.2056059

Yang, C., Sharkey, J. D., Reed, L. A., Chen, C., & Dowdy, E. (2018). Bullying victimization and student engagement in elementary, middle, and high schools: Moderating role of school climate. *School Psychology Quarterly*, *33*(1), 54–64. https://doi.org/10.1037/spq0000250

Zieger, L., Sims, S., & Jerrim, J. (2019). Comparing teachers' job satisfaction across countries: A multiple-pairwise measurement invariance approach. *Educational Measurement: Issues and Practice*, *38*(3), 75–85. https://doi.org/10.1111/emip.12254

Zullig, K. J., Huebner, E. S., & Patton, J. M. (2011). Relationships among school climate domains and school satisfaction. *Psychology in the Schools*, *48*(2), 133–145. https://doi.org/10.1002/pits.20532

Zullig, K. J., Koopman, T. M., Patton, J. M., & Ubbes, V. A. (2010). School climate: Historical review, instrument development, and school assessment. *Journal of Psychoeducational Assessment*, *28*(2), 139–152. https://doi.org/10.1177/0734282909344205

Appendix 1

Final sample sizes per country with country codes

Table 1

Sample Size by Country

	IDCNTRY	CNTRY CODE	Schools	Teachers	The average cluster size
1	Austria	AUT	246	4255	17.30
2	Belgium	BEL	302	5257	17.41
3	Brazil	BRA	185	2447	13.23
4	Bulgaria	BGR	200	2862	14.31
5	Chile	CHL	179	1963	1.97
6	Chinese Taipei	TWN	200	3835	19.18
7	Colombia	COL	154	2398	15.57
8	Croatia	HRV	188	3358	17.86
9	Cyprus	CYP	88	1611	18.31
10	Czech Republic	CZE	219	3447	15.74
11	Denmark	DNK	141	2001	14.19
12	Estonia	EST	195	3004	15.41
13	Finland	FIN	148	2851	19.26
14	France	FRA	176	3006	17.08
15	Georgia	GEO	192	3101	16.15
16	Hungary	HUN	189	3245	17.17
17	Japan	JPN	196	3555	18.14
18	Kazakhstan	KAZ	331	6566	19.84
19	Korea	KOR	163	2931	17.98
20	Latvia	LVA	135	2315	17.15
21	Lithuania	LTU	195	3759	19.28
22	Malta	MLT	55	1656	3.11
23	Mexico	MEX	193	2926	15.16
24	New Zealand	NZL	185	2257	12.20
25	Norway	NOR	185	4154	22.45
26	Portugal	PRT	200	3676	18.38
27	Russian Federation	RUS	230	4011	17.44
28	Saudi Arabia	SAU	179	2744	15.33
29	Slovak Republic	SVK	176	3015	17.13
30	Slovenia	SVN	132	2094	15.86
31	South Africa	ZAF	170	2046	12.04
32	Spain	ESP	399	7407	18.56
33	Sweden	SWE	180	2782	15.46
34	Turkey	TUR	196	3952	2.16
35	United Arab Emirates	UAE	521	8648	16.60
36	United States	USA	165	2560	15.52
37	Viet Nam	VNM	196	3825	19.52
			7384	125520	17.12

Appendix 2

Basic descriptive statistics for the pooled sample and for each country separately

Table 1

Basic Descriptive Statistic for the Entire Sample (Unweighted)

Variable name	n	mean	sd	min	max	skew	kurtosis	se
T3PLEADS	7078	11.16	1.92	3.88	15.86	0.00	0.10	0.02
T3PLEADP	7055	11.97	2.08	0.60	17.06	-0.16	1.48	0.02
FEMALE*	7384	0.46	0.50	0.00	1.00	0.14	-1.98	0.01
PUBLIC*	7384	0.79	0.41	0.00	1.00	-1.42	0.02	0.00
LACK_PERSON*	7384	0.47	0.50	0.00	1.00	0.11	-1.99	0.01
LACK_RESOUR*	7384	0.58	0.49	0.00	1.00	-0.34	-1.89	0.01
NENRSTUD	7081	2.70	1.38	1.00	5.00	0.38	-1.08	0.02
TC3G17A	7043	2.01	1.12	1.00	5.00	1.34	1.23	0.01
TC3G04A	7138	7.04	7.01	0.00	50.00	1.72	3.17	0.08
CLIMATEB	7384	-0.01	0.17	-1.10	0.72	-0.28	1.34	0.00
CLIMATEP	7384	0.13	0.33	-2.12	1.05	0.00	1.13	0.00
AVG_TT3G11A	7382	10.32	5.29	0.25	37.50	0.53	0.00	0.06

Note. *dummy coded. T3PLEADP- Participation among stakeholders scale; T3PLEADS- Instructional leadership scale; FEMALE- female principal in the school; PUBLIC- schools that are publicly managed; LACK_PERSON- lack of personal is not a problem in the school; LACK_RESOUR- lack of resources is not a problem in the school; NENRSTUD- number of enrolled students; TC3G17A-Students first language different from instr. language; TC3G04A- principal years of experience in this school; CLIMATEB- teacher ratings of school climate; CLIMATEP- principal ratings of school climate; AVG_TT3G11A- the average years of experience of teachers in the particular school.

Table 2

Instructional Leadership Scale (T3PLEADS)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	465	12.743	2.022	4.949	15.247	10.298	-0.372	-0.13	0.094
AUT	243	10.968	2.163	5.703	15.57	9.867	0.049	-0.296	0.139
BEL	287	11.03	1.575	7.199	15.521	8.322	0.299	-0.295	0.093
BGR	200	11.047	1.478	7.596	13.883	6.287	0.149	-0.023	0.105
BRA	181	11.752	2.041	5.024	15.095	10.071	-0.176	-0.03	0.152
CHL	168	10.958	2.177	3.879	14.076	10.197	-0.27	-0.481	0.168
COL	135	11.142	2.325	4.603	14.14	9.537	-0.233	-0.853	0.2
CYP	87	10.998	2.116	5.011	15.54	10.529	0.213	-0.071	0.227
CZE	217	10.986	1.516	6.645	14.63	7.985	-0.068	0.215	0.103
DNK	131	10.446	2.553	4.705	15.033	10.328	-0.026	-0.355	0.223
ESP	396	10.888	2.121	5.664	15.226	9.562	0.163	-0.472	0.107
EST	192	11.048	1.905	6.279	15.859	9.58	0.091	-0.079	0.138
FIN	148	11.093	1.661	6.512	15.44	8.928	0.21	0.089	0.137
FRA	171	11.057	2.028	5.761	15.684	9.923	0.154	0.299	0.155
GEO	174	11.166	1.412	6.159	14.104	7.945	-0.051	0.989	0.107
HRV	181	10.927	1.919	5.808	14.795	8.987	0.249	-0.445	0.143
HUN	180	11.052	1.089	8.056	14.031	5.975	-0.013	-0.313	0.081
JPN	195	11.043	1.31	6.987	14.962	7.975	0.296	-0.274	0.094
KAZ	331	10.935	1.483	7.009	13.083	6.074	-0.162	-0.69	0.081
KOR	146	11.198	2.492	5.519	15.62	10.101	0.14	-0.632	0.206
LTU	194	11.254	1.448	8.343	14.147	5.804	0.239	-0.175	0.104
LVA	133	11.107	1.433	7.045	14.089	7.044	0.004	0.179	0.124
MEX	193	10.797	2.097	6.232	14.121	7.889	-0.052	-0.929	0.151
MLT	50	11.062	2.128	5.328	15.577	10.249	-0.026	-0.091	0.301
NOR	152	11.054	1.68	5.823	15.369	9.546	-0.382	0.486	0.136
NZL	179	10.942	2.006	4.919	15.123	10.204	-0.358	0.601	0.15
PRT	200	10.673	2.019	5.457	14.881	9.424	0.191	-0.357	0.143
RUS	229	11.038	1.435	6.272	14.192	7.92	-0.045	0.337	0.095
SAU	167	11.089	2.193	4.331	14.374	10.043	-0.358	0.035	0.17
SVK	176	11.082	1.666	5.262	14.41	9.148	-0.261	0.752	0.126
SVN	116	10.646	1.869	6.409	14.686	8.277	0.238	-0.442	0.173
SWE	159	10.147	1.832	5.146	14.485	9.339	0.317	-0.257	0.145
TUR	190	11.333	1.792	6.728	14.887	8.159	0.032	-0.098	0.13
TWN	199	11.353	1.773	6.354	14.827	8.473	0.023	-0.032	0.126
USA	158	11.132	2.174	5.69	15.305	9.615	0.003	-0.084	0.173
VNM	196	12.122	1.185	8.976	14.663	5.687	0.31	0.587	0.085
ZAF	159	11.017	1.943	4.224	14.07	9.846	-0.345	0.093	0.154

Table 3

Distributed Leadership Scale (T3PLEADP)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	461	11.988	2.414	1.597	16.172	14.575	-0.355	2.049	0.112
AUT	241	11.991	2.003	6.681	15.263	8.582	-0.123	-0.478	0.129
BEL	283	12.116	2.129	5.492	16.965	11.473	-0.083	0.393	0.127
BGR	200	12.003	1.537	8.363	15.467	7.104	0.412	0.404	0.109
BRA	182	11.704	2.42	3.389	16.016	12.627	0.08	0.354	0.179
CHL	167	11.784	2.846	2.573	16.744	14.171	-0.282	0.375	0.22
COL	134	12.005	2.404	5.447	14.972	9.525	-0.224	-0.72	0.208
CYP	87	12.173	2.311	4.289	17.057	12.768	-0.122	1.271	0.248
CZE	217	11.941	1.334	6.596	15.373	8.777	0.191	1.156	0.091
DNK	132	11.946	1.42	7.718	15.653	7.935	0.181	0.297	0.124
ESP	396	11.748	2.637	1.991	16	14.009	-0.173	0.344	0.133
EST	192	11.995	1.947	7.422	15.368	7.946	0.373	-0.904	0.14
FIN	148	12.009	1.535	8.629	16.025	7.396	0.268	-0.019	0.126
FRA	170	12.111	2.211	4.905	16.652	11.747	-0.02	0.964	0.17
GEO	175	12.006	2.042	5.928	15.436	9.508	0.495	-0.428	0.154
HRV	179	12.008	2.103	6.477	16.262	9.785	0.445	0.612	0.157
HUN	180	11.896	2.021	7.749	15.796	8.047	0.331	-0.019	0.151
JPN	195	12.005	1.29	8.034	16.658	8.624	0.132	1.303	0.092
KAZ	331	12.198	1.875	4.515	15.91	11.395	0.019	0.455	0.103
KOR	145	11.871	2.127	5.362	14.788	9.426	-0.143	-0.558	0.177
LTU	194	11.946	2.219	0.791	15.241	14.45	-0.374	2.088	0.159
LVA	133	11.973	1.755	7.664	15.373	7.709	0.685	-0.347	0.152
MEX	193	12.049	2.317	2.988	16.04	13.052	-0.454	1.109	0.167
MLT	50	12.105	1.897	8.177	16.204	8.027	0.172	-0.184	0.268
NOR	152	12.003	1.436	5.328	16.092	10.764	-0.618	4.259	0.116
NZL	178	11.652	1.566	7.041	15.814	8.773	0.372	1.532	0.117
PRT	200	11.973	2.108	4.813	16.137	11.324	-0.049	1.02	0.149
RUS	230	11.908	2.416	5.586	15.512	9.926	0.107	-0.481	0.159
SAU	167	11.982	2.611	3.624	16.927	13.303	-0.781	1.433	0.202
SVK	176	12.039	1.359	8.676	15.523	6.847	0.243	0.547	0.102
SVN	115	11.997	1.816	7.133	15.823	8.69	0.436	0.568	0.169
SWE	155	12.069	1.322	8.365	15.669	7.304	0.032	0.348	0.106
TUR	190	12.012	2.586	1.214	16	14.786	-0.747	2.895	0.188
TWN	199	11.975	1.501	7.863	15.924	8.061	-0.022	0.762	0.106
USA	155	12.044	1.834	7.637	16.166	8.529	0.358	0.173	0.147
VNM	195	11.992	2.009	2.766	16.466	13.7	-0.53	2.315	0.144
ZAF	158	11.853	2.541	0.602	16.782	16.18	-0.167	1.814	0.202

Table 4

FEMALE (Dummy Coded, 1=Female, 0= Male)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	0.466	0.499	0	1	1	0.134	-1.986	0.022
AUT	246	0.459	0.499	0	1	1	0.162	-1.982	0.032
BEL	302	0.464	0.499	0	1	1	0.145	-1.985	0.029
BGR	200	0.735	0.442	0	1	1	-1.057	-0.887	0.031
BRA	185	0.746	0.437	0	1	1	-1.121	-0.748	0.032
CHL	179	0.453	0.499	0	1	1	0.189	-1.975	0.037
COL	154	0.331	0.472	0	1	1	0.71	-1.505	0.038
CYP	88	0.5	0.503	0	1	1	0	-2.023	0.054
CZE	219	0.479	0.501	0	1	1	0.082	-2.002	0.034
DNK	141	0.355	0.48	0	1	1	0.601	-1.65	0.04
ESP	399	0.426	0.495	0	1	1	0.298	-1.916	0.025
EST	195	0.564	0.497	0	1	1	-0.257	-1.944	0.036
FIN	148	0.48	0.501	0	1	1	0.08	-2.007	0.041
FRA	176	0.449	0.499	0	1	1	0.204	-1.97	0.038
GEO	192	0.562	0.497	0	1	1	-0.25	-1.948	0.036
HRV	188	0.505	0.501	0	1	1	-0.021	-2.01	0.037
HUN	189	0.593	0.493	0	1	1	-0.374	-1.87	0.036
JPN	196	0.077	0.267	0	1	1	3.161	8.036	0.019
KAZ	331	0.592	0.492	0	1	1	-0.373	-1.866	0.027
KOR	163	0.209	0.408	0	1	1	1.421	0.02	0.032
LTU	195	0.497	0.501	0	1	1	0.01	-2.01	0.036
LVA	135	0.785	0.412	0	1	1	-1.373	-0.114	0.035
MEX	193	0.347	0.477	0	1	1	0.637	-1.602	0.034
MLT	55	0.418	0.498	0	1	1	0.323	-1.93	0.067
NOR	185	0.443	0.498	0	1	1	0.227	-1.959	0.037
NZL	185	0.416	0.494	0	1	1	0.337	-1.896	0.036
PRT	200	0.425	0.496	0	1	1	0.301	-1.919	0.035
RUS	230	0.696	0.461	0	1	1	-0.845	-1.292	0.03
SAU	179	0.536	0.5	0	1	1	-0.144	-1.99	0.037
SVK	176	0.665	0.473	0	1	1	-0.692	-1.53	0.036
SVN	132	0.561	0.498	0	1	1	-0.241	-1.956	0.043
SWE	180	0.589	0.493	0	1	1	-0.358	-1.882	0.037
TUR	196	0.087	0.282	0	1	1	2.914	6.526	0.02
TWN	200	0.295	0.457	0	1	1	0.892	-1.21	0.032
USA	165	0.37	0.484	0	1	1	0.535	-1.724	0.038
VNM	196	0.276	0.448	0	1	1	0.997	-1.011	0.032
ZAF	170	0.276	0.449	0	1	1	0.991	-1.024	0.034

Table 5

PUBLIC (Dummy Coded, 1= Public, 0= Private)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	0.43	0.496	0	1	1	0.282	-1.924	0.022
AUT	246	0.85	0.358	0	1	1	-1.944	1.787	0.023
BEL	302	0.368	0.483	0	1	1	0.547	-1.707	0.028
BGR	200	0.975	0.157	0	1	1	-6.039	34.646	0.011
BRA	185	0.692	0.463	0	1	1	-0.824	-1.327	0.034
CHL	179	0.486	0.501	0	1	1	0.055	-2.008	0.037
COL	154	0.695	0.462	0	1	1	-0.838	-1.306	0.037
CYP	88	0.807	0.397	0	1	1	-1.528	0.339	0.042
CZE	219	0.945	0.228	0	1	1	-3.886	13.159	0.015
DNK	141	0.73	0.445	0	1	1	-1.028	-0.95	0.037
ESP	399	0.664	0.473	0	1	1	-0.693	-1.524	0.024
EST	195	0.944	0.231	0	1	1	-3.816	12.626	0.017
FIN	148	0.919	0.274	0	1	1	-3.038	7.281	0.023
FRA	176	0.83	0.377	0	1	1	-1.738	1.026	0.028
GEO	192	0.839	0.369	0	1	1	-1.826	1.341	0.027
HRV	188	0.947	0.225	0	1	1	-3.95	13.677	0.016
HUN	189	0.799	0.402	0	1	1	-1.48	0.191	0.029
JPN	196	0.893	0.31	0	1	1	-2.521	4.377	0.022
KAZ	331	0.867	0.34	0	1	1	-2.153	2.642	0.019
KOR	163	0.773	0.42	0	1	1	-1.292	-0.334	0.033
LTU	195	0.979	0.142	0	1	1	-6.713	43.292	0.01
LVA	135	0.948	0.223	0	1	1	-3.997	14.084	0.019
MEX	193	0.777	0.417	0	1	1	-1.322	-0.254	0.03
MLT	55	0.382	0.49	0	1	1	0.473	-1.808	0.066
NOR	185	0.762	0.427	0	1	1	-1.222	-0.511	0.031
NZL	185	0.908	0.29	0	1	1	-2.803	5.887	0.021
PRT	200	0.87	0.337	0	1	1	-2.184	2.783	0.024
RUS	230	0.987	0.114	0	1	1	-8.528	71.032	0.007
SAU	179	0.955	0.207	0	1	1	-4.37	17.194	0.015
SVK	176	0.875	0.332	0	1	1	-2.248	3.073	0.025
SVN	132	0.871	0.336	0	1	1	-2.191	2.823	0.029
SWE	180	0.744	0.437	0	1	1	-1.112	-0.769	0.033
TUR	196	0.918	0.275	0	1	1	-3.033	7.234	0.02
TWN	200	0.895	0.307	0	1	1	-2.558	4.565	0.022
USA	165	0.879	0.327	0	1	1	-2.3	3.311	0.025
VNM	196	0.923	0.267	0	1	1	-3.161	8.036	0.019
ZAF	170	0.718	0.451	0	1	1	-0.958	-1.088	0.035

Table 6

LACK_PERSON (Dummy Coded, Not a Problem=0, Else)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	0.599	0.491	0	1	1	-0.402	-1.842	0.021
AUT	246	0.211	0.409	0	1	1	1.405	-0.026	0.026
BEL	302	0.692	0.462	0	1	1	-0.828	-1.319	0.027
BGR	200	0.315	0.466	0	1	1	0.791	-1.382	0.033
BRA	185	0.703	0.458	0	1	1	-0.88	-1.233	0.034
CHL	179	0.419	0.495	0	1	1	0.326	-1.905	0.037
COL	154	0.844	0.364	0	1	1	-1.879	1.542	0.029
CYP	88	0.341	0.477	0	1	1	0.66	-1.582	0.051
CZE	219	0.388	0.488	0	1	1	0.456	-1.8	0.033
DNK	141	0.489	0.502	0	1	1	0.042	-2.012	0.042
ESP	399	0.346	0.476	0	1	1	0.646	-1.587	0.024
EST	195	0.508	0.501	0	1	1	-0.031	-2.009	0.036
FIN	148	0.155	0.364	0	1	1	1.883	1.557	0.03
FRA	176	0.761	0.427	0	1	1	-1.216	-0.524	0.032
GEO	192	0.328	0.471	0	1	1	0.726	-1.48	0.034
HRV	188	0.255	0.437	0	1	1	1.113	-0.764	0.032
HUN	189	0.487	0.501	0	1	1	0.053	-2.008	0.036
JPN	196	0.526	0.501	0	1	1	-0.101	-2	0.036
KAZ	331	0.332	0.472	0	1	1	0.709	-1.502	0.026
KOR	163	0.436	0.497	0	1	1	0.257	-1.946	0.039
LTU	195	0.303	0.461	0	1	1	0.853	-1.279	0.033
LVA	135	0.489	0.502	0	1	1	0.044	-2.013	0.043
MEX	193	0.503	0.501	0	1	1	-0.01	-2.01	0.036
MLT	55	0.436	0.501	0	1	1	0.25	-1.973	0.067
NOR	185	0.335	0.473	0	1	1	0.693	-1.528	0.035
NZL	185	0.411	0.493	0	1	1	0.36	-1.881	0.036
PRT	200	0.6	0.491	0	1	1	-0.405	-1.845	0.035
RUS	230	0.261	0.44	0	1	1	1.082	-0.833	0.029
SAU	179	0.726	0.447	0	1	1	-1.006	-0.993	0.033
SVK	176	0.295	0.458	0	1	1	0.889	-1.216	0.034
SVN	132	0.379	0.487	0	1	1	0.494	-1.769	0.042
SWE	180	1	0	1	1	0	NA	NA	0
TUR	196	0.469	0.5	0	1	1	0.122	-1.995	0.036
TWN	200	0.215	0.412	0	1	1	1.377	-0.104	0.029
USA	165	0.327	0.471	0	1	1	0.73	-1.477	0.037
VNM	196	0.883	0.323	0	1	1	-2.36	3.587	0.023
ZAF	170	0.641	0.481	0	1	1	-0.583	-1.669	0.037

Table 7

LACK_RESOUR (Dummy Coded, Not a Problem=0, Else)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	0.589	0.492	0	1	1	-0.362	-1.873	0.022
AUT	246	0.488	0.501	0	1	1	0.048	-2.006	0.032
BEL	302	0.656	0.476	0	1	1	-0.652	-1.58	0.027
BGR	200	0.495	0.501	0	1	1	0.02	-2.01	0.035
BRA	185	0.822	0.384	0	1	1	-1.667	0.782	0.028
CHL	179	0.492	0.501	0	1	1	0.033	-2.01	0.037
COL	154	0.883	0.322	0	1	1	-2.362	3.601	0.026
CYP	88	0.386	0.49	0	1	1	0.459	-1.81	0.052
CZE	219	0.493	0.501	0	1	1	0.027	-2.008	0.034
DNK	141	0.61	0.49	0	1	1	-0.446	-1.814	0.041
ESP	399	0.539	0.499	0	1	1	-0.155	-1.981	0.025
EST	195	0.503	0.501	0	1	1	-0.01	-2.01	0.036
FIN	148	0.439	0.498	0	1	1	0.243	-1.954	0.041
FRA	176	0.631	0.484	0	1	1	-0.537	-1.721	0.036
GEO	192	0.615	0.488	0	1	1	-0.467	-1.791	0.035
HRV	188	0.553	0.498	0	1	1	-0.212	-1.965	0.036
HUN	189	0.741	0.439	0	1	1	-1.09	-0.816	0.032
JPN	196	0.597	0.492	0	1	1	-0.392	-1.856	0.035
KAZ	331	0.523	0.5	0	1	1	-0.09	-1.998	0.027
KOR	163	0.601	0.491	0	1	1	-0.41	-1.843	0.038
LTU	195	0.544	0.499	0	1	1	-0.174	-1.98	0.036
LVA	135	0.533	0.501	0	1	1	-0.132	-1.997	0.043
MEX	193	0.751	0.433	0	1	1	-1.154	-0.672	0.031
MLT	55	0.455	0.503	0	1	1	0.178	-2.004	0.068
NOR	185	0.465	0.5	0	1	1	0.14	-1.991	0.037
NZL	185	0.432	0.497	0	1	1	0.271	-1.937	0.037
PRT	200	0.775	0.419	0	1	1	-1.307	-0.293	0.03
RUS	230	0.496	0.501	0	1	1	0.017	-2.008	0.033
SAU	179	0.86	0.348	0	1	1	-2.062	2.263	0.026
SVK	176	0.653	0.477	0	1	1	-0.639	-1.6	0.036
SVN	132	0.295	0.458	0	1	1	0.886	-1.223	0.04
SWE	180	0.433	0.497	0	1	1	0.267	-1.939	0.037
TUR	196	0.612	0.488	0	1	1	-0.457	-1.8	0.035
TWN	200	0.355	0.48	0	1	1	0.602	-1.646	0.034
USA	165	0.43	0.497	0	1	1	0.279	-1.934	0.039
VNM	196	0.867	0.34	0	1	1	-2.149	2.633	0.024
ZAF	170	0.824	0.382	0	1	1	-1.682	0.835	0.029

Table 8

NENRSTUD (Number of Enrolled Students)

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	467	3.443	1.411	1	5	4	-0.205	-1.397	0.065
AUT	241	2.274	1.144	1	5	4	0.469	-0.866	0.074
BEL	288	3.219	1.324	1	5	4	-0.117	-1.147	0.078
BGR	196	2.219	1.308	1	5	4	0.756	-0.63	0.093
BRA	182	2.841	1.411	1	5	4	0.235	-1.234	0.105
CHL	168	2.542	1.422	1	5	4	0.5	-1.063	0.11
COL	136	4.331	1.155	1	5	4	-1.607	1.481	0.099
CYP	87	2.011	0.896	1	5	4	1.322	2.45	0.096
CZE	217	2.235	0.955	1	5	4	0.345	-0.42	0.065
DNK	132	2.5	1.023	1	5	4	0.213	-0.477	0.089
ESP	396	3.093	1.246	1	5	4	0.09	-1.033	0.063
EST	194	1.933	1.264	1	5	4	1.03	-0.353	0.091
FIN	148	2.142	0.99	1	5	4	0.678	-0.196	0.081
FRA	173	2.786	0.962	1	5	4	0.434	0.001	0.073
GEO	176	2.114	1.45	1	5	4	1.001	-0.481	0.109
HRV	181	2.586	1.09	1	5	4	0.306	-0.675	0.081
HUN	181	2.088	1.04	1	5	4	0.827	0.098	0.077
JPN	191	2.246	1.019	1	5	4	0.654	-0.002	0.074
KAZ	331	2.867	1.547	1	5	4	0.212	-1.452	0.085
KOR	148	2.831	1.203	1	5	4	0.045	-0.922	0.099
LTU	193	2.301	1.165	1	5	4	0.701	-0.304	0.084
LVA	133	2.414	1.232	1	5	4	0.488	-0.788	0.107
MEX	186	2.478	1.364	1	5	4	0.547	-0.863	0.1
MLT	50	2.36	0.875	1	4	3	0.333	-0.614	0.124
NOR	154	1.643	0.683	1	4	3	0.703	-0.16	0.055
NZL	177	3.017	1.44	1	5	4	0.107	-1.296	0.108
PRT	200	3.595	1.36	1	5	4	-0.354	-1.288	0.096
RUS	230	2.974	1.538	1	5	4	-0.021	-1.473	0.101
SAU	169	1.586	0.736	1	4	3	1.085	0.598	0.057
SVK	176	2.011	0.997	1	5	4	0.631	-0.443	0.075
SVN	116	2.198	0.925	1	5	4	0.517	0.034	0.086
SWE	161	2.292	0.985	1	5	4	0.604	-0.047	0.078
TUR	191	2.822	1.569	1	5	4	0.237	-1.493	0.114
TWN	199	3.628	1.505	1	5	4	-0.586	-1.189	0.107
USA	159	3.447	1.31	1	5	4	-0.22	-1.173	0.104
VNM	196	2.898	1.351	1	5	4	0.321	-1.158	0.097
ZAF	158	3.323	1.455	1	5	4	-0.222	-1.353	0.116

Table 9

TC3G17A-Students First Language Different from Instr. Language

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	443	2.684	1.762	1	5	4	0.359	-1.668	0.084
AUT	243	2.753	1.023	1	5	4	0.92	-0.102	0.066
BEL	281	2.601	0.936	1	5	4	1.124	0.477	0.056
BGR	198	2.889	1.541	1	5	4	0.31	-1.477	0.109
BRA	182	1.148	0.44	1	5	4	4.578	31.14	0.033
CHL	168	1.363	0.613	1	5	4	2.23	7.705	0.047
COL	136	1.221	0.617	1	5	4	3.962	19.081	0.053
CYP	86	2.64	1.051	1	5	4	0.988	0.02	0.113
CZE	215	1.726	0.542	1	4	3	0.268	1.506	0.037
DNK	131	2.359	0.745	1	5	4	1.317	1.779	0.065
ESP	394	2.135	0.741	1	5	4	1.539	4.121	0.037
EST	194	1.825	0.789	1	5	4	1.452	3.829	0.057
FIN	148	2.115	0.685	1	4	3	0.86	1.37	0.056
FRA	170	2.165	0.782	1	5	4	1.698	3.979	0.06
GEO	175	1.474	0.787	1	5	4	1.846	3.372	0.059
HRV	180	1.661	0.778	1	5	4	1.724	4.616	0.058
HUN	182	1.297	0.536	1	5	4	2.463	10.943	0.04
JPN	194	1.448	0.539	1	3	2	0.6	-0.857	0.039
KAZ	331	1.758	1.15	1	5	4	1.563	1.381	0.063
KOR	148	1.216	0.578	1	5	4	4.156	22.412	0.047
LTU	193	1.42	0.774	1	5	4	2.346	6.084	0.056
LVA	132	2.182	0.979	1	5	4	1.183	1.335	0.085
MEX	191	1.225	0.577	1	5	4	3.376	14.043	0.042
MLT	50	2.08	0.778	1	5	4	1.394	3.233	0.11
NOR	154	2.234	0.748	1	5	4	1.739	4.403	0.06
NZL	176	2.318	0.822	1	5	4	1.387	2.356	0.062
PRT	199	1.935	0.483	1	4	3	0.094	2.398	0.034
RUS	230	1.87	1.094	1	5	4	1.653	2.262	0.072
SAU	169	1.284	0.619	1	5	4	2.727	9.318	0.048
SVK	176	1.648	0.969	1	5	4	1.871	3.371	0.073
SVN	115	2.017	0.635	1	5	4	1.209	4.458	0.059
SWE	157	2.828	0.969	1	5	4	0.766	-0.336	0.077
TUR	191	1.995	1.185	1	5	4	1.387	1.137	0.086
TWN	197	2.269	1.037	1	5	4	1.225	1.21	0.074
USA	159	2.403	1.032	1	5	4	1.205	0.857	0.082
VNM	196	1.872	1.464	1	5	4	1.411	0.315	0.105
ZAF	159	3.44	1.637	1	5	4	-0.325	-1.616	0.13

Table 10

TC3G04A- Principal Years of Experience in This School

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	471	5.482	5.882	0	37	37	2.116	5.549	0.271
AUT	243	7.823	5.921	0	43	43	1.872	6.225	0.38
BEL	300	6.943	6.391	0	39	39	1.905	5.137	0.369
BGR	198	10.899	8.875	0	37	37	0.638	-0.657	0.631
BRA	183	7.027	8.305	0	45	45	1.838	3.423	0.614
CHL	168	7.601	7.993	0	46	46	2.08	5.007	0.617
COL	135	7.948	9.165	0	50	50	2.071	5.063	0.789
CYP	87	4.333	4.896	0	26	26	2.528	6.508	0.525
CZE	217	10.71	7.128	1	34	33	0.641	-0.207	0.484
DNK	133	6.654	6.397	0	39	39	1.587	3.661	0.555
ESP	396	6.972	6.632	0	39	39	1.978	4.439	0.333
EST	195	9.944	8.471	0	36	36	0.95	0.135	0.607
FIN	148	7.284	6.471	0	32	32	1.246	1.246	0.532
FRA	174	3.667	3.21	0	23	23	2.858	12.717	0.243
GEO	176	8.693	7.634	0	45	45	2.024	5.045	0.575
HRV	180	9.644	7.464	0	34	34	0.907	-0.035	0.556
HUN	182	8.962	7.123	0	38	38	1.189	1.375	0.528
JPN	195	2.856	2.636	0	23	23	3.896	21.465	0.189
KAZ	330	5.803	5.821	0	35	35	1.709	2.97	0.32
KOR	148	2.176	2.906	0	34	34	8.935	94.311	0.239
LTU	194	14.108	9.413	0	38	38	0.36	-0.82	0.676
LVA	134	11.925	9.406	0	35	35	0.592	-0.758	0.813
MEX	193	5.824	7.378	0	28	28	1.599	1.453	0.531
MLT	51	5.686	5.03	0	27	27	1.991	4.856	0.704
NOR	153	5.484	4.626	0	22	22	1.303	1.384	0.374
NZL	182	6.577	6.314	0	31	31	1.301	1.163	0.468
PRT	200	8.72	7.933	0	40	40	1.606	2.581	0.561
RUS	230	9.852	8.224	0	38	38	1.175	0.903	0.542
SAU	174	5.287	5.46	0	24	24	1.679	1.996	0.414
SVK	176	9.261	6.991	0	34	34	0.992	0.655	0.527
SVN	117	9.47	7.564	0	31	31	0.846	0.008	0.699
SWE	166	4.151	3.704	0	18	18	1.389	1.774	0.287
TUR	194	3.191	2	0	12	12	1.25	2.261	0.144
TWN	199	3.945	2.987	0	26	26	2.773	15.001	0.212
USA	159	4.761	4.785	0	30	30	2.447	7.761	0.379
VNM	196	4.883	3.985	0	28	28	1.794	5.394	0.285
ZAF	161	5.652	5.435	0	28	28	1.675	3.054	0.428

Table 11

AVG_TT3G11A- The Average Years of Experience of Teachers in the Particular School

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	5.339	2.527	0.714	18	17.286	1.038	2.014	0.111
AUT	246	13.017	3.841	1.889	25	23.111	0.003	0.188	0.245
BEL	302	12.239	3.655	1.944	27.438	25.493	0.36	0.863	0.21
BGR	200	13.785	4.937	1.667	26.625	24.958	0.078	-0.27	0.349
BRA	185	8.072	3.642	1.333	21.529	20.196	0.6	0.334	0.268
CHL	179	7.802	3.699	1	19.857	18.857	0.725	0.237	0.276
COL	154	9.871	4.646	1.4	24	22.6	0.437	-0.121	0.374
CYP	88	4.139	2.479	1.167	13.25	12.083	1.76	2.656	0.264
CZE	219	12.803	3.94	4.286	29.5	25.214	0.595	1.488	0.266
DNK	141	10.188	2.949	4	21.667	17.667	0.366	0.575	0.248
ESP	399	8.939	4.34	1.05	21.2	20.15	0.659	-0.314	0.217
EST	195	14.547	5.572	2	27.696	25.696	-0.173	-0.587	0.399
FIN	148	10.241	2.598	4	18.588	14.588	0.199	0.14	0.214
FRA	176	9.047	3.007	2.4	20.8	18.4	0.452	0.743	0.227
GEO	192	18.278	5.77	2.364	35.444	33.081	-0.402	0.429	0.416
HRV	188	11.319	3.109	3.062	21.727	18.665	0.08	0.232	0.227
HUN	189	13.726	4.268	1.7	26	24.3	0.036	0.052	0.31
JPN	196	4.604	3.173	1.778	22	20.222	2.747	7.67	0.227
KAZ	331	10.264	4.339	1.6	22.45	20.85	0.21	-0.258	0.238
KOR	163	4.875	5.191	1.65	27.588	25.938	2.368	4.897	0.407
LTU	195	16.541	4.48	2.412	28.65	26.238	-0.456	0.67	0.321
LVA	135	16.115	4.991	4.65	29.222	24.572	0.131	-0.4	0.43
MEX	193	9.893	4.307	0.667	21.421	20.754	0.149	-0.209	0.31
MLT	55	7.568	3.039	1.5	14.464	12.964	0.046	-0.315	0.41
NOR	185	10.418	3.375	2	21.3	19.3	0.136	0.245	0.248
NZL	183	7.712	3.237	1	21.5	20.5	0.698	1.946	0.239
PRT	200	11.169	4.123	2	24.138	22.138	0.238	-0.61	0.292
RUS	230	15.149	5.179	2.9	30.125	27.225	0.13	-0.056	0.342
SAU	179	6.457	2.983	1.091	15.647	14.556	0.512	-0.104	0.223
SVK	176	12.442	3.739	2.273	21.4	19.127	-0.332	-0.273	0.282
SVN	132	17.139	3.591	4.143	26.25	22.107	-0.35	0.921	0.313
SWE	180	8.242	3.181	2	20.75	18.75	0.444	0.142	0.237
TUR	196	3.972	1.717	1	12.571	11.571	1.159	2.759	0.123
TWN	200	11.275	2.944	3.211	18.842	15.632	-0.378	-0.041	0.208
USA	165	8.338	4.245	1.941	37.5	35.559	2.217	11.797	0.33
VNM	196	10.672	3.504	2	20.25	18.25	-0.027	-0.072	0.25
ZAF	170	8.944	4.201	0.25	21.778	21.528	0.776	0.469	0.322

Table 12

CLIMATEB – Climate Reported by Teachers at the School Level

CNTRY	n	mean	sd	min	max	range	skew	kurtosis	se
ARE	521	0	0.172	-0.684	0.427	1.111	-0.356	0.213	0.008
AUT	246	-0.048	0.203	-0.539	0.521	1.06	0.449	-0.183	0.013
BEL	302	-0.002	0.175	-0.573	0.401	0.974	-0.624	0.5	0.01
BGR	200	-0.01	0.128	-0.416	0.407	0.823	0.173	0.269	0.009
BRA	185	-0.016	0.157	-0.481	0.519	1	-0.196	0.417	0.012
CHL	179	-0.004	0.209	-0.562	0.474	1.036	-0.204	-0.42	0.016
COL	154	-0.03	0.155	-0.504	0.397	0.901	-0.065	0.258	0.012
CYP	88	0.001	0.091	-0.204	0.206	0.41	0.291	-0.144	0.01
CZE	219	-0.013	0.138	-0.467	0.308	0.775	-0.245	-0.384	0.009
DNK	141	-0.027	0.2	-0.464	0.529	0.993	0.308	-0.384	0.017
ESP	399	-0.024	0.135	-0.402	0.383	0.785	0.04	-0.197	0.007
EST	195	0	0.13	-0.395	0.422	0.817	0.143	0.253	0.009
FIN	148	0.002	0.184	-0.387	0.545	0.932	0.062	-0.336	0.015
FRA	176	-0.015	0.135	-0.363	0.325	0.688	-0.201	-0.214	0.01
GEO	192	0	0.128	-0.417	0.345	0.762	-0.03	0.213	0.009
HRV	188	-0.01	0.136	-0.398	0.348	0.746	0.033	-0.134	0.01
HUN	189	-0.024	0.188	-0.528	0.698	1.226	-0.002	0.369	0.014
JPN	196	-0.017	0.198	-0.589	0.525	1.114	0.05	0.125	0.014
KAZ	331	-0.004	0.084	-0.273	0.241	0.514	-0.012	0.231	0.005
KOR	163	-0.012	0.185	-0.434	0.594	1.028	0.256	0.48	0.014
LTU	195	-0.011	0.123	-0.309	0.466	0.775	0.716	1.524	0.009
LVA	135	-0.005	0.127	-0.419	0.307	0.726	-0.184	0.32	0.011
MEX	193	-0.019	0.205	-0.786	0.466	1.252	-0.334	0.715	0.015
MLT	55	0.001	0.194	-0.406	0.369	0.775	-0.129	-0.861	0.026
NOR	185	-0.017	0.225	-1.095	0.562	1.657	-0.628	2.403	0.017
NZL	185	-0.104	0.29	-0.836	0.556	1.392	-0.185	-0.408	0.021
PRT	200	-0.004	0.093	-0.302	0.318	0.62	-0.193	0.849	0.007
RUS	230	0.014	0.124	-0.347	0.41	0.757	0.448	0.757	0.008
SAU	179	-0.008	0.105	-0.242	0.269	0.511	-0.13	-0.553	0.008
SVK	176	-0.016	0.164	-0.421	0.41	0.831	-0.149	-0.543	0.012
SVN	132	-0.004	0.124	-0.317	0.317	0.634	0.007	-0.239	0.011
SWE	180	0.006	0.21	-0.557	0.484	1.041	-0.052	-0.302	0.016
TUR	196	-0.036	0.201	-0.656	0.451	1.107	-0.308	-0.022	0.014
TWN	200	0.004	0.108	-0.242	0.319	0.561	0.245	-0.104	0.008
USA	165	-0.101	0.225	-0.712	0.428	1.14	-0.14	-0.29	0.018
VNM	196	-0.001	0.113	-0.34	0.312	0.652	0.356	0.305	0.008
ZAF	170	-0.01	0.227	-0.829	0.716	1.545	-0.041	1.197	0.017

Table 13

CLIMATEP – Climate Reported by Principals

CNTRY	mean	sd	min	max	range	skew	kurtosis	se
ARE	0.16	0.415	-2.122	0.785	2.907	-0.254	0.608	0.018
AUT	0.109	0.273	-0.825	0.776	1.601	-0.315	0.227	0.017
BEL	0.057	0.213	-0.718	0.583	1.301	-0.29	0.723	0.012
BGR	-0.033	0.307	-1.576	0.551	2.127	-0.075	2.215	0.022
BRA	0.041	0.348	-1.146	0.822	1.968	0.202	0.787	0.026
CHL	0.275	0.406	-1.777	0.96	2.737	-0.572	2.462	0.03
COL	0.2	0.282	-0.501	0.742	1.243	0.23	-0.149	0.023
CYP	0.152	0.284	-0.347	0.715	1.062	0.668	-0.861	0.03
CZE	0.111	0.208	-0.587	0.713	1.3	0.189	1.866	0.014
DNK	0.086	0.314	-0.521	0.984	1.505	0.555	0.053	0.026
ESP	0.046	0.318	-1.507	0.747	2.254	-0.023	1.499	0.016
EST	0.021	0.298	-0.791	0.806	1.597	0.455	0.734	0.021
FIN	0.15	0.386	-1.097	0.997	2.094	-0.108	-0.155	0.032
FRA	-0.013	0.296	-0.916	0.641	1.557	-0.428	0.425	0.022
GEO	-0.069	0.308	-1.157	0.6	1.757	0.36	0.362	0.022
HRV	0.119	0.301	-1.138	0.92	2.058	0.006	1.086	0.022
HUN	0.125	0.34	-0.827	0.919	1.746	0.103	-0.052	0.025
JPN	0.246	0.222	-0.297	0.93	1.227	0.759	0.994	0.016
KAZ	0.142	0.281	-0.58	0.664	1.244	0.346	-0.98	0.015
KOR	0.505	0.386	-0.394	1.05	1.444	-0.028	-1.272	0.03
LTU	0.098	0.236	-0.891	0.606	1.497	-0.092	1.036	0.017
LVA	0.003	0.326	-0.937	0.962	1.899	0.513	2.787	0.028
MEX	0.132	0.288	-0.645	0.784	1.429	-0.043	0.049	0.021
MLT	0.04	0.308	-0.934	0.851	1.785	-0.611	1.806	0.042
NOR	0.031	0.325	-0.933	0.872	1.805	0.063	0.589	0.024
NZL	0.121	0.357	-1.14	0.816	1.956	-0.205	0.438	0.026
PRT	0.174	0.229	-0.425	0.69	1.115	0.353	-0.123	0.016
RUS	0.11	0.301	-0.627	0.721	1.348	0.274	-0.496	0.02
SAU	0.269	0.315	-0.788	0.735	1.523	-0.221	-0.727	0.024
SVK	0.09	0.306	-0.612	0.88	1.492	-0.119	-0.024	0.023
SVN	0.102	0.228	-0.545	0.739	1.284	0.382	1.251	0.02
SWE	0.058	0.364	-1.278	0.859	2.137	-0.373	0.385	0.027
TUR	0.189	0.415	-1.613	0.917	2.53	-0.628	2.314	0.03
TWN	0.192	0.262	-0.579	0.8	1.379	0.483	0.511	0.019
USA	0.235	0.336	-0.534	0.964	1.498	0.057	0.057	0.026
VNM	0.378	0.347	-1.851	0.779	2.63	-1.641	6.98	0.025
ZAF	0.126	0.378	-1.655	0.922	2.577	-0.495	2.848	0.029

Appendix 3

ICCs, Model fit and reliability omega coefficients

Table 1

Intraclass Correlation Coefficients 1 (ICC1) and Intraclass Correlation Coefficients 2 (ICC2) for Teacher Ratings of School Climate by Country

		TT3G48F*		TT3G48G*		TT3G48H*		TT3G49E*	
	Average cluster size	ICC1	ICC2	ICC1	ICC2	ICC1	ICC2	ICC1	ICC2
ARE	16.28	0.109	0.666	0.137	0.721	0.138	0.723	0.085	0.602
AUT	16.76	0.199	0.806	0.162	0.764	0.183	0.790	0.217	0.823
BEL	16.75	0.152	0.750	0.150	0.747	0.164	0.767	0.231	0.834
BGR	14.11	0.097	0.602	0.124	0.666	0.155	0.721	0.183	0.760
BRA	12.81	0.130	0.657	0.143	0.681	0.150	0.693	0.172	0.727
CHL	10.86	0.155	0.666	0.196	0.726	0.206	0.738	0.127	0.612
COL	15.32	0.157	0.740	0.170	0.758	0.142	0.717	0.187	0.779
CYP	17.81	0.082	0.614	0.111	0.690	0.090	0.638	0.088	0.632
CZE	15.52	0.142	0.720	0.094	0.617	0.136	0.710	0.164	0.753
DNK	13.35	0.185	0.752	0.144	0.692	0.196	0.765	0.182	0.748
ESP	18.35	0.143	0.754	0.107	0.687	0.136	0.743	0.199	0.820
EST	15.09	0.114	0.660	0.115	0.662	0.113	0.658	0.152	0.730
FIN	18.96	0.199	0.825	0.156	0.778	0.141	0.757	0.138	0.752
FRA	16.10	0.101	0.644	0.125	0.697	0.152	0.743	0.216	0.816
GEO	15.80	0.111	0.664	0.129	0.701	0.179	0.775	0.172	0.766
HRV	17.45	0.074	0.582	0.101	0.662	0.131	0.725	0.152	0.758
HUN	16.87	0.182	0.790	0.151	0.750	0.163	0.767	0.178	0.785
JPN	18.02	0.160	0.774	0.206	0.824	0.181	0.799	0.178	0.796
KAZ	19.81	0.062	0.567	0.041	0.459	0.095	0.675	0.087	0.654
KOR	17.69	0.173	0.787	0.141	0.744	0.192	0.808	0.157	0.767
LTU	19.17	0.068	0.583	0.094	0.665	0.098	0.676	0.160	0.785
LVA	16.53	0.074	0.569	0.118	0.689	0.066	0.539	0.091	0.623

MEX	15.10	0.111	0.653	0.160	0.742	0.176	0.763	0.248	0.833
MLT	28.69	0.116	0.790	0.164	0.849	0.065	0.666	0.110	0.780
NOR	21.25	0.244	0.873	0.212	0.851	0.152	0.792	0.225	0.861
NZL	11.46	0.265	0.805	0.345	0.858	0.236	0.780	0.209	0.752
PRT	17.94	0.096	0.656	0.090	0.640	0.112	0.694	0.085	0.625
RUS	17.34	0.120	0.703	0.110	0.682	0.162	0.770	0.159	0.766
SAU	13.63	0.063	0.478	0.067	0.495	0.081	0.546	0.162	0.725
SVK	16.83	0.112	0.680	0.103	0.659	0.136	0.726	0.179	0.786
SVN	15.45	0.063	0.510	0.082	0.580	0.108	0.652	0.148	0.729
SWE	14.06	0.243	0.819	0.181	0.757	0.226	0.804	0.231	0.809
TUR	19.96	0.104	0.699	0.137	0.760	0.144	0.771	0.170	0.803
TWN	19.08	0.095	0.667	0.079	0.621	0.139	0.755	0.123	0.728
USA	14.55	0.144	0.710	0.249	0.828	0.189	0.772	0.076	0.545
VNM	19.50	0.172	0.802	0.135	0.753	0.148	0.772	0.124	0.734
ZAF	11.92	0.169	0.708	0.196	0.744	0.139	0.658	0.206	0.756
AVG	16.65	0.134	0.695	0.141	0.708	0.146	0.724	0.161	0.744

Note. TT3G48F- School staff share common set of beliefs about teaching and learning; TT3G48G- Enforcing rules for student behaviour consistently throughout school; TT3G48H- This school encourages staff to lead new initiatives; TT3G49E- Teachers can rely on each other.

Table 2

The Model Fit of Multilevel Confirmatory Factor Analysis (MCFA) of Teacher Ratings of School Climate with Saturated Structure at Level 1. The Results are Displayed by Country.

Model	Chi2	df	CFI	TLI	RMSEA	SRMRw	SRMRb	Omega_b
Entire sample	14.986	2	0.999	0.991	0.007	0.001	0.029	
ARE	2.796	2	1	0.999	0.007	0	0.011	0.949
AUT	5.615	2	0.999	0.993	0.021	0.001	0.016	0.940
BEL	16.507	2	0.994	0.964	0.038	0.002	0.055	0.883
BGR	1.259	2	1	1	0	0.001	0.019	0.921
BRA	1.612	2	1	1	0	0.001	0.018	0.907
CHL	0.361	2	1	1	0	0.001	0.008	0.939
COL	0.305	2	1	1	0	0.001	0.014	0.909
CYP	6.561	2	0.996	0.975	0.038	0.002	0.058	0.876
CZE	0.562	2	1	1	0	0	0.013	0.923
DNK	1.053	2	1	1	0	0.001	0.014	0.926
ESP	0.953	2	1	1	0	0	0.018	0.867
EST	5.014	2	0.999	0.993	0.023	0.001	0.024	0.927
FIN	1.673	2	1	1	0	0.001	0.017	0.935
FRA	15.223	2	0.991	0.944	0.048	0.002	0.068	0.849
GEO	0.671	2	1	1	0	0	0.008	0.945
HRV	1.66	2	1	1	0	0.001	0.024	0.883
HUN	10.049	2	0.997	0.984	0.036	0.001	0.026	0.940
JPN	1.855	2	1	1	0	0.001	0.011	0.946
KAZ	0.737	2	1	1	0	0	0.009	0.972
KOR	10.665	2	0.997	0.983	0.039	0.002	0.03	0.954
LTU	0.783	2	1	1	0	0.001	0.016	0.901
LVA	11.292	2	0.994	0.966	0.046	0.001	0.024	0.932
MEX	2.781	2	1	0.997	0.012	0.001	0.02	0.926

MLT	3.788	2	0.999	0.991	0.024	0.001	0.052	0.889
NOR	2.611	2	1	0.999	0.009	0.001	0.012	0.964
NZL	0.145	2	1	1	0	0.001	0.006	0.960
PRT	0.629	2	1	1	0	0	0.015	0.869
RUS	878.048	2	0.591	0	0.331	0.001	0.013	0.945
SAU	13.456	2	0.993	0.961	0.048	0.003	0.072	0.920
SVK	0.396	2	1	1	0	0.001	0.012	0.916
SVN	6.675	2	0.998	0.985	0.034	0.001	0.034	0.934
SWE	5.328	2	0.998	0.988	0.026	0.001	0.02	0.956
TUR	13.669	2	0.995	0.97	0.039	0.002	0.037	0.933
TWN	0.908	2	1	1	0	0	0.013	0.920
USA	0.538	2	1	1	0	0.001	0.025	0.867
VNM	4.932	2	0.998	0.988	0.02	0.001	0.037	0.872
ZAF	5.493	2	0.998	0.986	0.029	0.001	0.016	0.970

Table 3

The Model Fit Statistics of Confirmatory Factor Analysis (CFA) of Principal Ratings of School Climate with Saturated Structure at Level 1. The Results are Displayed by Country.

Model	Chi2	df	CFI	TLI	RMSEA	SRMRw	Omega
Entire sample	13.277	2	0.976	0.929	0.028	0.027	0.747
ARE	12.446	2	0.97	0.911	0.106	0.023	0.830
AUT	1.127	2	1	1	0	0.016	0.652
BEL	5.478	2	0.973	0.918	0.078	0.035	0.690
BGR	2.732	2	0.994	0.981	0.043	0.023	0.783
BRA	5.287	2	0.965	0.896	0.095	0.029	0.745
CHL	3.774	2	0.985	0.954	0.073	0.022	0.782
COL*	13.002	2	0.849	0.547	0.201	0.07	0.777
CYP	2.016	2	1	0.999	0.01	0.022	0.762
CZE	0.472	2	1	1	0	0.009	0.647
DNK	5.77	2	0.955	0.866	0.116	0.046	0.724
ESP	1.082	2	1	1	0	0.011	0.799
EST	1.659	2	1	1	0	0.017	0.691
FIN	3.107	2	0.991	0.972	0.061	0.022	0.772
FRA	0.845	2	1	1	0	0.014	0.771
GEO	5.727	2	0.972	0.916	0.103	0.029	0.770
HRV	6.356	2	0.93	0.791	0.108	0.046	0.734
HUN*	23.596	2	0.89	0.669	0.245	0.043	0.795
JPN	2.527	2	0.99	0.971	0.037	0.024	0.625
KAZ	4.806	2	0.977	0.93	0.065	0.024	0.779
KOR	1.223	2	1	1	0	0.014	0.826
LTU	5.142	2	0.948	0.844	0.09	0.036	0.678
LVA*	15.835	2	0.846	0.539	0.226	0.045	0.734
MEX	0.473	2	1	1	0	0.01	0.714
MLT*’	NA	NA	NA	NA	NA	0.139	0.552
NOR	3.14	2	0.987	0.962	0.06	0.032	0.737

NZL	3.649	2	0.974	0.922	0.067	0.036	0.759
PRT	0.384	2	1	1	0	0.009	0.665
RUS	1.459	2	1	1	0	0.012	0.811
SAU	1.643	2	1	1	0	0.02	0.767
SVK	6.355	2	0.91	0.729	0.11	0.037	0.659
SVN*	10.317	2	0.889	0.668	0.189	0.055	0.676
SWE	1.418	2	1	1	0	0.019	0.777
TUR	0.841	2	1	1	0	0.015	0.788
TWN	1.134	2	1	1	0	0.015	0.733
USA*’	NA	NA	NA	NA	NA	0.087	0.673
VNM	5.257	2	0.966	0.897	0.091	0.027	0.726
ZAF	1.598	2	1	1	0	0.019	0.788

Note. Models with * did not fit data according to any statistics. Models with ‘ were not calculated due to negative correction factor.

Appendix 4

Within country analysis on the association between school climate as reported by principals and school leadership, and school climate as reported by teachers and school leadership

Table 1

Standardized Regression Coefficients Showing the Association between Leadership Style (T3PLEADS and T3PLEADP), Climate Reported by Principals (CLIMATEP), (Model 1)

CNTRY	INTERCEPT	T3PLEADS	T3PLEADP	R^2
ARE	-2.871*(0.290)	0.142*(0.052)	0.472*(0.057)	0.283*(0.049)
AUT	-2.038*(0.426)	0.099(0.097)	0.357*(0.085)	0.149*(0.051)
BEL	-1.602*(0.681)	0.011(0.084)	0.300*(0.078)	0.090(0.047)
BGR	-2.604*(0.449)	0.149(0.080)	0.395*(0.090)	0.206*(0.069)
BRA	-5.280*(0.594)	0.265*(0.074)	0.404*(0.081)	0.270*(0.067)
CHL	-2.212*(0.340)	0.158*(0.078)	0.475*(0.065)	0.284*(0.055)
COL	-2.074*(0.484)	0.032(0.099)	0.503*(0.102)	0.260*(0.093)
CYP	-1.710*(0.703)	0.310*(0.110)	0.136(0.111)	0.140(0.081)
CZE	-2.767*(0.670)	0.116(0.082)	0.277*(0.067)	0.102*(0.040)
DNK	-2.977*(1.065)	0.197*(0.071)	0.304*(0.145)	0.144(0.082)
ESP	-1.652*(0.508)	0.248*(0.069)	0.144(0.086)	0.096(0.053)
EST	-2.300*(0.569)	0.032(0.063)	0.354*(0.073)	0.131*(0.054)
FIN	-3.046*(0.679)	0.088(0.077)	0.358*(0.081)	0.151*(0.061)
FRA	-1.125(0.757)	0.135(0.089)	0.074(0.090)	0.029(0.035)
GEO	-3.140*(0.608)	0.045(0.070)	0.452*(0.075)	0.214*(0.068)
HRV	-2.640*(0.595)	0.081(0.100)	0.461*(0.091)	0.239*(0.080)
HUN	-1.651(0.871)	0.030(0.104)	0.302*(0.082)	0.098*(0.044)
JPN	-2.515*(0.903)	0.153(0.087)	0.240*(0.088)	0.094*(0.046)
KAZ	-2.953*(0.550)	0.195*(0.068)	0.328*(0.081)	0.165*(0.054)
KOR	-2.881*(0.383)	0.212*(0.070)	0.537*(0.070)	0.376*(0.069)

LTU	-3.427*(0.533)	0.131*(0.061)	0.521*(0.072)	0.302*(0.075)
LVA	-3.932*(0.834)	0.175(0.102)	0.337*(0.127)	0.158*(0.077)
MEX	-1.473*(0.427)	0.240*(0.084)	0.155*(0.072)	0.096*(0.042)
MLT	-1.515(1.191)	0.304*(0.145)	0.008(0.160)	0.094(0.091)
NOR	-3.100*(0.894)	0.228*(0.112)	0.205*(0.078)	0.114(0.067)
NZL	-3.667*(0.921)	0.160(0.092)	0.443*(0.130)	0.214*(0.105)
PRT	-2.382*(0.485)	0.135(0.080)	0.418*(0.065)	0.207*(0.057)
RUS	-0.340(0.965)	-0.052(0.125)	0.241(0.153)	0.052(0.066)
SAU	-1.606*(0.624)	0.288*(0.075)	0.191*(0.094)	0.134*(0.064)
SVK	-1.298(1.111)	-0.011(0.086)	0.197(0.101)	0.039(0.040)
SVN	-2.030*(0.661)	0.065(0.087)	0.315*(0.085)	0.117*(0.058)
SWE	-2.763*(1.173)	-0.075(0.117)	0.353*(0.103)	0.122(0.070)
TUR	-2.353*(0.578)	0.081(0.082)	0.491*(0.168)	0.265(0.150)
TWN	-3.542*(0.642)	0.264*(0.066)	0.312*(0.076)	0.211*(0.063)
USA	-0.892(1.024)	0.194(0.169)	0.136(0.112)	0.084(0.095)
VNM	-2.502*(0.926)	0.196*(0.092)	0.277*(0.123)	0.130(0.069)
ZAF	-2.873*(0.562)	0.235*(0.086)	0.350*(0.090)	0.193*(0.070)

Table 2

Standardized Regression Coefficients Showing the Association between Leadership Style (T3PLEADS and T3PLEADP), Climate Reported by Teachers (CLIMATEB), (Model 1)

CNTRY	INTERCEPT	T3PLEADS	T3PLEADP	R^2
ARE	-1.883*(0.322)	0.076(0.047)	0.283*(0.050)	0.099*(0.030)
AUT	-0.293(0.572)	-0.158(0.107)	0.190*(0.091)	0.051(0.048)
BEL	-2.255*(0.614)	0.084(0.067)	0.288*(0.088)	0.094(0.050)
BGR	-0.213(0.486)	-0.020(0.082)	0.064(0.061)	0.004(0.007)
BRA	-2.697*(0.848)	0.182*(0.089)	0.164*(0.078)	0.070(0.046)
CHL	-1.643*(0.424)	0.018(0.078)	0.363*(0.066)	0.135*(0.047)
COL	-1.709*(0.596)	0.047(0.131)	0.283*(0.118)	0.087(0.058)

CYP	-1.078(0.610)	0.094(0.122)	0.118(0.092)	0.029(0.033)
CZE	-0.912(0.696)	-0.024(0.072)	0.119(0.065)	0.014(0.015)
DNK	-2.854*(0.942)	0.135(0.082)	0.277*(0.130)	0.103(0.068)
ESP	-1.157*(0.451)	0.135(0.071)	0.103(0.077)	0.034(0.027)
EST	-0.571(0.571)	-0.003(0.071)	0.096(0.077)	0.009(0.014)
FIN	-1.244(0.891)	-0.072(0.075)	0.222*(0.090)	0.046(0.037)
FRA	0.478(0.555)	-0.108(0.084)	0.025(0.075)	0.011(0.017)
GEO	-1.222(0.779)	-0.001(0.080)	0.214*(0.094)	0.046(0.041)
HRV	-2.466*(0.618)	0.081(0.087)	0.362*(0.098)	0.153*(0.072)
HUN	-0.226(0.940)	-0.067(0.095)	0.156(0.082)	0.022(0.023)
JPN	-2.187*(0.936)	0.080(0.093)	0.155(0.083)	0.035(0.029)
KAZ	-1.895*(0.665)	0.138(0.078)	0.131(0.069)	0.042(0.029)
KOR	-0.648(0.706)	-0.043(0.104)	0.139(0.100)	0.019(0.026)
LTU	-1.238(0.899)	0.047(0.097)	0.159(0.089)	0.029(0.031)
LVA	-2.394*(0.834)	0.130(0.087)	0.178(0.115)	0.054(0.041)
MEX	-1.213*(0.448)	0.232*(0.082)	0.001(0.075)	0.054(0.036)
MLT	-0.698(0.952)	0.002(0.128)	0.112(0.128)	0.013(0.028)
NOR	-2.590*(1.078)	0.236(0.133)	0.135(0.077)	0.088(0.075)
NZL	-3.079*(0.808)	0.082(0.083)	0.344*(0.110)	0.121(0.072)
PRT	-0.830(0.530)	0.230*(0.072)	-0.063(0.063)	0.053(0.032)
RUS	0.420(0.898)	-0.105(0.124)	0.086(0.117)	0.012(0.025)
SAU	-1.358*(0.551)	0.088(0.075)	0.187*(0.086)	0.047(0.038)
SVK	-1.938*(0.803)	0.044(0.067)	0.187*(0.083)	0.037(0.031)
SVN	-1.437*(0.726)	0.015(0.096)	0.203*(0.089)	0.043(0.037)
SWE	-1.968(1.136)	-0.064(0.125)	0.249*(0.091)	0.061(0.043)
TUR	-0.257(0.779)	0.093(0.106)	-0.074(0.111)	0.011(0.023)
TWN	-1.809*(0.670)	0.136(0.073)	0.114(0.078)	0.040(0.029)
USA	-1.026(0.894)	-0.053(0.114)	0.183(0.140)	0.026(0.040)
VNM	-2.327*(0.922)	0.206*(0.077)	0.024(0.091)	0.044(0.033)
ZAF	-1.56*(0.708)	0.045(0.089)	0.239*(0.094)	0.061(0.048)

Appendix 5

Measurement invariance testing across teachers and principals by country

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
1	ARE CONFIGURAL	34,694	15	1,13		0,998	0,012	0	0,027						0
2	ARE METRIC	57,345	18	1,0944	3	0,996	0,016	0,001	0,082	-0,002	0,004	0,001	0,055	25,70	1
3	ARE SCALAR	133,393	21	1,0751	3	0,988	0,025	0,001	0,147	-0,008	0,009	0,065	0,065	84,07	1
4	AUT CONFIGURAL	35,405	15	1,2778		0,994	0,018	0,001	0,07						0
5	AUT METRIC	33,085	18	1,455	3	0,995	0,014	0,001	0,084	0,001	-0,004	0	0,014	1,24	0
6	AUT SCALAR	73,736	21	1,5223	3	0,984	0,024	0,001	0,176	-0,011	0,01	0,092	0,092	33,28	1
7	BEL CONFIGURAL	47,224	15	1,4064		0,988	0,02	0,002	0,072						0
8	BEL METRIC	81,178	18	1,3686	3	0,977	0,026	0,002	0,133	-0,011	0,006	0	0,061	37,88	1
9	BEL SCALAR	163,555	21	1,3333	3	0,948	0,036	0,002	0,274	-0,029	0,01	0,141	0,141	95,38	1
10	BGR CONFIGURAL	14,2	15	1,5901		1	0	0,001	0,051						0
11	BGR METRIC	14,201	18	1,6188	3	1	0	0,001	0,052	0	0	0	0,001	0,23	0
12	BGR SCALAR	38,87	21	1,593	3	0,992	0,017	0,002	0,088	-0,008	0,017	0,036	0,036	27,07	1
13	BRA CONFIGURAL	21,978	15	1,2192		0,997	0,014	0,001	0,037						0
14	BRA METRIC	26,754	18	1,2263	3	0,996	0,014	0,002	0,065	-0,001	0	0,001	0,028	4,77	0
15	BRA SCALAR	55,171	21	1,2267	3	0,983	0,026	0,001	0,135	-0,013	0,012	0,07	0,07	28,37	1

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
16	CHL CONFIGURAL	14,279	15	1,1265		1	0	0,001	0,034						0
17	CHL METRIC	19,261	18	1,1429	3	0,999	0,006	0,002	0,07	-0,001	0,006	0,001	0,036	4,84	0
18	CHL SCALAR	53,859	21	1,1257	3	0,98	0,028	0,003	0,152	-0,019	0,022	0,082	0,082	37,77	1
19	COL CONFIGURAL	26,004	15	1,893		0,991	0,018	0,001	0,062						0
20	COL METRIC	31,127	18	1,9927	3	0,99	0,017	0,002	0,092	-0,001	-0,001	0,001	0,03	5,14	0
21	COL SCALAR	45,244	21	2,0329	3	0,981	0,022	0,002	0,127	-0,009	0,005	0,035	0,035	13,17	1
22	CYP CONFIGURAL	22,287	15	0,9118		0,995	0,017	0,002	0,079						0
23	CYP METRIC	28,537	18	0,9129	3	0,993	0,019	0,003	0,116	-0,002	0,002	0,001	0,037	6,24	0
24	CYP SCALAR	42,032	21	0,9348	3	0,985	0,025	0,002	0,162	-0,008	0,006	0,046	0,046	12,42	1
25	CZE CONFIGURAL	14,757	15	1,1103		1	0	0,001	0,052						0
26	CZE METRIC	23,36	18	1,1284	3	0,998	0,009	0,001	0,076	-0,002	0,009	0	0,024	8,18	1
27	CZE SCALAR	63,607	21	1,1019	3	0,984	0,024	0,002	0,14	-0,014	0,015	0,064	0,064	46,38	1
28	DNK CONFIGURAL	22,365	15	1,1734		0,993	0,016	0,001	0,054						0
29	DNK METRIC	26,041	18	1,188	3	0,993	0,015	0,001	0,066	0	-0,001	0	0,012	3,72	0
30	DNK SCALAR	68,998	21	1,1604	3	0,957	0,034	0,002	0,186	-0,036	0,019	0,12	0,12	49,39	1
31	ESP CONFIGURAL	15,39	15	2,1252		1	0,002	0,001	0,038						0
32	ESP METRIC	19,866	18	2,1501	3	0,999	0,004	0,001	0,059	-0,001	0,002	0	0,021	4,40	0
33	ESP SCALAR	104,55	21	2,1902	3	0,971	0,023	0,001	0,192	-0,028	0,019	0,133	0,133	76,63	1

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
34	EST CONFIGURAL	23,267	15	0,8893		0,998	0,014	0,001	0,053						0
35	EST METRIC	31,01	18	0,8871	3	0,996	0,016	0,001	0,083	-0,002	0,002	0	0,03	7,78	0
36	EST SCALAR	102,57	21	0,8891	3	0,976	0,036	0,001	0,247	-0,02	0,02	0,164	0,164	70,68	1
37	FIN CONFIGURAL	20,986	15	1,0588		0,997	0,012	0,001	0,063						0
38	FIN METRIC	22,76	18	1,0784	3	0,998	0,01	0,001	0,074	0,001	-0,002	0	0,011	1,98	0
39	FIN SCALAR	44,589	21	1,062	3	0,99	0,02	0,001	0,136	-0,008	0,01	0,062	0,062	23,67	1
40	FRA CONFIGURAL	23,042	15	1,1613		0,995	0,013	0,002	0,062						0
41	FRA METRIC	29,379	18	1,1781	3	0,993	0,015	0,003	0,088	-0,002	0,002	0,001	0,026	6,22	0
42	FRA SCALAR	120,712	21	1,1467	3	0,941	0,04	0,003	0,22	-0,052	0,025	0,132	0,132	108,33	1
43	GEO CONFIGURAL	12,582	15	1,1856		1	0	0	0,045						0
44	GEO METRIC	16,631	18	1,2064	3	1	0	0,001	0,072	0	0	0,001	0,027	3,93	0
45	GEO SCALAR	32,593	21	1,1904	3	0,996	0,013	0,001	0,089	-0,004	0,013	0,017	0,017	17,12	1
46	HRV CONFIGURAL	27,231	15	1,6615		0,995	0,016	0,001	0,082						0
47	HRV METRIC	30,621	18	1,7732	3	0,995	0,014	0,002	0,086	0	-0,002	0,001	0,004	3,88	0
48	HRV SCALAR	82,108	21	1,7199	3	0,975	0,029	0,003	0,229	-0,02	0,015	0,143	0,143	62,08	1
49	HUN CONFIGURAL	40,808	15	1,1035		0,993	0,023	0,001	0,056						0
50	HUN METRIC	46,312	18	1,1454	3	0,992	0,022	0,002	0,079	-0,001	-0,001	0,001	0,023	5,91	0
51	HUN SCALAR	87,451	21	1,1432	3	0,981	0,031	0,001	0,145	-0,011	0,009	0,066	0,066	41,53	1

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
52	JPN CONFIGURAL	17,975	15	1,2171		0,999	0,007	0,001	0,041						0
53	JPN METRIC	18,042	18	1,3216	3	1	0,001	0,001	0,06	0,001	-0,006	0	0,019	1,07	0
54	JPN SCALAR	30,264	21	1,3389	3	0,997	0,011	0,001	0,083	-0,003	0,01	0,023	0,023	11,56	1
55	KAZ CONFIGURAL	37,593	15	1,5567		0,993	0,015	0,001	0,083						0
56	KAZ METRIC	39,963	18	1,6821	3	0,993	0,014	0,002	0,107	0	-0,001	0,001	0,024	3,77	0
57	KAZ SCALAR	47,1	21	1,6692	3	0,992	0,014	0,002	0,101	-0,001	0	-0,006	-0,006	7,16	0
58	KOR CONFIGURAL	27,394	15	1,3941		0,996	0,017	0,002	0,058						0
59	KOR METRIC	33,148	18	1,318	3	0,995	0,017	0,002	0,087	-0,001	0	0	0,029	5,87	0
60	KOR SCALAR	40,044	21	1,3147	3	0,994	0,018	0,002	0,105	-0,001	0,001	0,018	0,018	6,92	0
61	LTU CONFIGURAL	33,8	15	1,3641		0,993	0,018	0,001	0,09						0
62	LTU METRIC	67,194	18	1,1922	3	0,982	0,027	0,002	0,154	-0,011	0,009	0,001	0,064	102,20	1
63	LTU SCALAR	91,68	21	1,3449	3	0,974	0,03	0,002	0,209	-0,008	0,003	0,055	0,055	19,10	1
64	LVA CONFIGURAL	35,418	15	0,9071		0,99	0,024	0,001	0,077						0
65	LVA METRIC	34,721	18	0,9765	3	0,992	0,02	0,002	0,084	0,002	-0,004	0,001	0,007	1,34	0
66	LVA SCALAR	68,385	21	1,0219	3	0,978	0,031	0,002	0,152	-0,014	0,011	0,068	0,068	27,80	1
67	MEX CONFIGURAL	20,417	15	1,3025		0,997	0,011	0,002	0,055						0
68	MEX METRIC	34,76	18	1,2926	3	0,992	0,018	0,002	0,102	-0,005	0,007	0	0,047	14,75	1
69	MEX SCALAR	82,444	21	1,2532	3	0,97	0,032	0,002	0,202	-0,022	0,014	0,1	0,1	57,42	1

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
70	MLT CONFIGURAL	30,221	15	0,8862		0,989	0,025	0,001	0,093						0
71	MLT METRIC	34,717	18	0,8633	3	0,988	0,024	0,002	0,13	-0,001	-0,001	0,001	0,037	4,26	0
72	MLT SCALAR	41,829	21	0,9358	3	0,985	0,025	0,002	0,192	-0,003	0,001	0,062	0,062	6,69	0
73	NOR CONFIGURAL	20,661	15	1,4464		0,998	0,01	0,001	0,049						0
74	NOR METRIC	29,923	18	1,4878	3	0,995	0,013	0,001	0,119	-0,003	0,003	0	0,07	8,64	1
75	NOR SCALAR	83,452	21	1,5024	3	0,975	0,027	0,001	0,299	-0,02	0,014	0,18	0,18	50,85	1
76	NZL CONFIGURAL	17,775	15	2,5136		0,997	0,009	0,004	0,058						0
77	NZL METRIC	39,142	18	2,4204	3	0,977	0,023	0,013	0,148	-0,02	0,014	0,009	0,09	25,61	1
78	NZL SCALAR	49,984	21	2,5382	3	0,969	0,025	0,011	0,164	-0,008	0,002	0,016	0,016	9,90	1
79	PRT CONFIGURAL	13,036	15	1,1079		1	0	0,001	0,047						0
80	PRT METRIC	13,549	18	1,159	3	1	0	0,001	0,054	0	0	0	0,007	0,89	0
81	PRT SCALAR	72,23	21	1,1014	3	0,98	0,026	0,006	0,126	-0,02	0,026	0,072	0,072	84,48	1
82	RUS CONFIGURAL	17,535	15	1,3441		0,999	0,006	0,001	0,035						0
83	RUS METRIC	20,003	18	1,7011	3	0,999	0,005	0,002	0,08	0	-0,001	0,001	0,045	3,00	0
84	RUS SCALAR	92,768	21	1,7136	3	0,971	0,029	0,001	0,2	-0,028	0,024	0,12	0,12	69,85	1
85	SAU CONFIGURAL	18,534	15	1,1586		0,998	0,009	0,003	0,048						0
86	SAU METRIC	19,822	18	1,27	3	0,999	0,006	0,003	0,075	0,001	-0,003	0	0,027	2,03	0
87	SAU SCALAR	55,117	21	1,2464	3	0,984	0,024	0,004	0,144	-0,015	0,018	0,069	0,069	39,40	1

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06 <0.08 Mplus	<0.01	>-0.015	>-0.03	>-0.03		
88	SVK CONFIGURAL	20,361	15	1,1653		0,997	0,011	0,001	0,048						0
89	SVK METRIC	22,216	18	1,3411	3	0,998	0,009	0,001	0,062	0,001	-0,002	0	0,014	2,73	0
90	SVK SCALAR	82,584	21	1,3404	3	0,968	0,031	0,001	0,263	-0,03	0,022	0,201	0,201	60,55	1
91	SVN CONFIGURAL	28,693	15	0,8989		0,993	0,021	0,001	0,068						0
92	SVN METRIC	33,18	18	0,9594	3	0,993	0,02	0,002	0,104	0	-0,001	0,001	0,036	4,79	0
93	SVN SCALAR	61,79	21	0,9446	3	0,98	0,03	0,002	0,216	-0,013	0,01	0,112	0,112	31,00	1
94	SWE CONFIGURAL	21,863	15	1,52		0,995	0,013	0,002	0,057						0
95	SWE METRIC	23,176	18	1,7264	3	0,996	0,01	0,002	0,09	0,001	-0,003	0	0,033	2,46	0
96	SWE SCALAR	36,662	21	1,9495	3	0,989	0,016	0,002	0,16	-0,007	0,006	0,07	0,07	9,57	1
97	TUR CONFIGURAL	26,931	15	1,7079		0,996	0,014	0,002	0,073						0
98	TUR METRIC	35,541	18	1,8171	3	0,994	0,016	0,002	0,111	-0,002	0,002	0	0,038	7,87	1
99	TUR SCALAR	52,02	21	1,8177	3	0,989	0,019	0,002	0,148	-0,005	0,003	0,037	0,037	16,46	1
100	TWN CONFIGURAL	15,077	15	1,212		1	0,001	0	0,052						0
101	TWN METRIC	17,58	18	1,1855	3	1	0	0,001	0,064	0	-0,001	0,001	0,012	2,44	0
102	TWN SCALAR	68,827	21	1,1062	3	0,986	0,024	0,003	0,176	-0,014	0,024	0,112	0,112	87,71	1
103	USA CONFIGURAL	50,168	15	1,3271		0,959	0,03	0,002	0,077						0
104	USA METRIC	30,408	18	2,7238	3	0,986	0,016	0,002	0,117	0,027	-0,014	0	0,04	1,67	0
105	USA SCALAR	36,141	21	3,1949	3	0,983	0,017	0,002	0,156	-0,003	0,001	0,039	0,039	5,42	0

		Chi2	df	Scaling correction	df Δ	CFI	RMSEA	SRMRw	SRMRb	ΔCFI	ΔRMSEA	ΔSRMRw	ΔSRMRb	Satorra-bentler	Critical value for 3df [7.81] 1= sig.
	TALIS criteria for model evaluation					>0.90	<0.08	<0.06	<0.06	<0.01	>-0.015	>-0.03	>-0.03		
									<0.08 Mplus						
106	VNM CONFIGURAL	27,459	15	1,4449		0,994	0,015	0,001	0,072						0
107	VNM METRIC	32,247	18	1,4618	3	0,993	0,014	0,001	0,095	-0,001	-0,001	0	0,023	4,83	0
108	VNM SCALAR	57,394	21	1,5054	3	0,981	0,021	0,002	0,124	-0,012	0,007	0,029	0,029	22,22	1
109	ZAF CONFIGURAL	18,591	15	1,198		0,998	0,011	0,002	0,048						0
110	ZAF METRIC	27,958	18	1,3121	3	0,995	0,016	0,003	0,106	-0,003	0,005	0,001	0,058	7,66	0
111	ZAF SCALAR	51,805	21	1,3097	3	0,984	0,027	0,003	0,188	-0,011	0,011	0,082	0,082	24,06	1

Article 3

Veletić, J., & Olsen, R. V. (2021). Exploring school leadership profiles across the world: A cluster analysis approach to TALIS 2018. *International Journal of Leadership in Education*. Scopus. https://doi.org/10.1080/13603124.2021.1953612.

Status: Published

III

INTERNATIONAL JOURNAL OF LEADERSHIP IN EDUCATION
https://doi.org/10.1080/13603124.2021.1953612

Routledge
Taylor & Francis Group

Exploring school leadership profiles across the world: a cluster analysis approach to TALIS 2018

Jelena Veletić and Rolf Vegar Olsen

Center for Educational Measurement, Faculty of Educational Sciences, University of Oslo, Norway

ABSTRACT

The purpose of this study is to examine leadership for learning practices across the world by establishing profiles of leadership at school and country levels. Consequently, the study brings to our attention the (ir)relevance of school and system features for leadership for learning. The paper contributes to the field through the use of an extensive exploratory approach across a varied set of school leadership measures collected from both teachers and principals and contextualized in 42 different educational systems participating in the Teaching and Learning International Survey (TALIS) 2018. Consequently, this work has the potential to generate hypotheses regarding the understanding of the complex nature of school leadership worldwide. Surprisingly, the findings reveal that clusters at the country level primarily do not reflect countries with geographical, linguistic, or political proximity. Such clusters were expected, given the evidence found in the literature that shows leadership to largely be determined by contextual, societal, and cultural values. Nevertheless, the analysis identifies five profiles of leadership across schools, the majority of which can be found in most countries participating in TALIS.

Introduction

Most of the studies in the area of school leadership are conducted within individual educational systems or larger geographical areas that are characterized by some shared features (e.g. Asia, U.S.), resulting in only a few international comparative studies in this field (Herborn et al., 2017; Mango, 2018). This likely indicates that school leadership differs as a function of cultural dimensions and other contextual features (Brewer et al., 2020; Hallinger, 2018; Jacobson & Johnson, 2011; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2016). The claim that leadership practices are embedded in culturally sensitive values and worldviews is also supported by findings from other disciplines that are concerned with leadership, such as management as well as occupational and organizational psychology (House et al., 2004).

Sensitivity to how school leadership is culturally embedded and contextually dependent is crucial in order to improve teaching and learning in schools (Knapp et al., 2014; Slater & Teddlie, 1992). Successful leadership, in practice, frequently

CONTACT Jelena Veletić jelena.veletic@cemo.uio.no Center for Educational Measurement, University of Oslo, Oslo, Norway

implies the integration of different leadership styles (Boyce & Bowers, 2018; Leithwood et al., 2008; Marks & Printy, 2003). Thus, leadership theories and models have been developing and adjusting to societal changes (Crow, 2006), consequently blurring clear boundaries between previously well-established leadership models (Brauckmann & Pashiardis, 2011). The relevant example referred to in this paper is the recent Leadership for Learning model, which integrates several precedent leadership frameworks – instructional leadership, transformational leadership, and distributed leadership (Hallinger, 2011; Murphy et al., 2007). The model is focused on learning at all levels and describes eight dimensions that encompass not only instruction and assessment but also organizational culture and social advocacy (Daniëls et al., 2019).

On the one hand, much of school leadership research focuses on associations between school leadership and other characteristics of schools (Krüger et al., 2007), teachers and classrooms (Boyce & Bowers, 2018; Printy, 2008; Tan et al., 2020), or students (Leithwood et al., 2010; Robinson et al., 2008). On the other hand, a person-centered approach to leadership is rarely employed. For instance, Urick and Bowers (2014) examined different types of principals but only in the U.S. context. In addition, pioneering work on leadership typologies around the world using TALIS data has recently been conducted, where at least three different profiles were identified at school and teacher levels through latent class analysis (Bowers, 2020).

The present study aims to provide a descriptive summary of leadership for learning measures that originate from both teachers and principals, scrutinizing them jointly in a single analysis. Accordingly, we recognize that leadership for learning is achieved through joint endeavors of various school stakeholders, mainly teachers and principals. We first employ a series of descriptive/exploratory analyses in order to assess whether the data at the country level are appropriate for subsequent analytical steps. Since the variation at the country level is found to be rather low, applying to only four variables, we focus our analysis on the school level. To identify groups of schools with unique leadership profiles, we employ K-means clustering (Everitt et al., 2011). Thus, we identify school characteristics that account for similarities/differences between clusters. By identifying clusters of schools (patterns of similarity within clusters and patterns of dissimilarity across clusters) that are summarized at the country level (percentages of schools belonging to the same cluster within a country), the presented analysis describes the unique and robust features of leadership at the system level. Seen together, inferences at the country level enable us to examine the heterogeneity of leadership practices across educational systems.

To properly account for a holistic Leadership for Learning model, this study uses data collected from the teachers and principals who participated in the most recent cycle of the Teaching and Learning International Survey (TALIS 2018). Similar studies that take a global comparative perspective on educational leadership are still limited, and our results will help improve understanding of how teachers and principals report about the broader characteristics of practices in schools. Moreover, these practices are reasonably assumed to reflect various theoretical dimensions of the Leadership for Learning framework. As such, our work aims to provide a basis for generating more targeted hypotheses for future research answering the following research questions:

(1) To what extent can *countries* across the world be classified into groups based on leadership for learning practice as reported by teachers and principals?
(2) To what extent can *schools* across the world be classified into groups based on leadership for learning practice as reported by teachers and principals?
(3) How is group membership of schools associated to demographic characteristics of schools and principals?

Theoretical background

From instructional leadership to leadership for learning

Improving school leadership by focusing on learning, monitoring teaching, building safe and effective learning environments, supporting teacher collaboration, acquiring and allocating resources, has been a promising approach employed in the overarching endeavor to improve education in general (Blitz & Modeste, 2015). Improving school leadership imposes tremendous demands on school leaders, further resulting in leadership practices in which functions and responsibilities are, to a large extent, distributed within (school management teams, teachers) and outside (collaboration with other schools and local community) of schools (Pont et al., 2008). Historically, the model of instructional leadership was considered to be of great significance for the improvement of teaching and learning for all by relating leadership to the larger educational agenda (Hallinger, 2005, 2009; Robinson et al., 2008). A core feature of instructional leadership has been the improvement of instruction and learning through the principal's direct engagement (Bossert et al., 1982; Hallinger & Murphy, 1985). Instructional leadership frameworks first emerged in the USA in the 1950s and has been a dominant construct of leadership grounded in practice (Hallinger, 2015). Principals were widely invited to become instructional leaders, which implied not only directly engaging with instruction as implemented in the classrooms, but also a focus on managerial, human resources, political, and institutional functions. This is frequently perceived as an unattainable ideal. Thus, this form of leadership had less and less sense and support in practice (Leithwood et al., 2012) and attention gradually shifted toward a shared instructional leadership perspective (Harris, 2004; Marks & Printy, 2003). Such perspectives were brought to the foreground by leadership frameworks such as distributed leadership (Gronn, 2002; Harris, 2009; Spillane et al., 2004) and transformational leadership (Day et al., 2016; Hallinger, 2003; Leithwood & Jantzi, 2006).

Leadership for learning frameworks appeared in the literature in the early 2000s. One group of authors, mainly coming from the U.S., used this term as a synonym for instructional leadership with some more detailed and broader description of what leadership practice entails while still keeping school improvement and effectiveness as a central objective (Hallinger, 2009; Murphy et al., 2007). Another group of authors, mainly from the UK, developed a leadership for learning framework characterized by different underlying assumptions and objectives (MacBeath, 2019), In common with instructional leadership this framework maintained a focus on learning, yet through a more collaborative perspective taking into account a wider range of leadership sources and broadening learning as something not only including the students, but the school as a whole (Townsend, 2019). Both conceptualizations of leadership were central in

educational reforms that took place worldwide in the early 2000s with an increased emphasis on accountability. MacBeath (2019) emphasizes the importance of terminology by explaining that 'instruction' place teacher, parent, or authority figure at the central stage, while 'learning' puts an emphasis on what learners do and how learning is made manifest. Thus, learning and leading are understood more as activities and not as roles, in which emotional and human aspects are emphasized. Thus, leadership for learning compared to instructional leadership emphasizes 1) capacity building of teachers and staff, 2) greater reliance on multiple forms of teacher leadership and teacher collaboration, as well as 3) more attention to school as a learning organization for all, not only students. Leadership for learning is more responsive to students, embraces a moral purpose of education, connects with agents outside of school, and neglects hierarchy (Dempster, 2019; Imig et al., 2019).

An operationalization of Leadership for Learning has been even more challenging and scholars did not agree on a single model to date. Following Bowers (2020), we acknowledge various attempts to describe leadership for learning domains that are, to a great extent, congruent with one another but differ in how broadly they capture leadership practices (Boyce & Bowers, 2018; Hallinger, 2011; Halverson et al., 2014; Murphy et al., 2007). The first such attempt was a model described by Murphy (2007), followed by subsequent models that share the same fundamental concepts. As a result, in this study, we focus on the model proposed by Murphy et al. (2007).

This model consists of eight major leadership for learning dimensions, which are further defined by several domains. These dimensions are: *vision for learning, instructional program, curricular program, assessment program, communities of learning, resource acquisition and usage, organizational culture*, and *social advocacy* (Murphy et al., 2007). *Vision for learning* implies that a great deal of time is dedicated to the development, articulation, implementation, and stewardship of ambitious goals that are focused on learning and achievement and are easily interpretable and measurable. *Instructional program* refers to the involvement in instruction and teaching, staff support, and protection of instructional time. By establishing high standards and expectations and by coordinating curriculum materials and assessments, the *curricular program* dimension is covered. Similarly, the *assessment program* dimension is covered through the crafting, implementing, and monitoring of assessments at classroom and school levels. Professional development, a culture of collaboration, and fairness are emphasized through the *learning communities* dimension. The *resource acquisition and usage* dimension is oriented toward locating and securing additional resources for schools from the broader school community using both formal and informal channels. Resource deployment and use should clearly be linked to school mission and goals. Continuous focus on school development and on a safe and orderly learning environment, as well as an emphasis on personal and group achievements and recognition, are features of the *organizational culture* dimension. Finally, the *social advocacy* dimension covers four domains – environmental context, diversity, ethics, and stakeholder involvement. The presented model is nicely described by MacBeath and Townsend (2011), who state that leadership for learning embraces much more than the improvement of student learning outcomes only – but that it is also concerned with teacher and leadership learning, creating a climate of creativity and growth by drawing attention to the dynamic connections, relationships, and mutual influences that shape both learning and teaching.

School leadership in the TALIS 2018 study

School leadership remains a top priority, according to the country ratings of the themes for inclusion in the TALIS 2018 study. The increasing interest in school leadership is recognized by the TALIS 2018 study, where richer measures for school environment can be found in both the school and teacher questionnaires (Ainley & Carstens, 2018). Since the first study, which was conducted in 2008, the thematic coverage of the subsequent TALIS surveys has changed in order to also reflect the recent trends and innovations in research on school leadership. As previously discussed, school leadership research has shifted its focus toward more distributed practices, involving stakeholders across all levels of the educational system. TALIS acknowledges the developments in the field by keeping the principal as a central character but also including a more collaborative perspective on leadership (Ainley & Carstens, 2018). Consequently, in addition to instructional leadership, which remains a main interest, two additional leadership conceptualizations are discussed in the TALIS 2018 conceptual framework: teacher leadership, where teachers take on leadership roles both within and outside of the classroom (Muijs & Harris, 2003), and system leadership, where principals take on leadership roles outside of the school. The latter brings the importance of the relation with the broader community to our attention (Ainley & Carstens, 2018; Schley & Schratz, 2011).

School leadership, as described by the TALIS 2018 conceptual framework, nicely encompasses all three important features of leadership for learning: the principal remains the central character (instructional leadership), the perspective on leadership is more collaborative and distributed (distributed leadership), and the broader social and system features are accounted for (system leadership).

The actual scales available in TALIS 2018 questionnaires, which directly deal with leadership, are the scales of *school leadership* and *participation among stakeholders* from the principal questionnaire and *participation among stakeholders* from the teacher questionnaire. However, other scales that are not exclusively described as school leadership scales can also be used to describe the school environment and the working conditions that are closely related to leadership for learning (e.g. academic pressure, team innovativeness, stakeholder involvement, and more). Conceptual mapping of the Leadership for Learning theory and the TALIS 2018 *items* from both teacher and principal questionnaires was performed by Bowers in a working paper about leadership typologies using the same data (Bowers, 2020).

Table 1 provides a broad conceptual map of how the TALIS 2018 *scales* correspond with the dimensions of the Leadership for Learning framework. The left-hand side of the table lists the eight theoretically-defined dimensions from the Leadership for Learning framework, while the right-hand side identifies the TALIS 2018 scales that partly reflect certain aspects of these dimensions. It should be noted that some scales from the TALIS 2018 study are identified as reflecting more than one theoretical dimension and that one of the dimensions in the theoretical framework does not have a corresponding scale in the empirical data. Hence, Table 1 provides a condensed picture of how TALIS 2018 represents a broad and coarse-grained operationalization of the core features of the Leadership for Learning framework. However, the table also serves to illustrate the fact that the match between the Leadership for Learning theory and TALIS is only a partial one. Accordingly, TALIS is a rather blunt – but nevertheless useful – instrument that can

Table 1. Conceptual mapping of the Leadership for Learning framework and TALIS 2018 scales.

Leadership for learning framework	TALIS 2018 scales[a]
Vision for learning	Academic pressure; Team innovativeness.
Instructional program	School leadership; Instructional autonomy.
Curricular program	Academic pressure; Instructional autonomy.
Assessment program	Instructional autonomy .
Communities of learning	Participation among stakeholders – principals; Participation among stakeholders – teachers; Teacher cooperation; Effective professional development.
Resource acquisition and use	
Organizational culture	Team innovativeness; Organizational innovativeness; Exchange and cooperation among teachers.
Social advocacy	Stakeholder involvement – partnership; Diversity beliefs; Participation among stakeholders.

[a]More about scales can be found in Table 3 in this paper.

be used to map out the characteristics of school practices that reasonably reflect the Leadership for Learning framework.

National and school contexts and their relevance for school leadership

It is likely not very useful to conceptualize school leadership as a universal phenomenon, independent of school context, educational policy, culture, national history, and values. Although weakly supported in the quantitative literature, the argument about the importance of cultural and national contexts for school leadership practices is widely accepted among practitioners and scholars (Clarke & O'Donoghue, 2016; Harris, 2020; Johnson et al., 2008). This claim is further supported by examples from practice in which, for instance, successful leaders in one environment did not necessarily succeed as leaders in another (Miller, 2018). Finally, research describes cases that illustrate how attempts to transfer educational policies for school governance and leadership from one educational system to another were unsuccessful (Harris, 2020; Hooge, 2020; Oplatka & Arar, 2017). Studies on how divergent national educational policies directly shape school leadership practices provide further evidence of cross-cultural differences (Hooge, 2020; Miller, 2018). Møller and Schratz (2009) expand the argument further to the socio-cultural, historical, and political contexts by discussing the differences, similarities, and conditions in four different regions – England, Scandinavia, German-speaking countries, and Eastern European countries. They conclude that leadership is culturally embedded and socially constructed and that the difference is even greater when countries do not share linguistic and common cultural heritage. However, empirical evidence about the importance of system features for leadership practices is still limited. Therefore, the current study applies quantitative analysis to system-level representative data in order to answer what has, over the years, primarily been supported by evidence from case studies and literature reviews.

Moreover, strong evidence exists for the importance of culture for leadership practice at the micro (school) level. Values, norms and traditions that shape organizational culture within schools are found to be strongly associated with school leadership practice (DuPont, 2009; Hallinger & Leithwood, 1998; Kalkan et al., 2020; Karada & Öztekin, 2018; Liu et al., 2021; Sahin, 2011). Together with the concept of school climate which

refers to shared perceptions and behaviors (Ashforth, 1985; Hoy, 1990; Wang & Degol, 2016), school culture might be one of the closest and tightly related factors that could explain possible variation in leadership practice across schools. When it comes to specific school contexts – such as school and principal demographic variables – the literature is generally inconsistent (Hallinger, 2005, 2008; Opdenakker & Damme, 2007). In a review paper on this matter, with a focus on instructional leadership, Hallinger (2008) concluded that school size, school performance rating, private schools, and level of the principal did not significantly account for differences in approaches to leadership, while gender and the number of years of experience of the principal were more frequently found to be significantly related to how instructional leadership is implemented. We were unable to identify similar studies that specifically refer to leadership for learning and therefore we examine to what extent 1) school demographics (such as school size, location, private/public, and number of students whose first language differs from the language of instruction) and 2) principal demographics (such as gender) are relevant for leadership.

Partially, the current study also responds to the criticism that much of the theoretical work on school leadership is derived from Western countries, predominantly from the U.S. This criticism calls for more studies to incorporate varied and globally relevant cultural, institutional, and economic settings (Hallinger, 2018; Oplatka & Arar, 2017; Walker & Dimmock, 2002). for learning practices.

Methods

Data and sample

This study used data from the third and currently last cycle of the TALIS study – the TALIS 2018 survey. TALIS is an international, large-scale survey that is concerned with teaching and learning conditions, learning environments, school leadership, and more (Ainley & Carstens, 2018). In TALIS 2018, 48 countries and economies took part in the core survey – that is, teachers and principals from the lower secondary level of education (ISCED Level 2).[1] TALIS applied a two-stage sampling design. Within a country, a random sample of 200 schools was identified and invited to participate in the study during the first stage, followed by drawing a random sample of 20 teachers from each of the selected schools. More details about the sampling design and outcomes can be found in the TALIS technical report (Organisation for Economic Co-operation and Development [OECD], 2018).

The total sample in this study included 7,427 schools from 43 countries.[2] Six countries were excluded from the analysis due to a large amount of systematically missing data (10.25% of the total sample). Table 2 provides an overview of the excluded countries as well as the reasons for their exclusion. For the remaining sample, listwise deletion was utilized. The effect of the listwise deletion varied between countries (from 0.68% to 35.94%) with an average of 11.86% of missing data per country.[3] A large portion of missing data resulted from all data missing for the school level (17.94% of data missing from the total data after county exclusion) or all data missing for the teacher level (20.72% of data missing from the total data after country exclusion).

Table 2. Countries excluded due to missing data.

Country	Reason for exclusion	Number of schools
Argentina	All data missing on variables 'Establishing student assessment policies' and 'Approving students for admission' – not administered.	134 schools
Iceland	Does not have data available at the school level.	122 schools
Japan	Does not have response option 'School governing board' on variables TC3G20I:TC3G20F.	196 schools
Sweden	All data missing on T3PCOM (stakeholders involvement) due to scale non-convergence.	183 schools
Denmark	All data missing on T3PCOM (stakeholders involvement) due to scale non-convergence.	148 schools
Hungary	All data on T3EFFPD (effective professional development) was withdrawn at Hungary's request because wording was not sufficiently clear for items TQ24–26.	189 schools

Measures

Six scales from the principal questionnaire and six scales from the teacher questionnaire were used in the analysis (see Table 3). The scales represent the factor scores calculated in the Confirmatory Factor Analysis (CFA) framework and already available as part of the TALIS database (OECD, 2018). TALIS conducted two types of CFA: 1) joint analysis of data from all participating countries and 2) separate analyses for each country's economy. The final scale modeling accounted for invariance levels across countries and ISCED levels. Hence, the final scales were modeled using the multigroup CFA (MGCFA) framework from which factor scores are obtained.

One of the scales originating from the principal questionnaire, Instructional autonomy in schools, is not included in the publicly available dataset. We derived this scale for the purpose of this study, by following procedures very similar to how are the other scales in TALIS produced. A set of six items for which principals responded in relation to the question of who has a significant responsibility for the following tasks: choosing which learning materials are used, deciding which courses are offered, determining the course content, approving students for admission to the school, establishing student assessment policies, and establishing student disciplinary policies. These items were first recoded into an ordinal scale with three levels: full autonomy (if internal evaluators, such as the principal, other members of the school management team, teachers, or the school governing board, were checked), mixed autonomy (if both internal and external evaluators are checked), and no autonomy (if only external evaluators were checked). Then, these six variables with three levels were used to calculate the unique factor score that represents the school autonomy for instructional practices variable.[4]

The final school file consisted of scores for twelve scales originating from both the principal and teacher questionnaires. Prior to the cluster analysis, all scales were standardized with an international mean of zero and a standard deviation of one, using so-called senate weights to ensure that all countries contribute equally. Subsequently, teacher variables were aggregated to the school level. As shown in the last column in Table 3, the scales used in this analysis achieved a different level of measurement invariance. Measurement invariance (MI) refers to the property that an instrument should function equally across a range of conditions regarded to be irrelevant to the attribute being measured (e.g. language, culture, item understanding) (Millsap, 2012). If this condition is not satisfied, then there is no sound basis for a comparison of (latent) mean scores across groups. Establishing the MI across such a large number of groups has been challenging and, as some authors argue, may represent an unrealistic goal (Byrne &

Table 3. TALIS 2018 list of scales used in the analysis.

Scale	# of items	Item stimulus	Item example	Response categories	MI*
Principal questionnaire					
T3PACAD (Academic pressure)	4	To what extent do the following statements apply to this school?	Teachers understand the school's curricular goals.	4-point scale (1 = *Not at all* to 4 = *A lot*)	M
T3PCOM (Stakeholder involvement, partnership)	3	To what extent do the following statements apply to this school?	Parents or guardians support student achievement.	4-point scale (1 = *Not at all* to 4 = *A lot*)	M
T3PLEADS (School leadership)	3	Please indicate how frequently you engaged in the following activities in this school during the last 12 months.	I took actions to support co-operation among teachers to develop new teaching practices.	4-point scale (1 = *Never or rarely* to 4 = *Very often*)	M
T3PLEADP (Participation among stakeholders, principals)	5	How strongly do you agree or disagree with these statements as applied to this school?	This school provides staff with opportunities to actively participate in school decisions.	4-point scale (1 = *Strongly disagree* to 4 = *Strongly agree*)	M
T3PORGIN (Organizational innovativeness)	4	How strongly do you agree or disagree with the following statements?	This school quickly identifies the need to do things differently.	4-point scale (1 = *Strongly disagree* to 4 = *Strongly agree*)	C
AUTONOMY** (Instructional autonomy in schools)	6	Level of autonomy regarding the following.	Choosing which learning materials are used.	3-point scale (1 = *No autonomy* to 3 = *full autonomy*)	
Teacher questionnaire					
T3COLES (Professional collaboration in lessons among teachers)	4	On average, how often do you do the following in this school?	Teach jointly as a team in the same class.	6-point scale (1 = *Never* to 6 = *Once a week or more*)	M
T3STAKE (Participation among stakeholders, teachers)	5	How strongly do you agree or disagree with these statements as applied to this school?	This school provides staff with opportunities to actively participate in school decisions.	4-point scale (1 = *Strongly disagree* to 4 = *Strongly agree*)	M
T3TEAM (Team innovativeness)	4	Thinking about the teachers in this school, how strongly do you agree or disagree with the following statements?	Most teachers in this school strive to develop new ideas for teaching and learning.	4-point scale (1 = *Strongly disagree* to 4 = *Strongly agree*)	S
T3EFFPD (Effective professional development)	3	Thinking of the prof. develop. activity that had the greatest positive impact on your teaching during the last 12 months, did it have any of the following characteristics?	It built on my prior knowledge.	Binary choice (1 = *Yes* and 2 = *No*)	C
T3EXCH (Exchange and cooperation among teachers)	4	On average, how often do you do the following in this school? Response options.	Exchange or develop teaching materials with colleagues.	6-point scale (1 = *Never* to 6 = *Once a week or more*)	C
T3DIVP (Diversity practices, teacher)	4	In this school, are the following practices in relation to diversity implemented?	Adopting teaching and learning practices that integrate global issues throughout the curriculum.	Binary choice (1 = *Yes* and 2 = *No*)	C

*Level of measurement invariance that the scale achieved C = Configural; M = Metric; S = Scalar.
** Factor score calculated for this study.

Vijver, 2010; Lubke & Muthén, 2004; Rutkowski & Svetina, 2014; Zieger et al., 2019). In the present study, we did not explicitly compare scales at the country level but, instead, the profiles of leadership for learning at the school level, which constitute a mitigating factor for the inclusion of scales that achieve a different MI level. Although we examined more than 40 educational systems, we also acknowledge the potential shortcomings that could result from scales that only achieve a configural level of invariance.

Statistical analysis

Data were first prepared using the IDB Analyzer 4.0, while the main analyses were conducted using the R studio (R Core Team, 2018; IMB Corp, 2017). The R package *cluster* (Maechler et al., 2019) was used for the cluster analyses, while the package *factoextra* was applied to extracting and visualizing the results (Kassambara & Mundt, 2020). Cluster analysis was regarded to be an appropriate method to use here, given the nature of the problem and the data. Furthermore, the fact that no prior hypothesis about the number and nature of the expected clusters could be reasonably stated suggested that a descriptive and exploratory approach is reasonable. Cluster analysis is a common label attached to a group of statistical techniques and it enables similar observations found in a dataset to be classified or grouped together. Simply stated, objects in the same cluster are similar to one other in relation to a set of characteristics, while objects in different clusters are dissimilar in terms of the same set of characteristics (Everitt et al., 2011). The starting point of cluster analysis is the calculation of a proximity/dissimilarity matrix, consisting of measures identifying the degree of similarity between objects. The choice of the proximity measure depends on the nature and scale of the data (Everitt et al., 2011). In this study, we wanted to identify clusters of schools with similar profiles across a set of leadership for learning characteristics. Accordingly, the Pearson correlation coefficient would be a proximity measure of choice. However, when data are standardized the results obtained from two proximity measures (Pearson correlation and Euclidian distance) are comparable. Thus, we used Euclidian distance as default measure with k-means (Kassambara, 2017). The procedure of clustering was similar to factor analysis but with two main differences: 1) cluster analysis groups *objects* based on the proximity of pairs or larger groups of objects (in our case, schools), while 2) factor analysis groups *variables* based on patterns of variations. Specifically, we used k-means clustering. In k-means clustering, each cluster is represented by its center (i.e. centroid), which corresponds to the mean profile of the objects assigned to the cluster. The main idea is that the total intra-cluster variation is minimized. This method clusters given data into a set of k groups, where k is the number of groups pre-specified by a researcher. Since we had no prior hypothesis about the number of clusters, the optimal number was selected based on 1) the elbow method that utilizes the total within-cluster sum of squares variation as a function of the number of clusters and on 2) the silhouette method that computes the average silhouette coefficient (sometimes referred to as silhouette width) for different values of k (Kassambara, 2017). In addition, a silhouette coefficient was used to validate the clustering solution – i.e. how well each object (in our case, school) is classified to the belonging cluster. Hence, each school was assigned a value that is referred to as the silhouette coefficient (*Si*).[5] A silhouette coefficient can take a value from – 1 to 1. A silhouette coefficient near 1 indicates that observation is well clustered in the belonging

cluster and is far away from other clusters. A negative silhouette coefficient might indicate that observation is also proximal to other clusters, and as such negative values identify objects which are not that well captured by the clustering solution. In evaluating the number of clusters, an optimum is found for the number of clusters resulting in the lowest average silhouette coefficient.

Results

Descriptive statistics

As a first step, we inspected the descriptive statistics for the variables included in the analysis. We ran unconditional three-level (teacher data) and two-level (principal data) models in order to scrutinize variance decomposition across levels – teachers in schools in countries (Snijders & Bosker, 1999). The intra-class correlation coefficients from the unconditional models for each variable are presented in Table 4. Given that the primary concern of this analysis, as stated in RQ1, was to explore phenomena that represent features of schools across countries, the expectation was that a meaningful variability in the data can be accounted for by countries. However, at the country level, only a few variables (school autonomy for instructional practices, effective professional development, team innovativeness, school leadership) showed significant and substantially meaningful variability across countries. However, a significant and substantially meaningful amount of variability was found for most teacher variables at the school level. This finding suggests that most of the variability in the variables of interest lies between schools rather than between countries. Consequently, the originally intended country-level cluster analysis was dropped. Instead, cluster analysis was conducted with schools as primary units.

Figure 1 visualizes and presents additional details about the (lack of) variability between countries through box-plots that describe the dispersion in the country mean scores for all variables. The relatively large intraclass correlations (ICCs) for the 'T3TEAM' (Team innovativeness) and 'AUTONOMY' (Instructional autonomy in schools) variables are indicated by wide boxes (representing the values for the 25th and

Table 4. Variance decomposition at teacher, school, and country levels presented by intraclass correlations (ICC).

Variable	Name	ICC school level	ICC country level
Three-level (teachers in schools in countries)			
T3STAKE	Participation among stakeholders	.141	.000
T3COLES	Professional collaboration in lessons among teachers	.122	.002
T3DIVP	Diversity practices, teachers	.104	.004
T3TEAM	Team innovativeness	.092	.060
T3EXCH	Exchange and cooperation among teachers	.080	.001
T3EFFPD	Effective professional development	.019	.213
Two-level (schools in countries)			
AUTONOMY	Instructional autonomy in schools		.357
T3PLEADS	School leadership		.049
T3PACAD	Academic pressure		.001
T3PCOM	Stakeholders involvement, partnership		.000
T3PLEADP	Participation among stakeholders, principals		*
T3PORGIN	Organizational innovativeness		*

*Model did not converge.

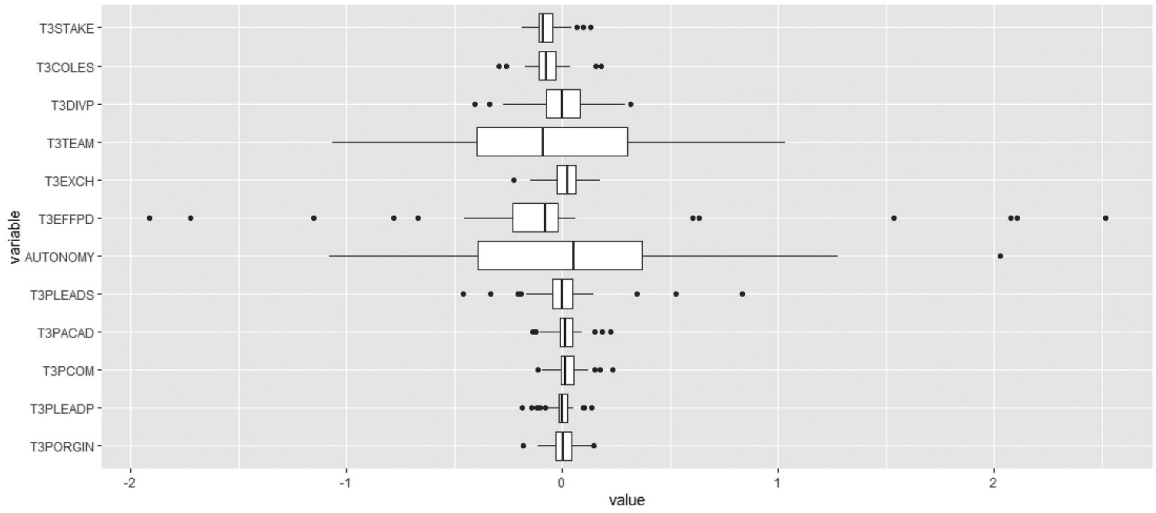


Figure 1. Box-plots showing the variability across countries (country averages).

75th percentiles). It is interesting to note that the high ICC for the 'T3EFFPD' (Effective professional development) variable relates to dispersion, where only a few countries have either extremely low or high scores. The box-plot also shows that the Organizational innovativeness ('T3PORGIN') and Participation among stakeholders, principals ('T3PLEADP') variables have close to zero variability at the country level, which is reassuring, given the fact that null models did not converge for these variables.

K-means clustering of schools: the five cluster solution

All variables from Table 3 are used as indicators of leadership for learning at the school level. First, in order to determine the optimal number of clusters, we used the Elbow method and the Silhouette method, which are illustrated in Figure 2. Both approaches suggested five clusters to be the optimal solution. The aim was to establish clusters that are distinct but also meaningful and interpretable. A manageable number of reasonably-sized clusters was also considered to be an advantage.

In response to the second RQ, we established five different leaderships for learning profiles across schools in 43 educational systems. Figure 3 presents the average values for each of the variables included in the analysis for each of the five clusters. In the following, we refer to these figures as cluster profiles. A first observation is that Cluster 1 and Cluster 4 have profiles that largely go in opposite directions. Cluster 1 is characterized by moderately low values for all variables, except for the school autonomy for instructional practices variable, while Cluster 4 has moderately high values for almost all variables, except for the professional collaboration in lessons among teachers variable. Cluster 3 and 5 profiles also represent mirror images. Cluster 3 is characterized by moderately low negative values for almost all variables reported by principals and by low positive values for most of the measures based on teacher reports. This is also the largest cluster, accommodating 28% of all schools in the sample. In contrast, Cluster 5 has moderately positive values for variables reported by principals and negative or neutral values for variables reported by teachers. It is interesting to note that in both Cluster 3 and Cluster 5, the reports about distributed leadership practice (i.e. participation among

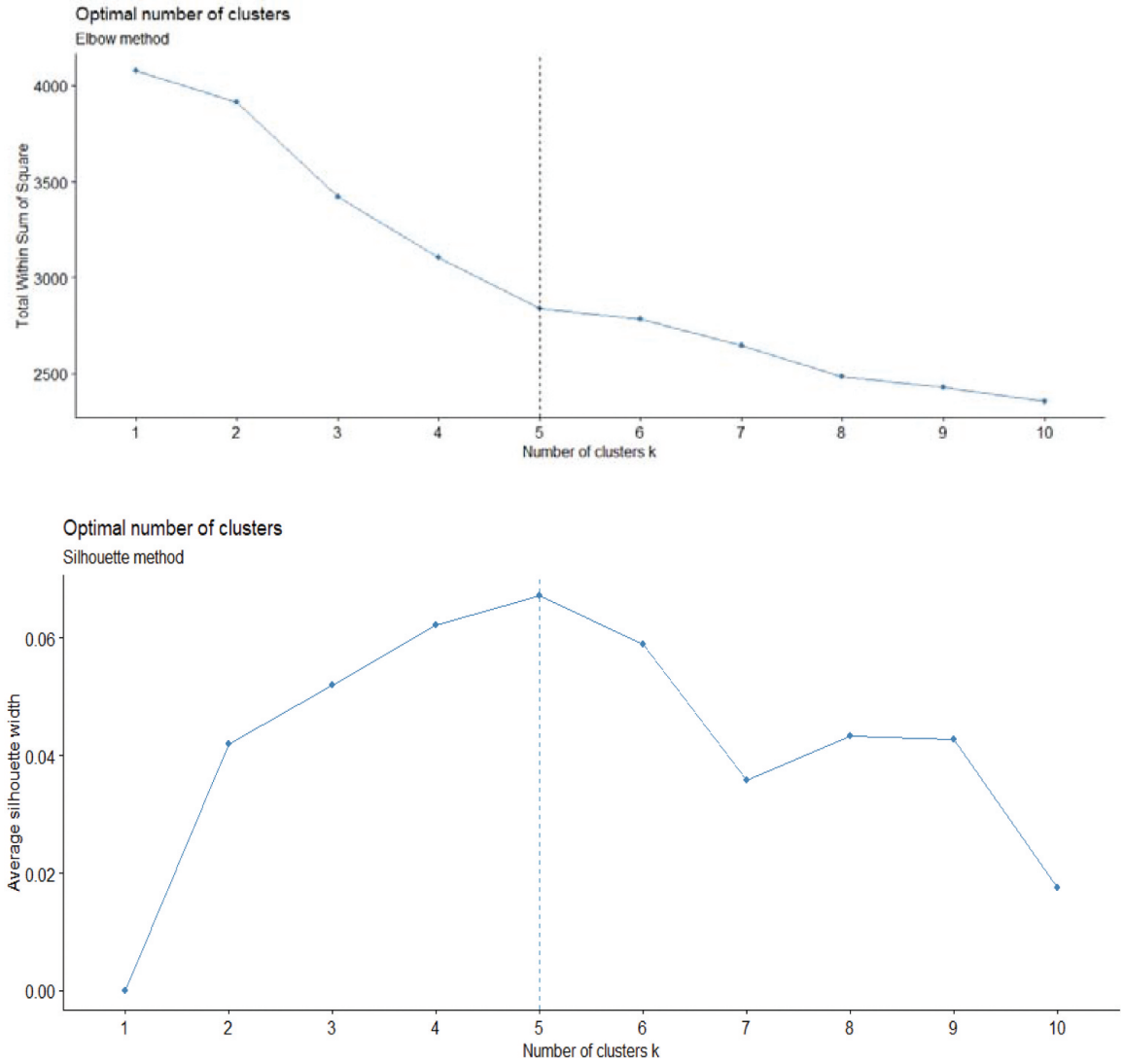


Figure 2. Graphs showing the optimal number of clusters.

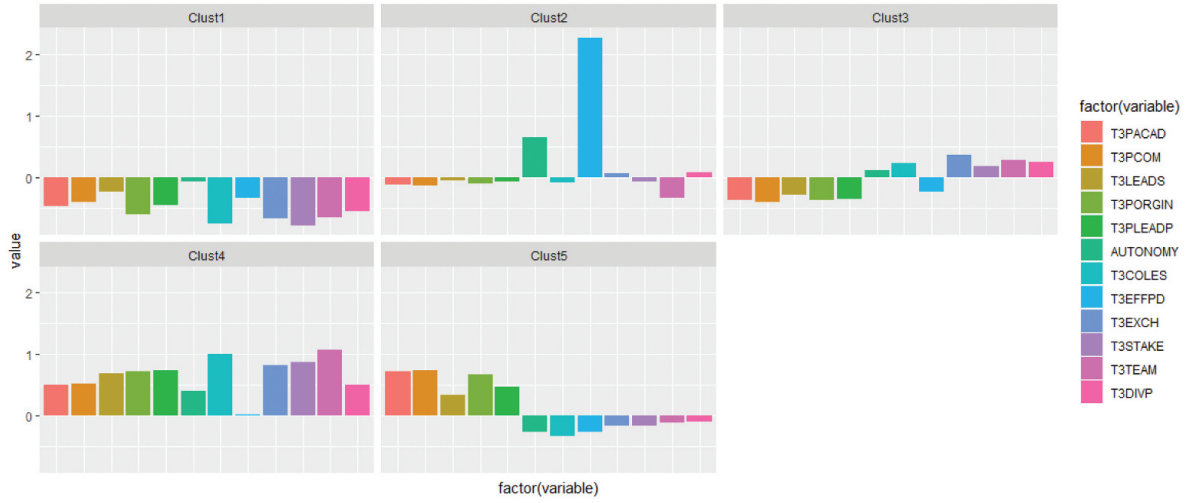


Figure 3. Cluster profiles.

stakeholders) differ between teachers and principals. In cluster 3, principals reported lower distributed leadership ('T3PLEADP') while teachers at the same time reported somewhat higher distributed leadership ('T3STAKE'). The opposite holds for Cluster 5. Thus, a potential gap in how leadership practice is perceived exists between teachers and principals. Note also that the school autonomy for instructional practices variable, reported by principals, moves together with the variables reported by teachers. Cluster 2 has a unique profile characterized by high values for the effective professional development and higher school autonomy for instructional practices variables. This cluster is the smallest in size, comprising 11% of all schools from the sample.

Cluster validation

Figure 4 shows the silhouette coefficients (silhouette widths) for each school as bars stacked next to one another, sorted from the school with the highest width to the left, for each of the five clusters, respectively. The red dotted line represents the average silhouette coefficient across the entire sample. Inspecting Figure 4 reveals that none of the schools in Cluster 3, the largest cluster, have negative widths. In total, 790 schools of the entire 7,427-school sample were identified with negative silhouette coefficients – 15% of the 1,769 schools in Cluster 1, 7% of the 833 schools in Cluster 2, 33% of the 993 schools in Cluster 4, and 7% of the 1,744 schools in Cluster 5. Across countries, NOR had the most schools with a negative silhouette coefficient, amounting to 19% of all its schools, followed by AUT, SAU, GEO, and LVA (see Appendix A for complete list and definitions of the abbreviations presented here) with more than 15% of schools having uncertain cluster membership. In both absolute and relative numbers, most of the uncertainties in classifications relate to Clusters 4 (all high) and 1 (all low). Although the five-clusters solution does not provide a perfect representation of the schools' leadership profiles, the vast majority of schools can reasonably well be categorized into one of the five suggested clusters. The clustering accounts for 25.9% of the variability in school profiles. For further analyses, we excluded schools with negative silhouette coefficients. By doing this, we purified clusters and relied on well-clustered schools that represent typical schools that belong to specific clusters. In conclusion, even when such classification does not work as an informative diagnostic for every single school, the presented cluster solution provides a useful birds-eye view of school leadership practices across the countries that participate

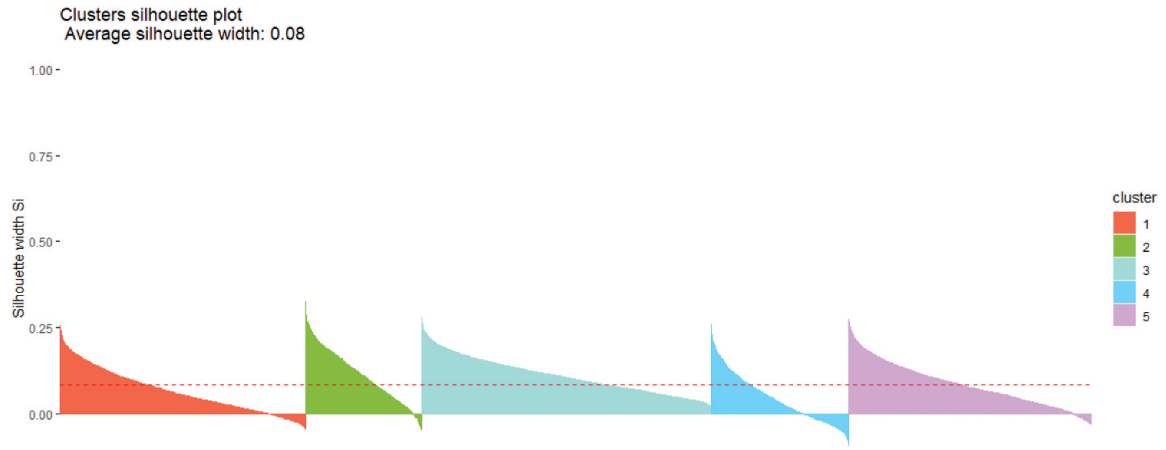


Figure 4. Clusters silhouette plot.

in TALIS. Again, it should be of interest to note that the variability in leadership practices, as reported by principals and teachers, is much larger within than it is across countries.

Dominant clusters within countries

Although we were unable to directly explore clusters of countries based on leadership for learning practice due to low variability at the country level, in response to RQ1 we provided the frequencies of clusters established at the school level within each country (see Appendix B). Table 5 represents a list of countries in which more than 30%[6] of schools belong to respective clusters. Countries listed in italic font have more than 30% their schools in more than one cluster and, therefore, appear in more than one column.

Note that the majority of countries have more than 30% of their schools in Cluster 3, which is the largest cluster. It is interesting to note that countries with more than 30% of schools in at least two clusters are primarily found in two contrasting clusters – Cluster 3 (low on principal, high on teacher variables) and Cluster 5 (high on principal, low on teacher variables) and these are AUS, ISR, ITA, MLT, LVA, USA, KOR, SVK, TWN, SGP, and BEL (see Appendix A for complete list and definitions of the abbreviations presented here). Some countries, such as PRT, ESP, VNM, and FRA, have the majority of their schools (more than 65%[7]) in Cluster 2, while GEO and SVN have more than 50% in Cluster 3. For these countries, we can say that leadership for learning practices are more homogeneous. In contrast, countries such as TUR, BEL, and MLT do not have a dominant leadership profile, meaning that leadership practices are more heterogeneous within these countries. Note that countries with a similar distribution of schools across clusters do not indicate countries that could easily be classified as representing geographically, linguistically, or politically proximal countries, with the possible exception of the four countries in which the majority of schools belong to Cluster 2.

Cluster composition

In order to describe schools within clusters and respond to our third RQ, we examined descriptive statistics regarding school background characteristics, including public vs. private schools, schools in urban vs. rural communities, percentage of students whose first language differs from the language of instruction, school size, and principal gender. Table 6 shows the percentage of schools across clusters in relation to the various school and principal characteristics as well as the total percentage of schools in each category. The chi-square test did not reveal any substantial differences across clusters with respect to the background characteristics of schools and principals. This is in line with previous research on instructional leadership. Instead, all clusters reflected the overall distributions of these characteristics.

Discussion and conclusion

In this paper, we established five distinct clusters of leadership for learning across schools that participated in TALIS 2018. Cluster 1 is characterized by weak leadership for learning practice at all levels where lower scores are obtained on all variables, reflecting the theoretical dimensions represented in the instruments. Countries such as BRA, CHL, COL, CYP, CZE, BEL, NZL, and NLD have more than 30% of their schools in this cluster. At the same time, in this cluster, 15% of all schools are uncertainly classified. Cluster 2 is

Table 5. List of countries with more than 30% of schools in respective clusters.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	< 30% in all clusters
NZL, BRA, CHL, CZE, COL, *NLD, CYP, BEL*	PRT, ESP, VNM, FRA	GEO, ROU, SVN, LVA, SAU, NOR, KAZ, RUS, CSH, BGR, *USA, KOR*, EST, LTU, HRV, ZAF, MEX, AUT, FIN, CAB, *SGP, CYP, BEL, CZE, MLT, ARE, AUS, ISR, SVK, TWN, ITA*	ARE	ENG, *AUS, ISR, ITA, NLD, MLT, LVA, USA, KOR, SVK, TWN, SGP,BEL*	TUR

Table 6. The distribution of schools across school and principal background characteristics.

	Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Public/Private						
Public	80.40%	22.17%	12.97%	32.42%	10.28%	22.15%
Private	19.60%	23.26%	11.63%	25.61%	9.62%	29.87%
Language						
None	32.97%	21.60%	7.00%	33.58%	14.28%	23.54%
1–10	44.37%	22.76%	16.63%	29.19%	7.49%	23.92%
11–30	10.14%	22.46%	11.68%	31.59%	6.89%	27.40%
31–60	5.27%	21.61%	5.19%	37.18%	7.78%	28.24%
More than 60	7.24%	26.62%	7.55%	33.33%	10.48%	22.01%
Location						
Village	15.87%	20.23%	11.67%	35.89%	11.87%	20.33%
Small town	20.58%	21.76%	17.03%	30.53%	9.00%	21.68%
Town	24.09%	22.44%	14.49%	28.72%	7.76%	26.60%
City	21.96%	21.73%	10.83%	31.86%	11.11%	24.47%
Large city	17.50%	25.95%	3.97%	32.30%	10.24%	27.54%
Sex of principal						
Female	48.78%	20.71%	9.54%	32.48%	11.52%	25.76%
Male	51.22%	24.27%	13.74%	30.46%	8.49%	23.03%
School size						
Under 250	22.38%	18.77%	22.37%	21.54%	14.13%	23.20%
250–500	12.01%	10.19%	27.61%	26.32%	14.71%	21.16%
500–749	31.41%	22.06%	24.78%	21.52%	11.35%	20.29%
750–1000	10.11%	27.15%	24.54%	18.10%	10.89%	19.33%
≥ 1000	24.09%	16.34%	22.52%	19.63%	13.51%	27.99%

characterized mostly by neutral leadership for learning practice with a substantial emphasis on school autonomy for instructional practices and teacher professional development. This is the smallest and most distinct cluster. Latin-speaking countries, such as ESP, FRA, and PRT, as well as VNM, have most of their schools in this cluster. Cluster 3 is characterized by the fact that the leadership for learning practices are balanced at the schools belonging to it, with all indicators being moderately represented. This is the cluster in which the majority of the schools are found for more than half of the countries in our sample. SVN and GEO, for example, have more than 50% of their schools in this cluster. Cluster 4 which describes strong leadership for learning practice at all levels is characterized by a stronger instructional/curricular/assessment program, stronger orientation toward learning, stronger organizational culture, and stronger social advocacy. This is also the smallest cluster in terms of the absolute number of schools globally. Accordingly, this cluster is not a dominant one, with ARE being an exception with more than 30% of its schools being in this cluster. Furthermore, this cluster contains 33.4% of all uncertainly classified schools, suggesting that this cluster is not perfectly empirically isolated from one or more of the other clusters. Cluster 5 is characterized by Leadership for learning practice that is oriented toward stronger dimensions related to instruction, curriculum, and assessment but balanced on organizational culture and communities for learning. English speaking countries ENG, USA, AUS, but also other countries such KOR, ITA, MLT have more than 30% schools in this cluster. It caught our attention that in the two biggest clusters (Cluster 3 and Cluster 5) principals and teachers reported differently about distributed leadership practice, creating a potential gap in how leadership is perceived within schools. This issue has been already investigated in the literature showing gaps in perceptions between teachers and principals not only when leadership is studied (Goff et al., 2014; Urick & Bowers, 2017) but also for other phenomena (Brezicha

et al., 2019). Further research could investigate how such gaps in perception shape school dynamics and climate.

Taken together, we can say that the distribution of schools that belong to a specific cluster at the country level does not reflect easily identifiable geographical, linguistic, or cultural similarities, which is contrary to our initial expectations given the theoretical background (Hallinger, 2018; Printy & Liu, 20210; Walker & Dimmock, 2002). However, we do find that some countries have more homogeneous leadership practices – ESP, FRA, and PRT constitute one such cluster of countries, as already mentioned. At the other extreme of this spectrum countries such as TUR, SAU, ARE, BRA, BGR, and CYP are found, representing countries with very heterogeneous leadership practices and schools that are relatively evenly distributed across all clusters. Although we cannot relate these findings to specific previous literature on leadership in each of these countries, a study by David and Abukari (2020) on school leadership in the United Arab Emirates provides an interesting example – concluding that there is a lack of robust national policies and strategies on school leadership in this country. This finding is consistent with the observed heterogeneous leadership practices in this part of the world.

In contrast to the qualitative literature, which shows that school leadership is dependent on wider societal norms (Harris, 2020; Møller & Schratz, 2009), we find that leadership for learning predominantly is a school-level phenomenon within each country. In the current study, this is first supported by our three-level analysis, which decomposed the variance across teachers, schools, and countries. For most variables, the proportion of variance between countries is very small. The relatively homogenous distribution across clusters in most countries in the sample further supports this finding. The two variables with marked between-country variance are related to the degree to which schools stimulate effective professional development and the extent to which teachers report having high instructional autonomy. Both of these characteristics are particularly prominent in the profile for Cluster 2. This cluster is also the only cluster completely dominated by only a few select countries that were previously categorized as Latin-speaking European countries. Furthermore, consistent with previous research, we do not find that local school context (private/public, urban/rural, etc.) is substantially related to cluster classifications. However, the substantial variability at the school level might be explained by other variables that are not examined in this study but have been proven of great importance for leadership practice, such as school climate and school culture. The scope of this paper and availability of data in the TALIS dataset did not allow for targeted analysis of school level factors that could account for the variability of school clusters. Moreover, the present study does not include student SES, student achievement, teacher experience and teacher effectiveness that are relevant factors when leadership is studied at the school level. Future international comparative studies of school leadership would benefit greatly from including empirical observations of such school characteristics.

In addition to the obvious issue of omitted variables, the limitations of this study primarily relate to the specific method applied, the mixture of teacher and principal scales in one joint single analysis, and the previously discussed issue of measurement invariance. Cluster analysis is a purely descriptive method and, as such, it only describes the data at hand by grouping unorganized observations into a specified number of clusters, including no prior hypothesis and no possibility to test the outcome according to model-

based assumptions. Another issue with cluster analysis is its inability to handle missing data. Consequently, only observations without missing data are included in the final analysis. Yet, in our case, the sample size was not substantially affected. Alternatively, latent profile analysis (LPA) could have been employed using the same data. However, this is a model-based approach that comes with certain assumptions, such as local independence for outcome variables. Specifically, this assumption implies that the associations between manifest variables included in the model are completely explained by their relationships with a latent variable (Hagenaars & McCutcheon, 2002). When this assumption does not hold, the model requires additional classes to be introduced, which results in continuous model fit improvement as additional classes are introduced. This is the case with our data as well, where latent profile analysis resulted (not reported) in an absurd situation – the inclusion of ever more classes improves the relative model fit. Another limitation relates to the use of aggregates of teacher data at the school level. This approach made it possible for us to represent leadership as a joint teacher and principal phenomenon. However, this approach does not fully account for the multilevel data structure (teachers nested in schools).

Furthermore, it should be noted that TALIS is not primarily a leadership survey. Rather, this is a broadly scoped study representing policy makers and researchers' diverse interests for studying a large set of school-related phenomena. This implies that the expert committees with the task to construct the questionnaires will have to make tough priorities of what to include or not in the final instruments. Therefore, the variables selected for inclusion in our analysis does not represent a rich and detailed perspective of one specific theory of leadership. Consequently, our analysis builds on data representing only a partial representation of the Leadership for Learning framework. It is therefore reasonable to assume that the specific solution presented in this paper is affected by omitting important variables. However, TALIS is currently the only study that collects this kind of data across a wide range of countries and, consequently, represents a unique source for studying leadership for learning worldwide.

Currently, comparative perspectives on how leadership is affected by cultural, ideological, and political values have largely been based on qualitative inquiry rooted in the analysis of policy documents and interviews with stakeholders. This literature presents interesting findings about how leadership is a culturally embedded practice. However, contrary to these findings, such patterns could not be identified in the unprecedentedly rich quantitative materials collected in the TALIS study, including data from more than 40 educational systems across the world. Instead, the five distinctly different leadership profiles, as reported collectively by teachers and principals, exist in almost all countries. Moreover, most countries are not dominated by only one or two of these profiles. This calls for additional and better-targeted research on educational leadership as a global phenomenon.

Notes

1. International Standard Classification of Education
2. A complete list of participating countries can be found in Appendix A.
3. Countries with more than 20% of data missing: SVN, KOR, NOR, COL, VNM, SAU, ISR, GEO.

4. Factor scores were calculated at the school level using Mplus version 8.4. on the pooled sample of schools. TYPE = COMPLEX and school weight were used to account for non-independence of observations and unequal probability of selection. The WLSMV estimator was used because data were categorical with three categories.
5. $Si = (Ci - Di)/\max(Di, Ci)$; where Di represents the average dissimilarity between each observation i and all other points within the same cluster; and Ci represents the dissimilarity between i and the cluster that is closest to i right after its own cluster.
6. The 30% cutoff point is a pragmatic choice. If all clusters were evenly distributed, each cluster would include 20% of the schools. Given that for some countries the number of schools can be as low as 150, a 95% confidence interval for the proportion 20% would span the interval from 13% to 27%. 30% is the next rounded number beyond this confidence interval, and hence, is chosen to indicate that a cluster is overrepresented in a country.
7. Full table with percentages available in Appendix B.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

Jelena Veletić and Rolf Vegar Olsen are part of the European training network OCCAM. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska-Curie grant agreement number 765400.

Notes on contributors

Jelena Veletić is a doctoral research fellow at the Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo, Norway. Her project is a part of the European Training Network (ETN) OCCAM which is a sub-call in the Marie Sklodowska-Curie Innovative Training Networks (MSCA ITN) of the European Commission's Horizon 2020 framework. Jelena's research relates to school leadership with a focus on studying variations of school leadership practices across systems, time- and school-level characteristics using large-scale assessment data.

Rolf Vegar Olsen is a professor at the Centre for Educational Measurement, Faculty of Educational Sciences, University of Oslo, Norway. He has a portfolio of research relating to national and international large-scale assessments. A main focus in his current research activities is to study how stakeholders can make use of data to inform and improve practices in classrooms and schools.

ORCID

Jelena Veletić http://orcid.org/0000-0002-3240-9674
Rolf Vegar Olsen http://orcid.org/0000-0002-9621-4083

References

Ainley, J., & Carstens, R. (2018). *Teaching and Learning International Survey (TALIS) 2018 conceptual framework* (OECD Education Working Papers, No. 187). https://doi.org/10.1787/799337c2-en

Ashforth, B. E. (1985). Climate formation: Issues and extensions. *The Academy of Management Review*, *10*(4), 837–847. https://doi.org/10.5465/amr.1985.4279106

Blitz, M. H., & Modeste, M. (2015). The differences across distributed leadership practices by school position according to the Comprehensive Assessment of Leadership for Learning (CALL). *Leadership and Policy in Schools*, *14*(3), 341–379. https://doi.org/10.1080/15700763.2015.1024328

Bossert, S. T., Dwyer, D. C., Rowan, B., & Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64. https://doi.org/10.1177/0013161X82018003004

Bowers, A. J. (2020). *Examining a congruency-typology model of leadership for learning using two-level latent class analysis with TALIS 2018* (OECD Education Working Papers No. 219). https://doi.org/10.1787/c963073b-en

Boyce, J., & Bowers, A. J. (2018). Toward an evolving conceptualization of instructional leadership as leadership for learning: Meta-narrative review of 109 quantitative studies across 25 years. *Journal of Educational Administration*, *56*(2), 161–182. https://doi.org/10.1108/JEA-06-2016-0064

Brauckmann, S., & Pashiardis, P. (2011). A validation study of the leadership styles of a holistic leadership theoretical framework. *International Journal of Educational Management*, *25*(1), 11–32. https://doi.org/10.1108/09513541111100099

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

Brezicha, K. F., Ikoma, S., Park, H., & LeTendre, G. K. (2019). The ownership perception gap: Exploring teacher job satisfaction and its relationship to teachers' and principals' perception of decision-making opportunities. *International Journal of Leadership in Education*, *23*(4), 428–456. https://doi.org/10.1080/13603124.2018.1562098

Byrne, B. M., & Vijver, F. J. R. van de. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132.

Clarke, S., & O'Donoghue, T. (2016). Educational leadership and context: A rendering of an inseparable relationship. *British Journal of Educational Studies*, *65*(2), 167–182. https://doi.org/10.1080/00071005.2016.1199772

Crow, G. M. (2006), "Complexity and the beginning principal in the United States: perspectives on socialization„. *Journal of Educational Administration*, *44*(4), 310–325. https://doi.org/10.1108/09578230610674930

Daniëls, E., Hondeghem, A., & Dochy, F. (2019). A review on leadership and leadership development in educational settings. *Educational Research Review*, *27*, 110–125. https://doi.org/10.1016/j.edurev.2019.02.003

David, S. A., & Abukari, A. (2020). Perspectives of teachers' on the selection and the development of the school leaders in the United Arab Emirates. *International Journal of Educational Management*, *34*(1), 56–69. https://doi.org/10.1108/IJEM-02-2019-0057

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

Dempster, N. (2019). Leadership for learning: Embracing purpose, people, pedagogy and place. In T. Townsend (Ed.), *Instructional leadership and leadership for learning in schools: Understanding theories of leading* (pp. 403–421). Springer International Publishing. https://doi.org/10.1007/978-3-030-23736-3_16

DuPont, J. P. (2009). *Teacher perceptions of the influence of principal instructional leadership on school culture: A case study of the American Embassy School in New Delhi, India*. Unpublished doctoral dissertation. Minnesota, MN: University of Minnesota. http://conservancy.umn.edu/handle/11299/50822

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Goff, P. T., Goldring, E., & Bickman, L. (2014). Predicting the gap: Perceptual congruence between American principals and their teachers' ratings of leadership effectiveness. *Educational Assessment, Evaluation and Accountability*, *4*(26), 333–359. https://doi.org/10.1007/s11092-014-9202-5

Gronn, P. (2002). Distributed leadership. In K. Leithwood, P. Hallinger, G. C. Furman, K. Riley, J. MacBeath, P. Gronn, & B. Mulford (Eds.), *Second international handbook of educational leadership and administration* (pp. 653–696). Springer Netherlands. https://doi.org/10.1007/978-94-010-0375-9_23

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge University Press.

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. (2008, March). *Methodologies for studying school leadership: A review of 25 years of research using the principal instructional management rating scale* [Paper presentation]. Annual meeting of the American Educational Research Association, New York.

Hallinger, P. (2009, September). *Leadership for the 21st century schools: From instructional leadership to leadership for learning* [Paper presentation]. Chair professors: Public lecture series, The Hong Kong Institute of Education, China. https://repository.eduhk.hk/en/publications/leadership-for-the-21st-century-schools-from-instructional-leader-3

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. (2015). The evolution of instructional leadership. In P. Hallinger & W.-C. Wang (Eds.), *Assessing instructional leadership with the principal instructional management rating scale* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-15533-3_1

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., & 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

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

Halverson, R., Kelley, C., & Shaw, J. (2014). A call for improved school leadership. *Phi Delta Kappan*, *95*(6), 57–60. https://doi.org/10.1177/003172171409500612

Harris, A. (2004). Distributed leadership and school improvement: Leading or misleading? *Educational Management Administration & Leadership*, *32*(1), 11–24. https://doi.org/10.1177/1741143204039297

Harris, A. (Ed.). (2009). *Distributed leadership: Different perspectives*. Springer.

Harris, A. (2020). Leading school and system improvement: Why context matters. *European Journal of Education*, *55*(2), 143–145. https://doi.org/10.1111/ejed.12393

Herborn, K., Mustafić, M., & Greiff, S. (2017). Mapping an experiment-based assessment of collaborative behavior onto collaborative problem solving in PISA 2015: A cluster analysis approach for collaborator profiles. *Journal of Educational Measurement*, *54*(1), 103–122. https://doi.org/10.1111/jedm.12135

Hooge, E. (2020). The school leader is a make-or-break factor of increased school autonomy. *European Journal of Education*, *55*(2), 151–153. https://doi.org/10.1111/ejed.12396

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. SAGE Publications.

Hoy, W. K. (1990). Organizational climate and culture: A conceptual analysis of the school workplace. *Journal of Educational and Psychological Consultation*, *1*(2), 149–168. https://doi.org/10.1207/s1532768xjepc0102_4

IMB Corp. (2017) . *IBM SPSS Statistics for Windows* (Version 26) [Computer software].

Imig, D., Holden, S., & Placek, D. (2019). Leadership for learning in the US. In T. Townsend (Ed.), *Instructional leadership and leadership for learning in schools: Understanding theories of leading* (pp. 105–131). Springer International Publishing. https://doi.org/10.1007/978-3-030-23736-3_5

Jacobson, S. L., & Johnson, L. (2011). Successful leadership for improved student learning in high needs schools: U.S. Perspectives from the International Successful School Principalship Project (ISSPP). In T. Townsend & J. MacBeath (Eds.), *International handbook of leadership for learning* (pp. 553–569). Springer Netherlands. https://doi.org/10.1007/978-94-007-1350-5_31

Johnson, L., Møller, J., Jacobson, S. L., & Wong, K. C. (2008). Cross-national Comparisons in the International Successful School Principalship Project (ISSPP): The USA, Norway and China. *Scandinavian Journal of Educational Research*, *52*(4), 407–422. https://doi.org/10.1080/00313830802184582

Kalkan, Ü., Altınay Aksal, F., Altınay Gazi, Z., Atasoy, R., & Dağlı, G. (2020). The relationship between school administrators' leadership Styles, school culture, and organizational image. *SAGE Open*, *10*(1), 2158244020902081. https://doi.org/10.1177/2158244020902081

Karada, E., & Öztekin, O. (2018). The effect of authentic leadership on school culture: A structural equation model. *International Journal of Educational Leadership and Management*, *6*(1), 40–75. https://doi.org/10.17583/ijelm.2018.2858

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning (Multivariate Analysis I)* (Vol. 1, 1st ed.). Statistical Tools For High-Throughput Data Analysis (STHDA). https://www.amazon.com/Practical-Guide-Cluster-Analysis-Unsupervised/dp/1542462703

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and visualize the results of multivariate data analyse (1.0.7)* [Computer software]. Software.

Knapp, M. S., Honig, M. I., Plecki, M. L., Portin, B. S., & Copland, M. A. (2014). *Learning-focused leadership in action: Improving instruction in schools and districts*. Routledge.

Krüger, M. L., 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

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., & Jantzi, D. (2006). Transformational school leadership for large-scale reform: Effects on students, teachers, and their classroom practices. *School Effectiveness and School Improvement*, *17*(2), 201–227. https://doi.org/10.1080/09243450600565829

Leithwood, K., Jantzi, D., & Steinbach, R. (2012). *Changing leadership for changing times*. Open University Press.

Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Educational Administration Quarterly*, *46*(5), 671–706. https://doi.org/10.1177/0013161X10377347

Liu, Y., Bellibaş, M. Ş., & Gümüş, S. (2021). 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*, *49*(3), 430–453. https://doi.org/10.1177/1741143220910438

Lubke, G. H., & Muthén, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(4), 514–534. https://doi.org/10.1207/s15328007sem1104_2

MacBeath, J., & Townsend, T. (2011). Thinking and acting both locally and globally: What do we know now and how do we continue to improve? In T. Townsend & J. MacBeath (Eds.),

International handbook of leadership for learning (pp. 1237–1254). Springer Netherlands. https://doi.org/10.1007/978-94-007-1350-5_66

MacBeath, J. (2019). Leadership for learning. In T. Townsend (Ed.), *Instructional leadership and leadership for learning in schools: Understanding theories of leading*(pp. 49–75). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-23736-3

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). *Cluster: Cluster analysis basics and extensions (R package version 2.1.0.)* [Computer software]. Software.

Mango, E. (2018). Beyond leadership. *Open Journal of Leadership*, *7*(1), 117–143. https://doi.org/10.4236/ojl.2018.71007

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

Miller, P. W. (2018). The nature of school leadership. In P. W. Miller (Ed.), *The nature of school leadership: Global practice perspectives* (pp. 165–185). Springer International Publishing. https://doi.org/10.1007/978-3-319-70105-9_9

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Møller, J., & Schratz, M. (2009). Leadership development in Europe. In J. Lumby, G. M. Crow, & P. Pashiardis (Eds.), *International handbook on the preparation and development of school leaders* (pp. 341–366). Routledge.

Muijs, D., & Harris, A. (2003). Teacher leadership—Improvement through empowerment?: An overview of the literature. *Educational Management & Administration*, *31*(4), 437–448. https://doi.org/10.1177/0263211X030314007

Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership & Management*, *27*(2), 179–201. https://doi.org/10.1080/13632430701237420

Murphy, J., Elliott, S. N., Goldring, E., & Porter, A. C. (2007). Leadership for learning: A research-based model and taxonomy of behaviors. *School Leadership & Management*, *27*(2), 179–201. https://doi.org/10.1080/13632430701237420

Opdenakker, M.-C., & Damme, J. V. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal*, *33*(2), 179–206. https://doi.org/10.1080/01411920701208233

Oplatka, I., & Arar, K. (2017). Context and implications document for: The research on educational leadership and management in the Arab world since the 1990s: A systematic review. *Review of Education*, *5*(3), 308–310. https://doi.org/10.1002/rev3.3096

Organisation for Economic Co-operation and Development. (2018). *TALIS 2018 technical report* [PDF file]. Paris: OECD Publishing. https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf

Pont, B., Nusche, D., & Moorman, H. (2008). *Improving school leadership, volume 1: Policy and practice* [Online]. OECD Library. https://read.oecd-ilibrary.org/education/improving-school-leadership_9789264044715-en

Printy, S., & Liu, Y. (2021). Distributed leadership globally: The interactive nature of principal and teacher leadership in 32 countries. *Educational Administration Quarterly, 57*(2), 290–325. https://doi.org/10.1177/0013161X20926548

Printy, S. M. (2008). Leadership for teacher learning: A community of practice perspective. *Educational Administration Quarterly*, *44*(2), 187–226. https://doi.org/10.1177/0013161X07312958

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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, L., & Svetina, D. (2014). Assessing the Hypothesis of Measurement Invariance in the Context of Large-Scale International Surveys. *Educational and Psychological Measurement*, *74* (1), 31–57. https://doi.org/10.1177/0013164413498257

Sahin, S. (2011). The relationship between instructional leadership style and school culture (Izmir case). *Educational Sciences: Theory and Practice*, *11*(4), 1920–1927. https://eric.ed.gov/?id=EJ962681

Schley, W., & Schratz, M. (2011). Developing leaders, building networks, changing schools through system leadership. In T. Townsend & J. MacBeath (Eds..), *International handbook of leadership for learning* (Vol. 25, pp. 267–295). Springer. https://link.springer.com/chapter/10.1007/978-94-007-1350-5_17

Slater, R. O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement*, *3*(4), 247–257. https://doi.org/10.1080/0924345920030402

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies*, *36*(1), 3–34. https://doi.org/10.1080/0022027032000106726

Tan, C. Y., Gao, L., & Shi, M. (2020). Second-order meta-analysis synthesizing the evidence on associations between school leadership and different school outcomes. *Educational Management Administration & Leadership*, 174114322093545. https://doi.org/10.1177/1741143220935456

Townsend, T. (2019). *Instructional Leadership and Leadership for Learning in Schools*: *Understanding Theories of Leading*. Springer International Publishing.

United Nations Educational, Scientific and Cultural Organization. (2016). *Leading better learning*: *School leadership and quality in the Education 2030 agenda*. UNESCO. https://silo.tips/download/leading-better-learning-school-leadership-and-quality-in-the-education-2030-agen

Urick, A., & Bowers, A. J. (2014). What are the different types of principals across the United States? A latent class analysis of principal perception of leadership. *Educational Administration Quarterly*, *50*(1), 96–134. https://doi.org/10.1177/0013161X13489019

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*, *18*(3), 249–269. https://doi.org/10.1080/15700763.2017.1384499

Walker, A., & Dimmock, C. (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, & B. Mulford (Eds.), *Second international handbook of educational leadership and administration* (pp. 167–202). Springer Netherlands. https://doi.org/10.1007/978-94-010-0375-9_7

Wang, M.-T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Zieger, L., Sims, S., & Jerrim, J. (2019). Comparing Teachers' Job Satisfaction across Countries: A Multiple-Pairwise Measurement Invariance Approach. *Educational Measurement: Issues and Practice*, *38*(3), 75–85. https://doi.org/10.1111/emip.12254

Appendices

Appendix A

List of countries participating in TALIS 2018

	Countries	Code
1	Argentina (Buenos Aires)	ABA
2	Australia	AUS
3	Austria	AUT
4	Belgium Flemish	BEL
5	Belgium French Communities	BFL
6	Brazil	BRA
7	Bulgaria	BGR
8	Canada (Alberta)	CAB
9	Chile	CHL
10	China (People's Republic of Shanghai)	CSH
11	Chinese Taipei	TWN
12	Cyprus	CYP
13	Colombia	COL
14	Croatia	HRV
15	Czech Republic	CZE
16	Denmark	DNK
17	Estonia	EST
18	Finland	FIN
19	France	FRA
20	Georgia	GEO
21	Hungary	HUN
22	Iceland	ISL
23	Israel	ISR
24	Italy	ITA
25	Japan	JPN
26	Kazakhstan	KAZ
27	Korea	KOR
28	Latvia	LVA
29	Lithuania	LTU
30	Malta	MLT
31	Mexico	MEX
32	Netherlands	NLD
33	New Zealand	NZL
34	Norway	NOR
35	Portugal	PRT
36	Romania	ROU
37	Russian Federation	RUS
38	Saudi Arabia	SAU
39	Singapore	SGP
40	Slovak Republic	SVK
41	Slovenia	SVN
42	South Africa	ZAF
43	Spain	ESP
44	Sweden	SWE
45	Turkey	TUR
46	United Arab Emirates	ARE
47	United Kingdom (England)	ENG
48	United States	USA
49	Vietnam	VNM

Appendix B

In this table, countries are sorted with increasing proportions of schools in Cluster 1. Entries in the table that represent more than 30% of schools in one of the clusters are colored as follows: Cluster 1

in red; Cluster 2 in green; Cluster 3 in mint; Cluster 4 in blue; Cluster 5 in lilac. In addition, the cells providing the country labels are colored to identify the most frequent cluster in each country. The distribution of clusters within countries differs. Some countries have three or more equally frequent profiles (e.g. Turkey, Brazil, Belgium, Singapore, Chinese Taipei), while others have the majority of their schools within one cluster (e.g. Portugal, Spain, Vietnam, France). Only one country (United Arab Emirates) has more than 30% of schools in Cluster 4, characterized by higher values on all variables.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
PRT	4.12%	90.21%	1.55%	1.55%	2.58%
ESP	4.42%	88.83%	1.04%	2.86%	2.86%
VNM	6.57%	78.10%	2.19%	6.57%	6.57%
FRA	7.64%	67.52%	8.92%	5.10%	10.83%
GEO	12.62%	0.00%	53.40%	11.65%	22.33%
ROU	12.87%	0.00%	48.54%	14.62%	23.98%
MLT	15.22%	2.17%	41.30%	2.17%	39.13%
SVN	17.39%	0.00%	52.17%	6.52%	23.91%
LVA	17.76%	0.00%	40.19%	11.21%	30.84%
SAU	17.95%	0.85%	40.17%	18.80%	22.22%
NOR	18.85%	0.82%	47.54%	9.02%	23.77%
ARE	19.12%	0.00%	30.15%	30.88%	19.85%
KAZ	20.66%	0.00%	41.33%	8.86%	29.15%
RUS	20.73%	0.52%	40.41%	11.92%	26.42%
TUR	21.53%	20.14%	25.00%	15.97%	17.36%
CSH	21.64%	0.00%	41.52%	11.11%	25.73%
BGR	22.42%	2.42%	41.82%	9.70%	23.64%
USA	22.90%	0.76%	35.11%	10.69%	30.53%
KOR	23.33%	0.00%	40.83%	5.83%	30.00%
EST	24.32%	0.00%	38.51%	7.43%	29.73%
AUS	24.86%	0.00%	32.97%	7.03%	35.14%
LTU	25.64%	0.00%	42.31%	8.33%	23.72%
ISR	26.09%	0.00%	29.57%	10.43%	33.91%
SVK	27.03%	0.00%	36.49%	4.73%	31.76%
TWN	27.37%	0.00%	34.08%	5.59%	32.96%
HRV	27.81%	0.00%	33.77%	9.93%	28.48%
ZAF	27.82%	0.00%	41.35%	10.53%	20.30%
MEX	28.10%	0.00%	40.52%	13.07%	18.30%
ITA	28.22%	0.00%	32.52%	4.91%	34.36%
AUT	28.36%	0.00%	33.33%	9.95%	28.36%
FIN	28.37%	0.00%	34.75%	7.09%	29.79%
CAB	28.43%	0.00%	31.37%	15.69%	24.51%
SGP	28.86%	0.00%	35.57%	2.68%	32.89%
ENG	29.69%	0.00%	23.44%	6.25%	40.63%
NZL	30.46%	1.32%	25.83%	14.57%	27.81%
BRA	31.21%	2.13%	28.37%	16.31%	21.99%
CYP	32.00%	1.33%	42.67%	6.67%	17.33%
BEL	32.10%	0.41%	30.86%	4.53%	32.10%
CHL	32.88%	0.00%	28.77%	13.01%	25.34%
CZE	33.87%	0.00%	33.33%	3.23%	29.57%
COL	34.51%	0.00%	28.32%	18.58%	18.58%
NLD	37.50%	0.00%	20.83%	1.04%	40.63%

ERRATA TO PART II

Page	Original	Edited
XII		
XIII		
XIV