Studies in Educational Evaluation 87 (2025) 101521

Contents lists available at ScienceDirect

Studies in Educational Evaluation

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

How school leadership and innovation shape instructional pathways to student achievement across nations: Evidence from multilevel structural equation modeling and decision tree analysis

Burak Aydın [a,b], Marjolein Fokkema [c], Nurullah Eryilmaz [b,d], Daniel Muijs [e], Ronny Scherer [f], Marcus Pietsch [b,g,*]

[a] School of Education, Ege University, Türkiye
[b] School of Education, Leuphana University, Germany
[c] Department of Methods & Statistics for Psychology, Leiden University, the Netherlands
[d] Research and Analysis Unit, International Association for the Evaluation of Educational Achievement, Germany
[e] School of Education, Queen's University Belfast, UK
[f] Centre for Educational Measurement (CEMO), University of Oslo, Norway
[g] School of Management & Technology, Leuphana University, Germany

ARTICLE INFO

Keywords:
Educational Leadership
Innovation
Multilevel Structural Equation Modelling
Multilevel Decision Trees
PISA-TALIS 2018

ABSTRACT

Educational leadership, innovation and teaching play essential roles in shaping student achievement. However, extant literature primarily has relied on linear modelling approaches and has not focused on substantively testing a theory. The present study employs multilevel structural equation modelling (ML-SEM) and multilevel decision trees (MLM trees) to investigate associations between school leadership, team innovation, cognitive activation and student achievement using PISA-TALIS 2018 linked data across seven countries: Australia; Colombia; Czech Republic; Denmark; Georgia; Malta; and Türkiye. The ML-SEM findings indicated no significant indirect effects from leadership on achievement. The MLM trees highlighted country-specific patterns in associations between school leadership, innovation and student achievement, revealing potential nonlinear relationships. These findings suggest that the relationship between leadership, instructional practices and achievement is highly context-dependent. The study contributes to the literature by offering a comparative analysis of ML-SEM and MLM trees, highlighting traditional linear models' limitations in educational research.

1. Introduction

Educational leadership is crucial in shaping school environments, fostering innovation and change in educational practices, and ultimately enhancing student learning outcomes (Blömeke et al., 2021; Liu et al., 2024; Pietsch et al., 2025). Traditionally, educational research predominantly has examined these relations through linear modelling approaches, assuming uniform effects across different educational contexts (Goff & Finch, 2016). However, emerging evidence suggests that the relations between leadership, instructional practices and student achievement are highly complex and often nonlinear, requiring more flexible methodological approaches to uncover hidden patterns (Hallinger, 2011; OECD, 2020).

Both school effectiveness (Creemers & Kyriakides, 2008) and educational leadership (Hallinger & Heck, 2010) literature has assumed that leadership influences teachers' instructional practices, which, in turn, affects student achievement. As instructional practices are not enacted in a vacuum but are inherently embedded in their surrounding contexts, scholars have increasingly emphasised leadership and other organisational factors as pivotal determinants of how these practices unfold in schools. Consequently, Praetorius et al. (2025), in a recent literature review, argue that teaching quality must be understood as a context-sensitive concept: variations in school climate, culture, structures, and resources can fundamentally shape both the opportunities and the constraints for teaching and instruction. In line with this perspective, an innovative school climate in particular is believed to foster the implementation of innovative teaching behaviours, thereby enhancing the classroom's instructional quality (Blömeke et al., 2021). In essence,

* Corresponding author at: School of Education, Leuphana University, Germany.
E-mail address: marcus.pietsch@leuphana.de (M. Pietsch).

https://doi.org/10.1016/j.stueduc.2025.101521

Available online 16 October 2025

an innovative climate is a work climate in which innovation is supported and incentivised (Mumford, 2000; Pietsch et al., 2024), and can be assessed by surveying shared perceptions at the organisational level (Anderson & West, 1998; Siegel & Kaemmerer, 1978) regarding the extent to which processes encourage and enable innovation (Newman et al., 2020).

Following this understanding, in educational research, collective teacher innovativeness—defined as teachers' collective efforts to develop, test, and implement new teaching strategies (Dedering & Pietsch, 2025)—is recognised widely as a key variable moderating and/or mediating the relation between educational leadership and innovative instructional practices (Leithwood & Jantzi, 2006; Liu et al., 2024; Pan et al., 2024). While relations between leadership and teaching practices (e.g., Hallinger & Heck, 2010; Sebastian et al., 2016), as well as teaching practices and student achievement (e.g., Hattie, 2009; Klieme et al., 2009; Neugebauer & Prediger, 2023; Serdyukov, 2017; Tourón et al., 2019), and an innovative climate's role in these specific relations (e.g., Halász, 2018; Caro et al., 2016; Thurlings et al., 2015; Moolenaar et al., 2010) have been the subject of empirical investigation, a comprehensive model integrating all these variables remains lacking. Furthermore, the literature on how and under what conditions leadership influences instructional innovation—particularly through teacher collaboration and school climate—remains inconsistent (e.g., Sun & Leithwood, 2015; Marks & Printy, 2003). One reason for this might be that despite these well-established associations, extant studies rarely have accounted for heterogeneity in educational contexts, such as varying levels of school achievement, which may shape how leadership and instructional practices are enacted and experienced (Baumert et al., 2010).

Accordingly, this study aims to address these limitations by employing both multilevel structural equation modelling (ML-SEM; Muthén & Asparouhov, 2011) and multilevel decision trees (MLM trees; Fokkema et al., 2021) to investigate associations between school leadership, and in particular aspects of distributed and instructional leadership practice, team innovation, cognitive activation and student achievement. Unlike extant studies that mainly have been theory-driven, we also employ a data-driven machine learning approach to examine nonlinear relationships (Hilbert et al., 2021), allowing us to test and refine underlying theories and challenge common explanations in organisational (Leavitt et al., 2021) and particularly school effectiveness (Hu et al., 2022) research. Using PISA-TALIS 2018 linked data from seven countries (Australia, Colombia, Czech Republic, Denmark, Georgia, Malta and Türkiye), we aim to understand how leadership and instructional practices shape student learning.

Specifically, we seek to answer the following research questions:

- RQ1: While considering each country and achievement domain separately, assuming that relations between predictor variables are linear and employing a multilevel structural equation model, considering that students are nested in schools:

o RQ1a: To what extent do instructional and distributed leadership exert a direct effect on team innovation?
o RQ1b: To what extent do instructional and distributed leadership exert an indirect effect on achievement via team innovation and cognitive activation?

- RQ2: For the PISA-TALIS 2018 linked dataset, without assuming linearity and by employing a multilevel decision tree approach, considering that students are nested in schools and schools are nested in countries:

o RQ2a: Are achievement subgroups formed as a function of leadership, team innovation and cognitive activation-observed scores?
o RQ2b: If subgroups exist, to what extent is the separation of subgroups heterogeneous across countries?

Integrating ML-SEM and MLM trees, the present study contributes to a growing body of research advocating for nonlinear and context-sensitive educational analyses. Our findings provide new insights into the complexity of leadership effects on instructional practices and student learning outcomes, informing both policy and practice in international education.

1.1. Literature review

1.1.1. Leadership and collective teacher innovativeness

School leadership plays a central role in shaping the professional environment within schools, particularly in fostering team innovation among teachers. Effective leadership is associated with facilitating collaboration, promoting innovative teaching practices and encouraging professional learning communities (Hallinger, 2011). Leadership is a broad concept, but the educational research literature has typically concentrated on three main forms of leadership in schools. Transformational leadership is a concept that emerged from the business literature, and that refers to the ability of the leader to enact positive change by inspiring and motivating followers to achieve common goals and fulfill their potential (Bass & Riggio, 2006). Instructional leadership is an education-specific concept, which entails a focus on improving teaching and learning, emphasizing a principal's direct involvement in curriculum development, instructional practices, and professional development to foster a culture of continuous learning (Hallinger, 2011). Distributed leadership is a concept that originates from educational research and practice, but has since become widely used across different sectors. Spillane (2006, p. 204), one of the originators of this leadership theory, defines distributed leadership as 'activities tied to the core work of the organization that are designed by organizational members to influence the motivation, knowledge, affect, or practices of other organizational members'. This situates leadership as a process and a property of an organization, rather than at the level of the individual leader. Distributed leadership has a number of key elements, such as shared leadership between staff across the school, organisational learning, developing leadership capacity and empowerment. It is the latter two on which we will focus on in this study.

Transformational, instructional and distributed leadership models have been studied widely in this context, emphasising how leadership behaviours create conditions for innovation by empowering teachers and fostering shared decision-making (Leithwood & Jantzi, 2006; Buyukgoze et al., 2022; Cao et al., 2025; Vermeulen et al., 2022). Studies suggest that overall leadership has a small to medium-size effect on student attainment (Tan, Dimmock & Walker, 2021; Karadağ & Sertel, 2025; Wu & Shen, 2022), and there is support for a mediation model in which teacher behaviours mediate that relationship. A recent meta-analysis (Papadakis et al., 2024) found that leadership had direct effects on teacher performance (defined as a range of in- and out-of-class teacher behaviours such as assessing student progress and collaborating with other teachers; instructional leadership: $r = .60$; transformational leadership: $r = .40$; distributed leadership: $r = .19$), which in turn was significantly related to student achievement. For student achievement, total effects were strongest for distributed leadership ($r = .58$), followed by instructional ($r = .40$) and transformational ($r = .33$). Mediation analyses revealed, however, that the instructional leadership effect was fully mediated by teacher performance, whereas transformational and distributed leadership showed both direct and indirect relations. Another meta-analysis suggested that the leadership activities most strongly associated with improved student achievement were facilitating improvements in curriculum, instruction, and assessment, all factors traditionally seen as improving teaching quality (Shen et al., 2020).

There are a number of studies pointing to a relationship between leadership and innovative practices in schools. Harris (2009) in her overview of research up to then finds innovation and knowledge creation to be two key benefits of distributed leadership in schools, while

O'Shea (2021), using TALIS data found that higher levels of distributed leadership were associated with higher levels of innovative practice. Instructional leadership practices, such as goal-setting and fostering collaboration have also been found to be related to innovative practice in a number of studies (Hojeij, 2024). Empirical studies using international large-scale assessment (ILSA) data suggest that leadership styles that prioritise teacher collaboration and instructional support contribute positively to innovative teaching practices (OECD, 2020). However, the findings on leadership's direct influence on team innovation remain mixed, with some studies suggesting that leadership effects are mediated by school culture and teacher engagement (Tschannen-Moran, 2014). This highlights the need for nonlinear approaches to better capture these complex interactions, as traditional linear models may fail to account for variations across educational contexts.

1.1.2. Collective teacher innovativeness and cognitive activation

Numerous frameworks have been developed to describe and assess instructional practices. For example, in the US, Danielson's (2007) Framework for Teaching is highly used. This framework divides teaching approaches into 22 components clustered into four factors: planning and preparation, the classroom environment, instruction, and professional responsibilities. In Europe a much-used framework is the Three Basic Dimensions model, which focuses more strongly on in-class behaviours and distinguishes classroom management, student support and cognitive activation (Praetorius et al., 2018). In this context, it is striking that the existing models and measurement instruments consistently incorporate only two overarching dimensions (Bell et al., 2019): student involvement and cognitive activation.

Cognitive activation is viewed as a key element in teacher's instructional practices (Bell et al., 2019) and has been linked consistently to improved student outcomes (Klieme, 2013). Cognitive activation is the process of engaging a person's mind in deep, constructive, and reflective thinking to foster deeper learning and understanding. It involves engaging students in challenging tasks, activating prior knowledge and examining multiple solutions to foster higher-order thinking and deep understanding (Klieme et al., 2009; Praetorius et al., 2018). However, for teachers, applying cognitive activation in the classroom requires a high level of professional knowledge (Baumert et al., 2010; Förtsch et al., 2016), continuous variation in instructional practices (Walshaw & Anthony, 2008) and dynamic adaptability (Burić & Kim, 2020).

Consequently, collective teacher innovativeness, which encompasses teachers' willingness to develop new instructional methods and engage in professional collaboration, is linked closely to cognitive activation in classrooms (Hattie, 2009). Extant studies have indicated that teachers working on collaborative and innovative teams are more likely to adopt cognitively activating instructional methods, as they share best practices and experiment with new pedagogical approaches (Sleegers et al., 2014). Recent cross-national research using TALIS 2018 confirms this association, indicating that school innovativeness—as perceived by teachers—is related positively to teacher collaboration, job satisfaction and cognitively activating teaching practices. Moreover, while teacher collaboration and job satisfaction effects were relatively homogeneous across countries, the strength of associations with cognitive activation and innovative practices has varied more substantially (Blömeke et al., 2021). Extant research using PISA and TALIS datasets has found that teacher collaboration is associated positively with student engagement in cognitively demanding tasks, yet cross-country differences remain underexamined (OECD, 2019). The challenge lies in understanding whether team innovation causally enhances cognitive activation or whether these relations vary across school systems, necessitating methodological approaches that can capture nonlinear dependencies and contextual variations.

1.1.3. Cognitive activation and student achievement

Cognitive activation has been studied widely for its potential to enhance student outcomes in international large-scale assessments, such as PISA, TIMSS and PIRLS. Teaching strategies that require students to engage in deep reasoning, justify their answers and solve open-ended tasks have been theorised as promoting higher-order thinking and academic performance (Klieme, 2013; Seidel & Shavelson, 2007; Zhang et al., 2021). Several studies have supported these associations. For example, Genç and Çolakoğlu (2021) found that cognitive activation was associated positively with students' mathematical literacy in the PISA 2012 Turkey sample, with both direct and indirect effects mediated by self-concept and interest.

However, the empirical evidence is not entirely consistent. Bellens et al. (2019)—using TIMSS 2015 data from Flanders, Germany and Norway—found that while cognitive activation was associated with achievement, the dimension did not relate to educational equity, and the construct's cross-cultural comparability was problematic.

These divergent findings highlight cognitive activation's complexity and suggest that its effectiveness may depend on contextual, cultural and implementation-specific factors. Consequently, there is a growing need for research that examines how these strategies function across diverse educational systems and student populations using context-sensitive and methodologically robust approaches.

While the literature largely has relied on the assumption of linear effects between instructional practices and student outcomes, more recent studies have challenged this simplification, suggesting that the relationship between cognitive activation and student achievement may be curvilinear and context-dependent. For example, (Teig et al., 2018), using TIMSS 2015 data from Norway, found that the relationship between inquiry-based teaching—closely related to cognitive activation—and science achievement followed a curvilinear pattern. Specifically, moderate levels of inquiry-based instruction were associated with higher student achievement, while excessive use exerted a negative effect, thereby rejecting the notion of "the more, the better."

1.1.4. Leadership and cognitive activation

There are some studies that have specifically linked school leadership with cognitive activation. For example, Bellibaş et al. (2025) studied teachers in the UAE, and found a positive relationship between instructional leadership in schools and maths achievement, which was mediated by development of cognitive activation by teachers. By contrast, a recent study in China found a negative relationship between principal instructional leadership and use of cognitive activation by teachers (Huang, Zhao & Zhou, 2024). Pietsch & Tulowitzki (2017) looked at instructional, transformational, transactional and laissez-fair leadership styles, and found that the use of cognitive activation strategies by teachers was fostered by a combination of leadership styles.

1.1.5. Literature gaps and this study's contributions

Despite the growing body of research on leadership, team innovation and cognitive activation, three critical literature gaps remain:

1. Linear modelling limitations: Extant studies largely have assumed linear relations, limiting insights into potentially complex interactions.
2. Cross-country variability: Most research has focussed on single-country contexts, leaving cross-national differences in leadership effects underexamined.
3. Indirect pathways: While leadership often is linked to student achievement, its indirect effects through team innovation and cognitive activation remain underexamined.

The present study addresses these literature gaps by employing both ML-SEM and MLM trees, allowing for a comparative analysis of linear vs. nonlinear modelling approaches. Leveraging PISA-TALIS 2018 linked data across seven countries, this research provides a nuanced understanding of how leadership, innovation and cognitive activation interact to shape student achievement globally.

2. Methods

2.1. Data

This study utilised PISA-TALIS 2018 linked data (OECD, 2021), a unique dataset that integrates student achievement data from PISA with teacher and school-level data from TALIS. The PISA-TALIS linked data aims to provide a comprehensive understanding of how school and classroom environments influence or relate to student learning outcomes.

Although the original dataset encompasses nine countries—Argentina, Australia, Colombia, Czech Republic, Denmark, Georgia, Malta, Türkiye and Vietnam—we excluded Argentina and Vietnam due to unavailability of data on crucial variables related to leadership, team innovation and cognitive activation.

The dataset encompasses information about students, teachers and schools, allowing for a multilevel analytical approach. The sample sizes for each country in our analysis are reported in Table 1. Each student in the dataset has 10 plausible values for achievement scores in reading, mathematics and science, which are utilised in the analysis. Achievement data included student and school identifiers, whereas teaching-related data included teacher and school identifiers; thus, school-level variables, such as cognitive activation and team innovation, are aggregated from teacher responses to facilitate data merging and ensure consistency in measurement across educational institutions (schools).

The PISA-TALIS 2018 linked data offer a rich source of information to examine associations between leadership, team innovation, cognitive activation and student achievement, while accounting for the data's nested structure (students within schools and schools within countries).

2.2. Measures

2.2.1. Inbstructional leadership

Instructional leadership was measured using the T3PLEADS scale, which captures leadership behaviours aimed at fostering teacher collaboration and instructional improvement. Respondents indicated how frequently they engaged in leadership activities over the past 12 months on a four-point Likert scale. The measure comprises three items: (1) supporting teacher cooperation for new teaching practices (TC3G22D); (2) ensuring that teachers take responsibility for skill improvement (TC3G22E); and (3) fostering a sense of responsibility for student learning outcomes (TC3G22F). Higher scores reflect stronger leadership engagement in these areas (OECD, 2020). These variables are of course limited indicators of instructional leadership which do not encompass the full range of the construct, but are nevertheless useful indicators thereof.

2.2.2. Distributed leadership

Distributed leadership was measured using the TC3G26 scale, which assesses the extent to which different stakeholders participate in school decision-making. Respondents rated their agreement with three statements on a four-point Likert scale. The items include: (1) providing staff with opportunities to participate actively in school decisions (TC3G26A); (2) providing parents or guardians with opportunities to participate actively in school decisions (TC3G26B); and (3) providing students with opportunities to participate actively in school decisions (TC3G26C). Higher scores indicate a stronger emphasis on shared leadership practices within the school (OECD, 2020). As with instructional leadership, these variables are a somewhat limited selection of aspects of the concept.

Table 1
Sample Sizes of Students and Schools in the PISA-TALIS 2018 Linked Dataset.

Country	Number of Students	Number of Schools
Australia	2563	131
Colombia	4720	152
Czech Republic	4290	182
Denmark	2320	102
Georgia	3292	134
Malta	3354	49
Türkiye	5573	143
Total	26,112	893

2.3. Team innovativeness

Team innovativeness was measured using the T3TEAM scale, which captures teachers' openness to new ideas and practices in their schools, i.e., collective teacher innovativeness. Respondents rated their agreement with four statements on a four-point Likert scale. The items include: (1) striving to develop new ideas for teaching and learning (TT3G32A); (2) being open to change (TT3G32B); (3) searching for new ways to solve problems (TT3G32C); and (4) providing practical support to each other in applying new ideas (TT3G32D). Higher scores indicated a stronger culture of innovation among teachers (OECD, 2020).

2.4. Cognitive activation

Cognitive activation was measured using the T3COGAC scale, which assesses the extent to which teachers engage students in critical thinking activities. Respondents rated their frequency using these practices on a four-point Likert scale. The items include: (1) presenting tasks for which there is no obvious solution (TT3G42E); (2) assigning tasks that require students to think critically (TT3G42F); (3) dividing students into small groups to solve problems (TT3G42G); and (4) asking students to decide on their own procedures for solving complex tasks (TT3G42H). Higher scores indicate greater emphasis on cognitively engaging instructional practices (OECD, 2020).

2.5. Mathematics, science and reading achievement

Student achievement in mathematics, science and reading (i.e., literacy) was assessed using the PISA 2018 cognitive tests. Each student received 10 plausible values for each domain to account for measurement uncertainty. Mathematics proficiency measured problem-solving and quantitative reasoning skills, science proficiency assessed the application of scientific knowledge and reading proficiency evaluated text comprehension and critical analysis abilities. Achievement scores were standardised with a mean of 500 and a standard deviation of 100 across OECD countries, and these scores were reported to have substantial intra-class correlation (ICC), indicating strong variation at the school level (OECD, 2019).

2.6. Analytical strategy

Before conducting the main analyses, we first assessed the reliability of all scales used in this study. Internal consistency was evaluated using Cronbach's alpha and McDonald's omega (Trizano-Hermosilla & Alvarado, 2016; Zinbarg et al., 2005). Furthermore, a single-level confirmatory factor analysis (CFA) was conducted to test the scales' factor structures. We then proceeded with the two chosen primary analytical methods, ML-SEM and MLM trees, which allowed us to examine both linear and nonlinear relationships while accounting for the data's nested structure.

2.7. Multilevel structural equation modelling (ML-SEM)

The first analytical approach was ML-SEM, which was used to test the hypothesised direct and indirect effects of leadership on student achievement via team innovation and cognitive activation. Given the data's hierarchical nature, in which students are nested within schools, we specified a two-level model:

Level 1 (Student Level): Student achievement scores in mathematics, reading and science were modelled as outcome variables.

Level 2 (School Level): Leadership, team innovation and cognitive activation were included as school-level predictors.

Fig. 1 depicts the structural model we tested. Using *Mplus* software (Version 8.6; Muthén & Muthén, 2021), we specified the availability of 10 plausible values (PVs) via the TYPE = IMPUTATION option and utilised full information maximum likelihood (FIML) estimation to address missing data (see Appendix for the input and output files). Confidence intervals were bootstrapped, and PVs were rescaled separately for each country and achievement domain to have a mean of zero and a standard deviation of one. To assess model fit, we first ran the model without weights and viewed CFI > .90, TLI > .90, RMSEA < .06 and SRMR < .08 as indicating acceptable fit (Chen, 2007; Hu & Bentler, 1999; Rutkowski & Svetina, 2014). The models with acceptable fit were rerun with school weights. For the final step of the ML-SEM, we employed a random-effects meta-analysis with restricted maximum likelihood to estimate the pooled regression coefficient across countries. Specifically, we fed the regression coefficients and the square of the corresponding standard errors to the *rma* function in the *metafor* software package (Viechtbauer, 2010).

2.7.1. Multilevel decision trees (MLM trees)

The second analytical approach employed a multilevel decision tree model to capture potential nonlinear relationships between school-level variables and student achievement, aligning with recent advancements in multilevel decision-tree methods (Fokkema et al., 2021). The model was fitted to a long-format dataset in which all 30 plausible values (PVs; 10 for each of the three domains) were scaled to have a mean of zero and a standard deviation of one, then were stacked. This approach enabled inclusion of all PVs for every student across all seven countries in a multilevel decision tree, with each PV weighted at 1/30 to avoid inflating power artificially. The multilevel decision accounted for the correlation of plausible values within students, students within schools and schools within countries. To obtain a tree with subgroups large enough to allow for interpretation and stable estimates, we specified a minimum node size (Zeileis et al., 2008) and arbitrarily selected 1000.

All analyses were conducted using the R software package *glmertree* (Version 0.2–6; Fokkema & Zeileis, 2024), which integrates *lme4* (Bates et al., 2015) for fitting mixed models and *partykit* (Hothorn & Zeileis, 2015) for recursive partitioning. This package allows for identification of subgroups in hierarchical data by combining decision trees' flexibility with linear mixed models' robustness (Fokkema et al., 2018). Inspired by Hajjem et al. (2017) and Sela and Simonoff (2012), to examine between-country variability, we implemented the following approach: First, we randomly selected 50 % of the schools and fitted the same tree model. The resulting tree structure then was used to predict node membership for the remaining 50 %, similar to cross-validation procedures in GLMM tree research. Subsequently, we estimated a random-intercepts, three-level model, allowing node effects/means to vary across countries using the second half of the dataset. Furthermore, we estimated a random-intercepts, three-level model with the second half to assess variance components (Fokkema et al., 2018). To assess the results' consistency, we computed correlations between tree-based node means and mixed-model-based node means (Fokkema et al., 2021). Finally, graphical representations were generated to visualise node structures and their variations across countries, following best practices in decision-tree research (Fokkema et al., 2021). By combining ML-SEM and MLM trees, we tested specific indirect effects among leadership, team innovativeness and cognitive activation, while also examining possible intersections between these constructs related to variations in achievement levels. The combination of these approaches enhanced our understanding of how leadership and instructional practices influence student achievement.

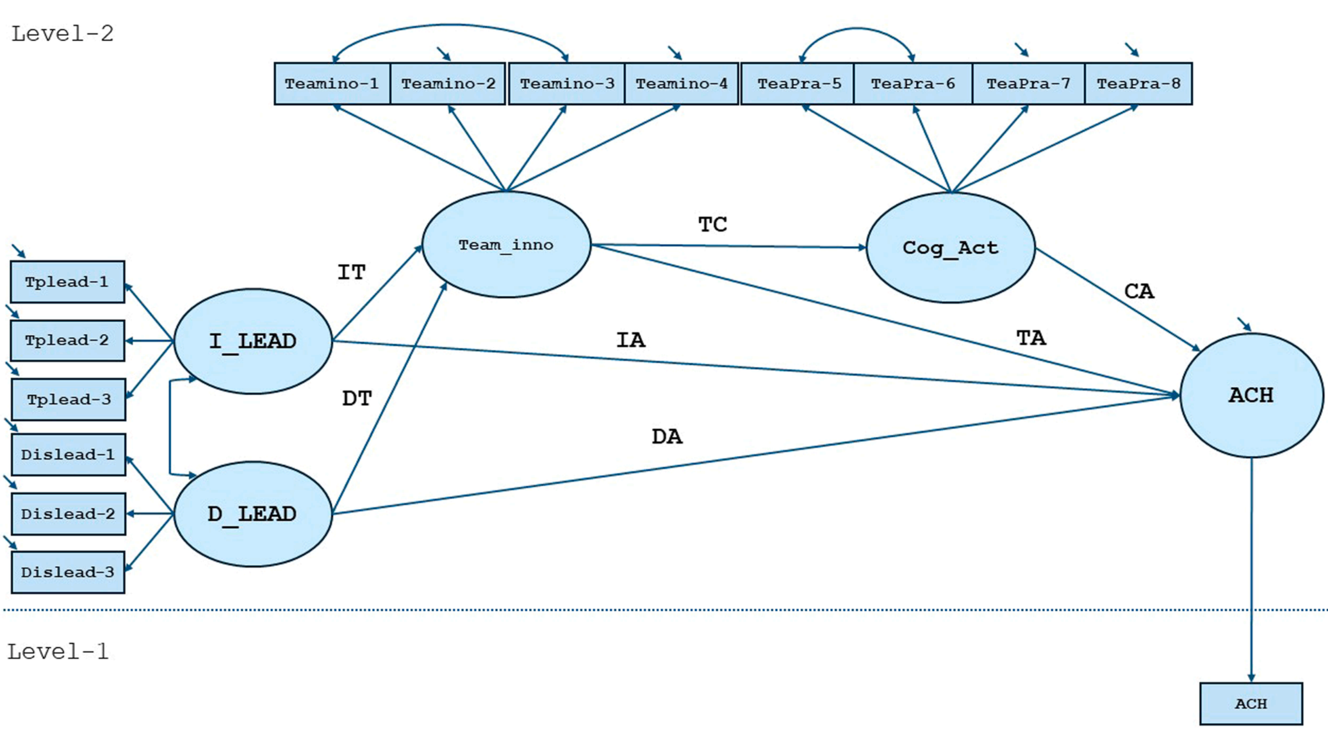

Fig. 1. *A structural equation model was tested separately for each country and each achievement domain. Note.* IA = Instructional Leadership Direct Effect on Achievement, DA = Distributed Leadership Direct Effect, TA = Team Innovativeness Direct Effect, CA = Cognitive Activation Direct Effect, ITCA = Indirect Effect of Instructional Leadership on Cognitive Activation via Team Innovativeness; DTCA = Indirect Effect of Distributed Leadership on Cognitive Activation via Team Innovativeness.

3. Results

3.1. Descriptive statistics, scale properties and correlations

To ensure the constructs' validity and reliability, we conducted single-level confirmatory factor analyses (CFAs) before proceeding with the multilevel analyses. The CFA and reliability results demonstrated strong model fit for team innovation (CFI =.99, TLI =.97, SRMR =.01, α =.95, ω =.95) and cognitive activation (CFI =.98, TLI =.93, SRMR =.03, α =.89, ω =.90). Furthermore, reliability analyses confirmed high internal consistency for instructional leadership (α =.82, ω =.81) and distributed leadership (α =.81, ω =.80), both of which were measured using three items. Therefore, these leadership constructs' models were identified and indicated perfect fit with the data. These findings demonstrate that the scales used in this study comprise satisfactory psychometric properties, justifying their inclusion in subsequent multilevel analyses.

3.2. ML-SEM results

Table 2 reports the sample size, model fit indices and path coefficients for the tested SEMs across seven countries. The ICC-1 values for the dependent variable (i.e., achievement) ranged between.13 and.60, with an average of.34. The model fit indicators suggest an excellent model fit for most countries, with RMSEA values below 0.02, indicating minimal approximation error. CFI and TLI values exceeded 0.90 for all countries except Malta (CFI =.78, TLI =.71), where model fit appeared poor. SRMR values confirmed acceptable fit, with $SRMR_{L1}$ = 0.00 at the student level and $SRMR_{L2}$ ranging between 0.05 and 0.12 at the school level. Examining the path coefficients, instructional leadership (IT) was related significantly to team innovation in Colombia (β =.33, *SE* =.17, *p* = .046), but this association is not statistically significant in other countries. Similarly, distributed leadership (DT) data indicate a significant association with team innovation in the Czech Republic data (β=.36, *SE*=.14 p = .009). The relation between team innovation and cognitive activation (TC) was statistically significant in the Czech Republic (β=.32, *SE*=.14, p = .021), Türkiye (β=.37, *SE*=.16, p = .024) and Georgia (β=.40, *SE*=.12, p = .001), suggesting that higher team innovation is associated with higher cognitive activation in these countries. Overall, while the ML-SEM model demonstrates a strong fit for most countries, variations exist in the relations between leadership, team innovation and cognitive activation across different educational systems. Meta-analyses of the regression coefficients revealed that the pooled coefficient for TC was positive and significant (pooled β=.29, *SE*=.06, p < .001), whereas it was insignificant for IT (pooled β=.11, *SE*=.06, p = .075) and DT (pooled β=.09, *SE*=.06, p = .116).

Table 3 also reports achievement-domain-specific path coefficients for reading across seven countries. Similar to the results for mathematics, most path estimates were non-significant, indicating weak or inconsistent relations between leadership, innovation, cognitive activation and reading achievement. The only statistically significant relation was observed in the Czech Republic (β=.30, SE =.07, p < .001), suggesting a positive association between cognitive activation and reading achievement. Other estimates, including ITCA (instructional leadership → cognitive activation) and DTCA (distributed leadership → cognitive activation), were small and non-significant across countries, with confidence intervals covering zero. These results indicate that the role of leadership, innovation and cognitive activation in reading achievement is not strongly established and varies across educational contexts.

Achievement-domain-specific path coefficients for science across seven countries also are reported in Table 3. As observed in the previous tables for mathematics and reading, most path estimates remain non-significant, indicating weak or inconsistent relations between leadership, innovation, cognitive activation and science achievement. The only statistically significant relation was found in the Czech Republic (β=.27, SE =.07, p < .001), suggesting a positive association between cognitive activation and science achievement. Meta-analyses of the regression coefficients revealed that the pooled coefficient for CA was positive and significant (pooled β=.09, *SE*=.03, p = .002), whereas all other estimates, including ITCA (instructional leadership → cognitive activation) and DTCA (distributed leadership → cognitive activation), were small and non-significant across countries, with confidence intervals covering zero. These results suggest that the associations between leadership, team innovation and cognitive activation with science achievement are weak or vary across different educational contexts.

3.3. MLM decision tree results

Fig. 2 illustrates the association of instructional leadership, distributed leadership, team innovation and cognitive activation with student achievement. Overall average achievement was zero, while average values for predictor variables were instructional leadership = 2.82 (SD = 0.59), distributed leadership = 3.22 (SD = 0.49), team innovation = 2.94 (SD = 0.30) and cognitive activation = 2.52 (SD = 0.35).

The third node (Node 3) registered the highest achievement score (0.303), approximately 0.3 standard deviations above the average across the seven countries. This node is characterised by schools in which instructional leadership is ≤ 2.67, and team innovation is ≤ 2.81, suggesting that lower instructional leadership and team innovation levels are associated with relatively higher achievement in this subgroup.

Conversely, the 19th node (Node 19) had the lowest achievement score (-0.59), which was the lowest among all subgroups. This node is formed when instructional leadership > 2.67, team innovation > 2.81, distributed leadership > 3.67 and cognitive activation > 2.55. This suggests that in schools with higher leadership, team innovation, distributed leadership and cognitive activation, achievement tends to be lower. These results highlight the nonlinear nature of leadership effects on achievement, indicating that certain combinations of leadership,

Table 2
Sample size, model fit and path coefficients that are common across achievement domains.

Country	n	J	RMSEA	CFI	TLI	$SRMR_{L1}$	$SRMR_{L2}$	IT		DT		TC	
								Est	SE	Est	SE	Est	SE
AUS	2563	131	.02	.92	.90	.00	.07	.08	.11	.11	.12	.40*	.22
COL	4720	152	.01	.99	.99	.00	.06	.33*	.16	.08	.13	.11	.32
CZE	4290	182	.01	.98	.97	.00	.05	−.01	.16	.36*	.14	.32*	.14
DNK	2320	102	.01	.96	.95	.00	.07	.31	.25	−.19	.34	.17	.15
GEO	3292	134	.01	.95	.94	.00	.06	.16	.20	.00	.13	.40*	.12
MLT	3354	49	.02	.78	.71	.00	.12	.14	.21	−.13	.18	−.08	.22
TUR	5573	143	.01	.94	.92	.00	.06	−.06	.16	.08	.15	.37*	.16
Pooled								.11	.06	.09	.06	.29*	.06

Note. n = Student Sample Size, J = Number of schools, RMSEA = root mean square error of approximation, CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, SRMR = standardised root mean square Residual, Est = Estimate, SE = Standard Error, IT = Instructional Leadership to Team Innovativeness, DT = Distributed Leadership to Team Innovativeness, TC = Team Innovativeness to Cognitive Activation. * p < .05

Table 3
Path coefficients for achievement.

Country	Ach	IA Est	IA SE	DA Est	DA SE	TA Est	TA SE	CA Est	CA SE	ITCA Est	ITCA 95 % CI	DTCA Est	DTCA 95 % CI
AUS	**Mathematics**	−.09	.06	.05	.07	−.02	.06	.13	.10	.004	[−.008,.017]	.006	[−.008,.020]
COL		.03	.08	−.07	.09	.08	.11	−.21	.17	−.008	[−.072,.055]	−.002	[−.022,.018]
CZE		−.11	.07	.06	.08	−.06	.07	.16	.10	.000	[−.016,.015]	.018	[−.009,.044]
DNK		−.04	.09	−.03	.10	−.02	.07	−.02	.09	−.001	[−.011,.008]	.001	[−.006,.008]
GEO		.14*	.07	−.07	.07	−.07	.06	.03	.06	.002	[−.008,.011]	.000	[−.003,.003]
MLT		−.13	.10	.12	.11	−.03	.05	−.09	.30	.001	[−.005,.006]	−.001	[−.006,.004]
TUR		−.05	.09	−.11	.10	.06	.08	.16	.13	−.004	[−.022,.015]	.005	[−.015,.025]
AUS	**Reading**	−.12	.09	.05	.08	.02	.06	.13	.09	.004	[−.007,.016]	.006	[−.008,.020]
COL		.02	.08	−.01	.08	.07	.11	−.20	.17	−.007	[−.062,.049]	−.002	[−.019,.016]
CZE		−.08	.06	.00	.07	−.06	.06	.30*	.07	.004	[−.010,.017]	.012	[−.008,.031]
DNK		−.09	.08	.02	.09	−.02	.06	−.01	.10	−.001	[−.011,.009]	.000	[−.006,.007]
GEO		.10	.06	−.08	.07	−.06	.05	.06	.06	.004	[−.008,.016]	.000	[−.006,.006]
MLT		−.12	.11	.13	.10	.03	.04	−.02	.26	.000	[−.006,.006]	.000	[−.006,.005]
TUR		−.08	.09	−.09	.11	.05	.08	.19	.14	−.004	[−.025,.017]	.005	[−.017,.028]
AUS	**Science**	−.10	.07	.05	.07	−.01	.06	.12	.10	.004	[−.008,.016]	.006	[−.008,.019]
COL		.04	.08	−.00	.08	.09	.09	−.14	.20	−.003	[−.029,.023]	.001	[−.009,.007]
CZE		−.08	.06	.01	.07	−.05	.06	.27*	.07	.003	[−.009,.016]	.011	[−.008,.029]
DNK		−.09	.09	.00	.11	−.01	.07	.01	.09	.000	[−.008,.009]	.000	[−.006,.005]
GEO		.11	.07	−.09	.07	−.05	.06	.05	.07	.003	[−.008,.014]	.000	[−.005,.005]
MLT		−.14	.10	.12	.09	−.00	.05	−.06	.24	.001	[−.005,.006]	−.001	[−.005,.004]
TUR		−.06	.09	−.11	.11	.02	.08	.18	.13	−.004	[−.024,.016]	.005	[−.016,.027]
Pooled[a]		−.04	.02	−.01	.02	−.01	.01	.09*	.03	.001	.001	.004	.001

Note. Est = Estimate, SE = Standard Error, CI = Confidence Interval, IA = Instructional Leadership Direct Effect, DA = Distributed Leadership Direct Effect, TA = Team Innovativeness Direct Effect, CA = Cognitive Activation Direct Effect, ITCA = Indirect Effect of Instructional Leadership on Cognitive Activation via Team Innovativeness; DTCA = Indirect Effect of Distributed Leadership on Cognitive Activation via Team Innovativeness. Estimates are standardised coefficients. * $p < .05$. Pooled[a] across three achievement domains and seven countries.

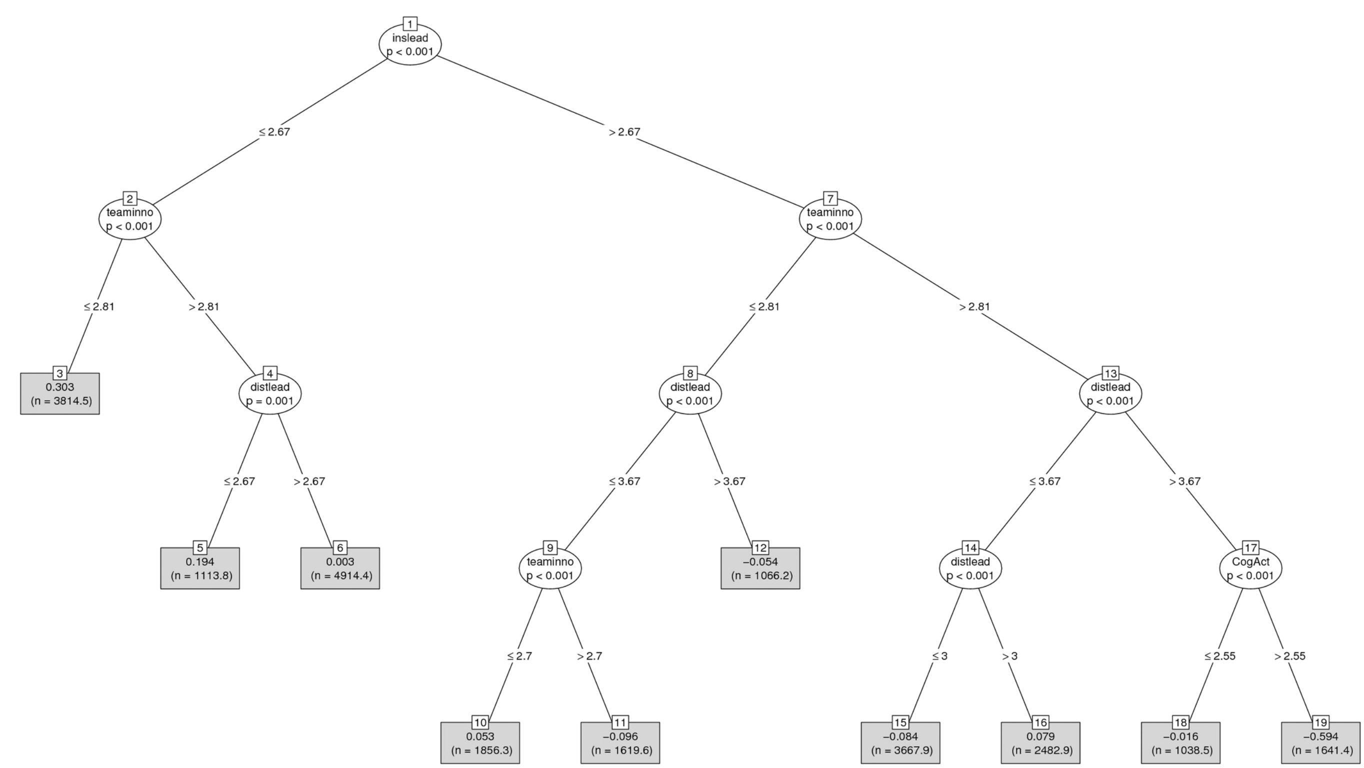

Fig. 2. *MLM tree with all observations. Note.* The multilevel decision tree (MLM tree) indicates splits by instructional leadership, team innovativeness, distributed leadership and cognitive activation. Terminal nodes display sample sizes with decimals due to weighting. All splits are statistically significant at $p < .001$.

innovation and instructional practices may yield different outcomes, depending on their interactions within school environments.

Fig. 3 reports the decision tree results using the randomly selected 50 % of the data. The tree structure reveals distinct subgroups based on distributed leadership, instructional leadership, team innovation and cognitive activation. The key finding is that the highest achievement was observed in the subgroup in which all four variables—distributed leadership, instructional leadership, team innovation and cognitive activation—were low. This suggests that in certain contexts, lower levels of leadership and instructional collaboration may be associated with higher student achievement, emphasising these relationships' complexity. Conversely, subgroups with higher levels of these factors

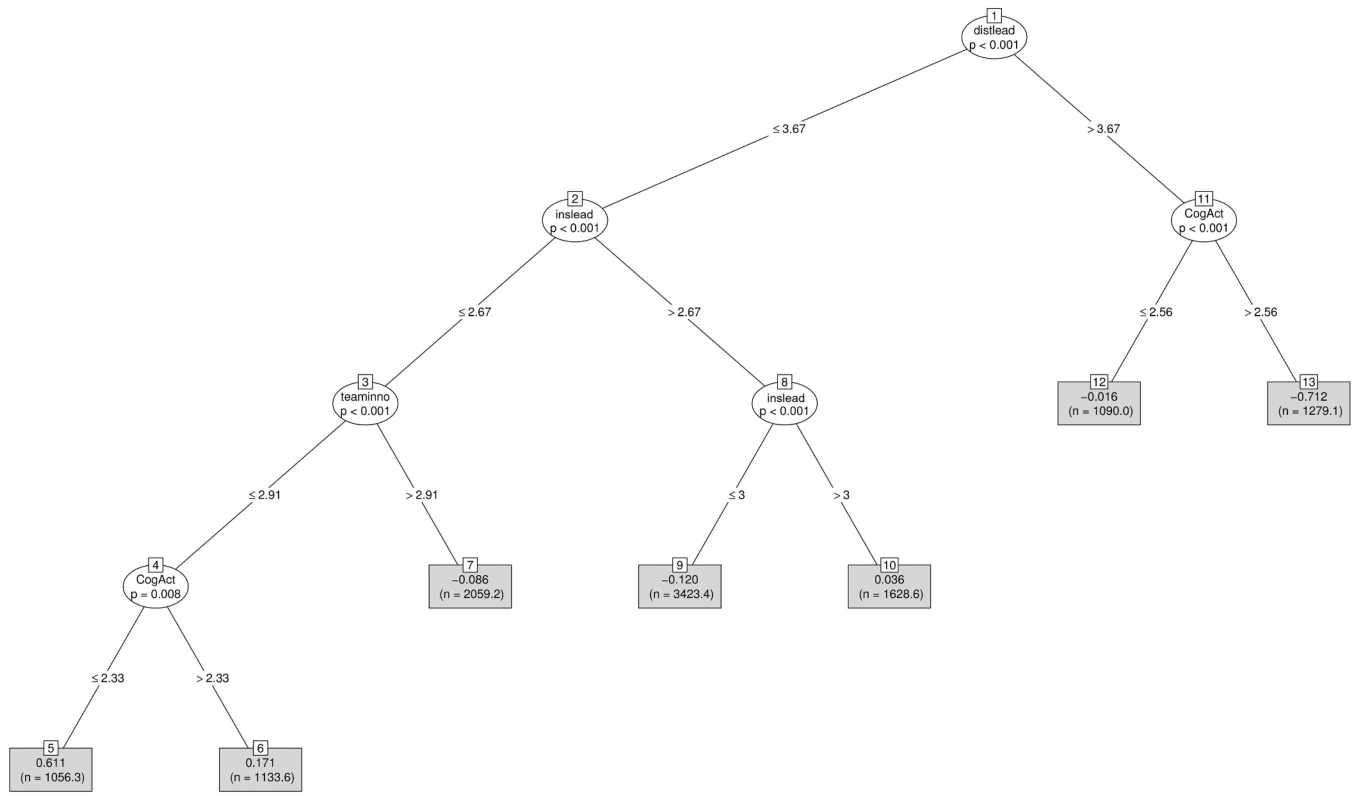

Fig. 3. MLM tree with the randomly selected 50 %. *Note.* The multilevel decision tree (MLM tree) indicates splits by instructional leadership, team innovativeness, distributed leadership and cognitive activation. Terminal nodes display sample sizes with decimals due to weighting. All splits are statistically significant at $p < .001$.

tend to have lower achievement, indicating potential nonlinear or context-dependent effects from leadership and instructional practices on student outcomes.

That said, for some nodes, school achievement averages within a country remained similar, such as in Node 13 (high distributed leadership, high cognitive activation), in which differences across countries were minimal. However, for other nodes, substantial variation in school achievement averages was observed across countries, as seen in Node 5

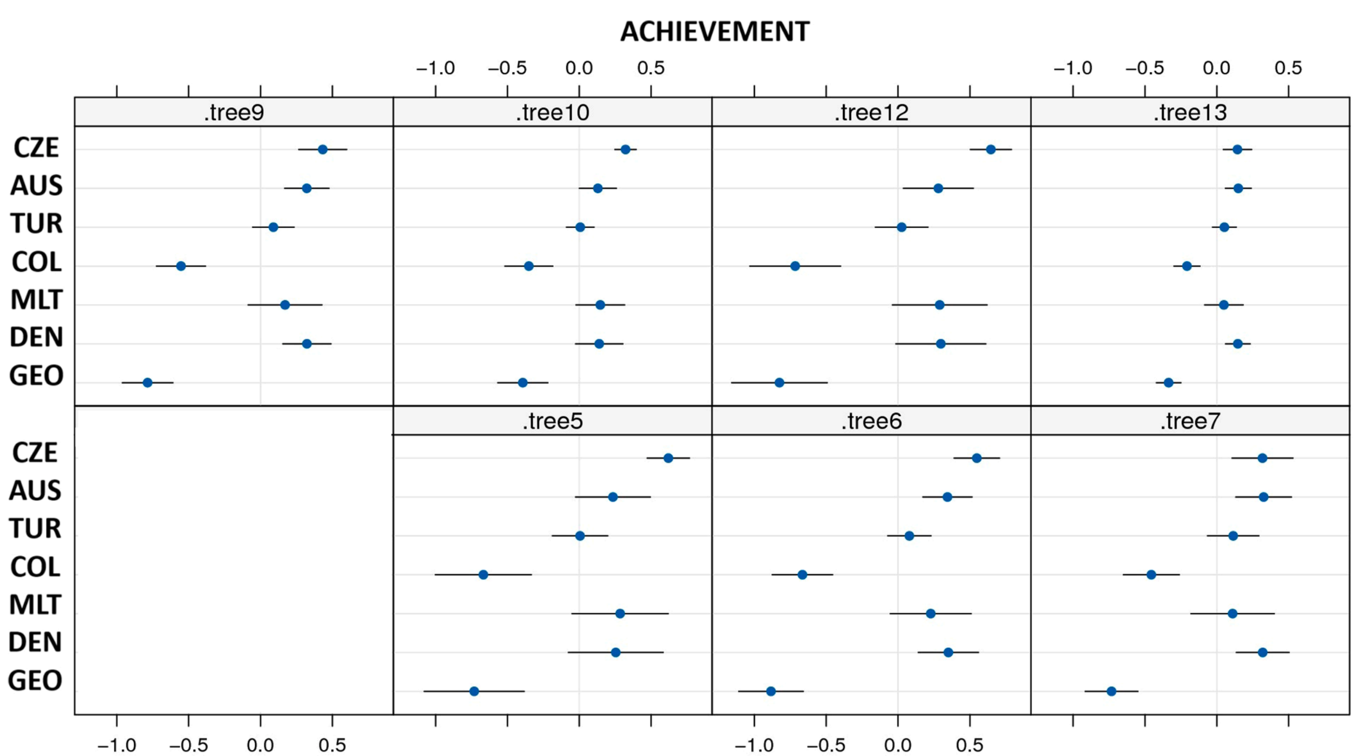

Fig. 4. Node means between countries using the second half of the data and the random slopes model.

(where distributed leadership, instructional leadership, team innovation and cognitive activation were all low). This suggests that while country-level trends remain stable, certain school-level conditions contribute to notable deviations in achievement within specific contexts. With the randomly selected 50 % of the data, the MLM trees resulted in an intraclass correlation coefficient of 0.228 due to variation among countries and a pseudo-R^2 value of 0.056 for the four predictors in the tree.

Fig. 4 indicates variability in the node means between countries using the remaining 50 % of the data and the random slopes model. Compared with a model in which node means are not allowed to vary across countries, the results indicate variation beyond the overall country intercepts (the seven intercepts). However, when evaluated visually, these variations do not alter the country intercepts' overall pattern substantially, as the blue dots generally are aligned across nodes, maintaining consistency in the observed relationships.

Our country-specific SEM results, in which relations are assumed to be linear, did not indicate indirect effects from leadership on student achievement. When applying a single MLM tree model, the results (reported in Fig. 2) highlight nonlinear patterns, offering an alternative perspective on the leadership-achievement relationship.

4. Discussion

This study examined the relations between instructional and distributed leadership, team innovation, cognitive activation and student achievement using ML-SEM and MLM trees. By leveraging PISA-TALIS 2018 linked data across seven countries, we examined both linear and nonlinear associations to better understand how leadership and instructional practices shape student outcomes.

4.1. Summary of key findings

The ML-SEM results suggest that leadership does not exert a significant indirect effect on student achievement via team innovation and cognitive activation. These findings contradict those from extant studies that have emphasised leadership's role in fostering instructional practices that improve student learning outcomes (Hallinger, 2011; Leithwood & Jantzi, 2006; Leithwood et al., 2020). The absence of indirect effects in this study suggests that traditional linear modelling approaches may not fully capture these relationships' complexity.

However, the MLM tree results reveal substantial heterogeneity in how leadership, team innovation and cognitive activation interact across schools. Notably, the highest achievement levels were observed in contexts in which leadership, team innovation and cognitive activation were low, challenging the assumption in the literature that stronger leadership and instructional collaboration necessarily enhance student learning. Moreover, country-specific MLM trees indicated significant variation in the formation of achievement subgroups, making it difficult to derive universal leadership effects. These findings underscore the context-dependent nature of leadership's role in education and align with recent research advocating for nonlinear approaches in educational studies (OECD, 2020).

4.2. Interpretation in relation to the literature

There is a longstanding critique that research in educational leadership has used overly simplistic linear effect models which do not capture the complexity of the relationships involved and the context of schools (e.g. Muijs, 2011). However, the assumption has usually been that using such models would lead to larger effect sizes than the small to moderate ones we typically find. These results suggest that this may not, in fact, be the case, but that more accurate models may instead problematise many of the relationships we have taken for granted, and which have been found in simpler models, such as the relationship between leadership, teacher behaviours and student outcomes (Papadakis et al., 2024). There are of course earlier studies which have raised some questions about these relationships (Charalambous & Praetorius, 2025). The impact of cognitive activation is, for example, mixed, as mentioned above, and the causality and contextuality of leadership effects have both been questioned (Tan, 2018; Noman & Gurr, 2020; Belchetz & Leithwood, 2007; Levačić, 2005; Martin et al., 2021).

These critiques have often been paid lip-service in the field, but have not greatly shaken the existing status quo view. However, these findings suggest that leadership's role in student achievement is more complex than traditionally assumed, providing empirical support for the view that schools are complex and nonlinear institutions (Morrison, 2012). As a consequence, while leadership often is associated with positive school outcomes, its impact may be contingent upon institutional, cultural and policy contexts. Extant studies using linear modelling frameworks have identified leadership as a driver of innovation (Pietsch & Mah, 2025) and instructional quality (Tschannen-Moran, 2014). However, our results may suggest that leadership's role may not be uniform across different education systems and that nonlinear patterns may better explain its effects or may provide further support to critiques of the current models of effective leadership and teaching employed in much of the literature.

4.3. Limitations

This study has several limitations that should be acknowledged. First, the cross-country variability in MLM tree results highlights the challenge of comparing leadership effects across different educational systems. While decision tree approaches allow for data-driven subgroup identification, the lack of formal statistical procedures for synthesising subgroup findings (e.g., meta-analytic techniques for node-level interpretations) limits the ability to draw generalizable conclusions. Future research should examine hybrid approaches, combining decision trees with meta-analytic frameworks to improve cross-country comparability.

Second, the ML-SEM approach assumes linearity, which may have restricted the ability to detect interaction effects or complex mediation pathways. While MLM trees provide a more flexible, data-driven alternative, they are sensitive to sample size variations and data distributions. Future studies could integrate machine learning approaches with causal inference techniques to refine identification of leadership effects in multilevel educational contexts.

An important limitation lies in the operationalisation of the distributed leadership variable in the PISA-TALIS dataset. The three statements primarily measure shared leadership practices, which is an incomplete definition of distributed leadership, which also encompasses organisational learning, developing leadership capacity and empowerment. The fact that the variables used here do not fully cover the concept may mean that what we are in fact measuring here is a more limited construct, which therefore limits the extent to which we can generalize to distributed leadership practices in schools. The same is true for the operationalisation of instructional leadership in this study. Here we have focused on variables related to leadership that fosters collaboration and improving teachers practice. This again is a limited conceptualisation of instructional leadership, which omits important aspects such as a focus on curriculum and assessment. The same limitations on generalisability therefore apply.

However, particularly our MLM analyses demonstrate that instructional and distributed leadership interact, thereby supporting the concept of shared instructional leadership (Printy & Marks, 2006). However, in contrast to the assumptions made in conventional literature (Marks & Printy, 2003; Printy et al., 2009; Zhan et al., 2020) no universal, generalisable effects of such integrative leadership can be demonstrated. Instead, the observed outcomes are the result of a multifaceted interaction with other factors, including team innovation and cognitive activation. In order to analyse the effectiveness of such leadership, it is thus important to recognise that a linear 'more is better' approach is not applicable in this regard. Instead, specific configurations must be considered in specific contexts (Bellibaş et al., 2024). Consequently, our findings suggest that leadership and instructional practices interact in ways not easily captured by traditional models. Future research should examine contextual moderators—such as policy environments, school autonomy and cultural factors—to further disentangle the mechanisms through which leadership influences student achievement.

5. Conclusion

Overall, this study contributes to the literature by highlighting linear modelling approaches' limitations in capturing the complex, nonlinear dynamics between leadership, instructional practices and student achievement. Our findings call into question some of the 'givens' in educational (leadership) research, and call for a shift towards nonlinear and context-sensitive methodologies to better understand how educational leadership operates in diverse contexts. By employing MLM trees alongside traditional ML-SEM models, this study provides a more nuanced perspective on leadership's role in shaping student learning outcomes. It is clear we cannot assume that more accurate modelling will lead to *higher* effect sizes. Instead, what we are seeing are highly complex and at times counter-intuitive patterns. Of course, more evidence is required before we jettison existing models of the relationship between leadership, teaching practices and attainment, especially as the dataset here has limitations in terms of operationalisation of concepts as well as its cross-sectional nature. Nevertheless, there is clearly a need for replication of complex statistical modelling with other datasets, so we can further test the robustness of the largely accepted causal model in which leadership is assumed to affect teaching, which in turn affects attainment. These findings therefore also have practical implications. A lot of existing principal and teacher evaluation systems, for example, make assumptions about the impact of instructional and distributive leadership approaches, innovation and teaching practices, and incorporate these into evaluation instruments (e.g. Darling-Hammond, 2015; Donaldson et al., 2021), which may be called into question if these findings are replicated elsewhere. The same would be true of many approaches to the professional development of school leaders, making these findings potentially far-reaching.

CRediT authorship contribution statement

Aydın Burak: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Marjolein Fokkema:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis. **Eryılmaz Nurullah:** Writing – review & editing, Writing – original draft, Validation. **Daniel Muijs:** Writing – review & editing, Writing – original draft, Validation, Investigation, Conceptualization. **Ronny Scherer:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Conceptualization. **Marcus Pietsch:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Consent for publication

All authors have reviewed and approved the final manuscript and consent to its publication in Studies in Educational Evaluation

Funding information

Marcus Pietsch is supported by the German Research Foundation (DFG) through a DFG Heisenberg Professorship (451458391) and project funding (531146345).

Declaration of Competing Interest

We do not have any competing interests.

Acknowledgments

None.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.stueduc.2025.101521.

Data Availability

Data and materials are available at: https://osf.io/uf3ns/

References

Anderson, N., & West, M. A. (1998). Measuring climate for work group innovation: Development and validation of the team climate inventory (TCI). *Journal of Organisational Behaviour, 19*, 235–258. https://doi.org/10.1002/(ISSN)1099-1379

Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership*. Psychology press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133–180. https://doi.org/10.3102/0002831209345157

Belchetz, D., & Leithwood, K. (2007). Successful leadership: Does context matter and if so, how?. *Successful principal leadership in times of change: An international perspective* (pp. 117–138). Netherlands: Dordrecht: Springer.

Bellibaş, M.Ş., Polatcan, M., & Akyürek, M.İ. (2024). Principal leadership typologies and their relationship with teacher self-efficacy and commitment: A latent profile mediation analysis. *Educational Management Administration Leadership, 52*(6), 1497–1518.

Bellibaş, M.Ş., Polatcan, M., & Alzouebi, K. (2025). Instructional leadership and student achievement across UAE schools: Mediating role of professional development and cognitive activation in teaching. *Educational Management Administration Leadership*. https://doi.org/10.1177/17411432241305702

Bell, C. A., Dobbelaer, M. J., Klette, K., & Visscher, A. (2019). Qualities of classroom observation systems. *School Effectiveness and School Improvement, 30*(1), 3–29.

Bellens, K., Van Damme, J., Van Den Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: Catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in flanders (Belgium), Germany, and Norway. *Large-Scale Assessments in Education, 7*, 1–27.

Blömeke, S., Nilsen, T., & Scherer, R. (2021). School innovativeness is associated with enhanced teacher collaboration, innovative classroom practices, and job satisfaction. *Journal of Educational Psychology, 113*(8), 1645.

Burić, I., & Kim, L. E. (2020). Teacher self-efficacy, instructional quality, and student motivational beliefs: An analysis using multilevel structural equation modeling. *Learning and Instruction, 66*, Article 101302.

Buyukgoze, H., Caliskan, O., & Gümüş, S. (2022). Linking distributed leadership with collective teacher innovativeness: The mediating roles of job satisfaction and professional collaboration. *Educational Management Administration Leadership, 52*(6), 1388–1409. https://doi.org/10.1177/17411432221130879

Cao, J., Huang, Y., Shao, X., & Zhong, Y. (2025). Driving innovation: Exploring the effect of distributed leadership on collective teacher innovativeness in Chinese schools. *The Asia-Pacific Education Researcher, 34*(1), 433–446.

Caro, D. H., Lenkeit, J., & Kyriakides, L. (2016). Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012. *Studies in Educational Evaluation, 49*, 30–41.

Charalambous, C. Y., & Praetorius, A. K. (2025). Are we doing okay? Making the case for critically examining small and inconsistent effects in research on teaching quality. *School Effectiveness and School Improvement, 36*(2), 153–163.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. Routledge.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. AsCD.

Darling-Hammond, L. (2015). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. Teachers College Press.

Dedering, K., & Pietsch, M. (2025). School leader trust and collective teacher innovativeness: On individual and organisational ambidexterity's mediating role. *Educational Review, 77*(2), 351–380.

Donaldson, M., Mavrogordato, M., Dougherty, S. M., Ghanem, R. A., & Youngs, P. (2021). *Principal evaluation under the elementary and secondary every student succeeds act: A Comprehensive policy review*.

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods, 50*, 2016–2034.

Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2021). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research, 31*(3), 329–341. https://doi.org/10.1080/10503307.2020.1785037

Fokkema, M., & Zeileis, A. (2024). *Generalized linear mixed model trees (Version 0.2-6) [R package glmertree]*. CRAN. ⟨https://CRAN.R-project.org/package=glmertree⟩.

Förtsch, C., Werner, S., von Kotzebue, L., & Neuhaus, B. J. (2016). Effects of biology teachers' professional knowledge and cognitive activation on students' achievement. *International Journal of Science Education, 38*(17), 2642–2666.

Genç, M., & Çolakoğlu, Ö. M. (2021). Modeling the effects of instructional quality on mathematical literacy performance from the students' perspective: PISA 2012 Turkey sample. *Education Science/Egitim ve Bilim, 46*(206).

Goff, P., & Finch, M. (2016). Challenges and opportunities for education leadership scholarship: A methodological critique. In A. J. Bowers, A. R. Shoho, & B. G. Barnett (Eds.), *Challenges and opportunities of educational leadership research and practice: The state of the field and its multiple futures* (pp. 119–146). Information Age.

Hajjem, A., Larocque, D., & Bellavance, F. (2017). Generalized mixed-effects regression trees. *Statistics Probability Letters, 126*, 114–118. https://doi.org/10.1016/j.spl.2017.02.033

Halász, G. (2018). Measuring innovation in education: The outcomes of a national education sector innovation survey. *European Journal of Education, 53*(4), 557–573.

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., & 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

Harris, A. (2009). Distributed leadership and knowledge creation. *Distributed Leadership according to the Evidence (pp 271-284) Routledge*.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education*. https://doi.org/10.1002/rev3.3310

Hojeij, Z. (2024). Educational leadership's role in fostering innovation and entrepreneurship in education: A narrative literature review. *Social Sciences Humanities Open, 10*, Article 101173.

Hothorn, T., & Zeileis, A. (2015). Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research, 16*(1), 3905–3909.

Huang, L., Zhao, D., & Zhou, S. (2024). Examining principal instructional leadership effects on the rural-urban instructional quality gap in China: The mechanism of teacher professional learning. *Studies in Educational Evaluation, 81*, Article 101346.

Hu, L. T., & 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

Hu, J., Peng, Y., & Ma, H. (2022). Examining the contextual factors of science effectiveness: A machine learning-based approach. *School Effectiveness and School Improvement, 33*(1), 21–50. https://doi.org/10.1080/09243453.2021.1929346

Karadağ, E., & Sertel, G. (2025). The effect of educational leadership on students' achievement: A cross-cultural meta-analysis research on studies between 2006 and 2024. *Educational Management Administration Leadership*. https://doi.org/10.1177/17411432251327645

Klieme, E. (2013). The role of instructional quality in explaining international differences in mathematics and science achievement. In In. M. Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 215–240). Springer.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German classrooms. In.T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Leavitt, K., Schabram, K., Hariharan, P., & Barnes, C. M. (2021). Ghost in the machine: On organizational theory in the age of machine learning. *Academy of Management Review, 46*(4), 750–777. https://doi.org/10.5465/amr.2019.0247

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., Sun, J., & Schumacker, R. (2020). How school leadership influences student learning: a test of "The four paths model". *Educational Administration Quarterly, 56*(4), 570–599.

Levačić, R. (2005). Educational leadership as a causal factor: Methodological issues in research on leadership 'effects'. *Educational Management Administration Leadership, 33*(2), 197–210.

Liu, S., Yin, H., Wang, Y., & Lu, J. (2024). Teacher innovation: Conceptualizations, methodologies, and theoretical framework. *Teaching and Teacher Education, 145*, Article 104611.

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.

Martin, R., Hughes, D. J., Epitropaki, O., & Thomas, G. (2021). In pursuit of causality in leadership training research: A review and pragmatic recommendations. *The Leadership Quarterly, 32*(5), Article 101375.

Moolenaar, N. M., Daly, A. J., & Sleegers, P. J. (2010). Occupying the principal position: Examining relationships between transformational leadership, social network position, and schools' innovative climate. *Educational Administration Quarterly, 46*(5), 623–670.

Morrison, K. (2012). *School leadership and complexity theory*. Routledge.

Muijs, D. (2011). Researching leadership. Towards a new paradigm. In T. Townsend, & J. MacBeath (Eds.), *International Handbook of Leadership for Learning*. Dordrecht: Springer.

Mumford, M. D. (2000). Managing creative people: Strategies and tactics for innovation. *Human Resource Management Review, 10*, 313–351. https://doi.org/10.1016/S1053/4822(99)00043-1

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. *Handbook of advanced multilevel analysis* (pp. 15–40). Routledge.

Muthén, L. K., & Muthén, B. O. (2021). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Neugebauer, P., & Prediger, S. (2023). Quality of teaching practices for all students: Multilevel analysis of language-responsive teaching for robust understanding. *International Journal of Science and Mathematics Education, 21*(3), 811–834.

Newman, A., Round, H., Wang, S., & Mount, M. (2020). Innovation climate: A systematic review of the literature and agenda for future research. *Journal of Occupational and Organizational Psychology, 93*(1), 73–109.

Noman, M., & Gurr, D. (2020). Contextual leadership and culture in education. *Oxford Research Encyclopedia of Education*.

OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

OECD. (2020). *TALIS 2018 technical report*. OECD Publishing. https://doi.org/10.1787/799337c2-en

OECD. (2021). *Positive, high-achieving students?: What schools and teachers can do, TALIS*. OECD Publishing. https://doi.org/10.1787/3b9551db-en

O'Shea, C. (2021). Distributed leadership and innovative teaching practices. *International Journal of Educational Research Open, 2*, Article 100088.

Pan, H. L., Lin, Y. C., & Chung, C. H. (2024). Teacher collaboration, school innovativeness and innovative teaching in Taiwan: Evidence from TALIS. *International Journal of Educational Research, 127*, Article 102383.

Papadakis, S., Kanadlı, S., Kardas, A., Tülübaş, T., Karakose, T., & Polat, H. (2024). Investigating the relationship between leadership for learning and student achievement through the mediation of teacher performance: A meta-analytic structural equation modeling (MASEM) approach. *Education Sciences, 14*(12), 1320. https://doi.org/10.3390/educsci14121320

Praetorius, A. K., Jentsch, A., Luoto, J., Keller, S. D., & Fauth, B. (2025). Context in research on teaching quality: Bringing a complex idea into the spotlight. *School Effectiveness and School Improvement, 36*(2), 263–296.

Praetorius, A. K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: The German framework of three basic dimensions. *ZDM: International Journal on Mathematics Education, 50*, 407–426.

Pietsch, M., & Tulowitzki, P. (2017). Disentangling school leadership and its ties to instructional practices – An empirical comparison of various leadership styles. *School Effectiveness and School Improvement, 28*(4), 629–649. https://doi.org/10.1080/09243453.2017.1363787

Pietsch, M., Cramer, C., Brown, C., Aydin, B., & Witthöft, J. (2024). Open innovation in schools: A new imperative for organising innovation in education? *Technology, Knowledge and Learning, 29*(2), 1051–1077.

Pietsch, M., Aydin, B., & Gümüş, S. (2025). Putting the instructional leadership-student achievement relation in context: A meta-analytical big data study across cultures and time. *Educational Evaluation and Policy Analysis, 47*(1), 29–64.

Pietsch, M., & Mah, D. K. (2025). Leading the AI transformation in schools: It starts with a digital mindset. *Educational Technology Research and Development, 73*(2), 1043–1069. https://doi.org/10.1007/s11423-024-10439-w

Printy, S. M., & Marks, H. M. (2006). Shared leadership for teacher and student learning. *Theory into Practice, 45*, 125–132.

Printy, S. M., Marks, H. M., & Bowers, A. J. (2009). Integrated leadership: How principals and teachers share transformational and instructional influence. *Journal of School Leadership, 19*(5), 504–532.

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

Sebastian, J., Allensworth, E., & Huang, H. (2016). The role of teacher leadership in how principals influence classroom instruction and student learning. *American Journal of Education, 123*(1), 69–108.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Sela, R. J., & Simonoff, J. S. (2012). RE-EM trees: A data-mining approach for longitudinal and clustered data. *Machine Learning, 86*(2), 169–207. https://doi.org/10.1007/s10994-011-5258-3

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it. *Journal of Research in Innovative Teaching Learning, 10*(1), 4–33.

Shen, J., Wu, H., Reeves, P., Zheng, Y., Ryan, L., & Anderson, D. (2020). The association between teacher leadership and student achievement: A meta-analysis. *Educational Research Review, 31*, Article 100357.

Siegel, S. M., & Kaemmerer, W. F. (1978). Measuring the perceived support for innovation in organisations. *Journal of Applied Psychology, 63*, 553–562. https://doi.org/10.1037/0021-9010.63.5.553

Sleegers, P., den Brok, P., Verbiest, E., Moolenaar, N., & Daly, A. (2014). Toward conceptual clarity: A multidimensional, multilevel model of professional learning communities in Dutch elementary schools. *Elementary School Journal, 115*(1), 100–124. https://doi.org/10.1086/676902

Spillane, James P. (2006). *Distributed leadership* (1st ed., pp. 11–12). San Francisco: Jossey-Bass.

Sun, J., & Leithwood, K. (2015). Direction-setting school leadership practices: A meta-analytical review of evidence about their influence. *School Effectiveness and School Improvement, 26*(4), 499–523.

Tan, C. Y. (2018). Examining school leadership effects on student achievement: The role of contextual challenges and constraints. *Cambridge Journal of Education, 48*(1), 21–45.

Tan, C. Y., Dimmock, C., & Walker, A. (2021). How school leadership practices relate to student outcomes: Insights from a three-level meta-analysis. *Educational Management Administration Leadership, 52*(1), 6–27. https://doi.org/10.1177/17411432211061445 (Original work published 2024).

Teig, N., Scherer, R., & Nilsen, T. (2018). More isn't always better: The curvilinear relationship between inquiry-based teaching and student achievement in science. *Learning and Instruction, 56*, 20–29. https://doi.org/10.1016/j.learninstruc.2018.02.006

Thurlings, M., Evers, A. T., & Vermeulen, M. (2015). Toward a model of explaining teachers' innovative behavior: A literature review. *Review of Educational Research, 85*(3), 430–471.

Tourón, J., Navarro-Asencio, E., Lizasoain, L., López-González, E., & García-San Pedro, M. J. (2019). How teachers' practices and students' attitudes towards technology affect mathematics achievement: Results and insights from PISA 2012. *Research Papers in Education, 34*(3), 263–275.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*, 769. https://doi.org/10.3389/fpsyg.2016.00769

Tschannen-Moran, M. (2014). *Trust matters: Leadership for successful schools* (*2nd ed.*). Jossey-Bass.

Vermeulen, M., Kreijns, K., & Evers, A. T. (2022). Transformational leadership, leader-member exchange and school learning climate: Impact on teachers' innovative behaviour in the Netherlands. *Educational Management Administration Leadership, 50*(3), 491–510.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research, 78*(3), 516–551.

Wu, H., & Shen, J. (2022). The association between principal leadership and student achievement: A multivariate meta-meta-analysis. *Educational Research Review, 35*, Article 100423.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics, 17*(2), 492–514. https://doi.org/10.1198/106186008X319331

Zhan, X., Anthony, A. B., Goddard, R., & Beard, K. S. (2020). Development, factor structure, and reliability of the shared instructional leadership scale in public secondary schools. *Educational Management Administration Leadership, 51*(1), 1–20. https://doi.org/10.1177/1741143220963103

Zhang, D., Wang, C., & Yang, Y. (2021). The association between cognitive activation and mathematics achievement: A multiple mediation model. *Educational Psychology, 41*(6), 695–711.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's alpha, Revelle's beta, McDonald's omega: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7