



Longitudinal effects of a professional development program on principals' leadership practices: results of a baseline latent change score model

André Meyer & Dirk Richter

To cite this article: André Meyer & Dirk Richter (08 Oct 2024): Longitudinal effects of a professional development program on principals' leadership practices: results of a baseline latent change score model, *Professional Development in Education*, DOI: [10.1080/19415257.2024.2413119](https://doi.org/10.1080/19415257.2024.2413119)

To link to this article: <https://doi.org/10.1080/19415257.2024.2413119>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 08 Oct 2024.



[Submit your article to this journal](#)





[View related articles](#)



[View Crossmark data](#)

Longitudinal effects of a professional development program on principals' leadership practices: results of a baseline latent change score model

André Meyer ^a and Dirk Richter ^b

^aDepartment of Primary Education, University of Potsdam, Potsdam, Germany; ^bDepartment of Education, University of Potsdam, Potsdam, Germany

ABSTRACT

Principal leadership plays an important role in teacher practices and student achievement. Professional development (PD) can help principals to improve their leadership practices – yet few studies to date have examined the effects of formal PD on principals. The present study uses matched data from $n = 62$ principals and $n = 1112$ teachers to investigate the longitudinal effects of an 18-month PD program regarding school improvement for principals in Germany. Results from baseline latent change score modelling suggest positive changes in principals' self-reported leadership practices aimed at improving instructional quality at their schools and promoting teacher development. The positive change in practices is partly reflected in the reports of teachers in the respective schools. We find evidence for characteristics of high-quality PD programs for principals, such as joint participation of leadership pairs, group discussions, internships and coaching, that might help practitioners to design future PD activities for principals. Research should further examine characteristics of effective PD for principals taking control groups and confounding variables into account.

ARTICLE HISTORY

Received 23 May 2023
Accepted 26 September 2024

KEYWORDS

Principal leadership;
leadership development;
professional development;
latent change model;
baseline latent change score
modelling

Introduction

Principals' leadership practices are crucial in the classroom, impacting teachers' instructional quality and, ultimately, student achievement (Robinson *et al.* 2008, Grissom *et al.* 2021). However, research indicates that principals in some countries lack key skills for school improvement, such as goal-setting, supporting teachers' learning, and implementing improvement initiatives (Johnson *et al.* 2021, Klein and Schwanenberg 2022). Given the importance of these skills, principals need PD programs that equip them effectively. Yet, few studies have examined how formal PD programs influence principals' leadership practices, especially in school improvement.

This study investigates how principals evaluated an intensive 18-month PD program and its impact on their leadership practices in school improvement. The program, designed for leadership pairs (the principal and a leadership team member), included workshops, coaching, internships, and participation in professional learning networks, facilitated by experienced trainers with school leadership backgrounds.

CONTACT André Meyer  andre.meyer@uni-potsdam.de  Department of Primary Education Pedagogics, University of Potsdam, Karl-Liebknecht-Strasse 24-25, Potsdam 14476, Germany

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

We focus on how principals' leadership practices, particularly in instructional improvement and staff development, evolved during the program, assessed from both principals' and teachers' perspectives. Specifically, we address: (1) How do principals rate the overall quality and usefulness of the PD program's design features? (2) How do principals' self-reported leadership practices in instructional improvement and staff development change during the program? (3) How do teachers perceive changes in their principals' leadership practices in these areas?

The following sections provide the theoretical background on the importance of principals' leadership practices and the need for effective PD programs, followed by empirical findings and key features of the PD program studied.

Theoretical background

The relevance of principals' leadership practices and professional learning

For more than three decades, studies have been examining the leadership practices of principals and how they relate to characteristics of schools, teachers and students (see Grissom *et al.* 2021, Wu and Shen 2022). This research has shown that principals can support teachers in multiple ways to reflect upon and improve their instruction (Tuytens and Devos 2017). Principals can help teachers identify their need for PD, for example, through classroom visits and instructional feedback (Thillmann *et al.* 2015). They can help teachers improve the quality of their teaching by designing, monitoring and evaluating their schools' instructional programs (Sebastian *et al.* 2017, Bellibaş *et al.* 2021). They can implement structures and time slots for meaningful teacher-principal collaboration and involve teachers in decisions on school improvement (Meyer *et al.* 2022). The diverse ways in which principals interact with teachers affect both teachers' instructional quality and students' achievement (Grissom *et al.* 2021, Tan *et al.* 2021, Wu and Shen 2022). As Grissom *et al.* (2021) summarise in their meta-analysis of 219 international studies, 'replacing a below-average elementary school principal with an above-average principal would result in an additional 2.9 months of maths learning and 2.7 months of reading learning each year for students in that school.'

Many studies on principal leadership have investigated specific leadership styles (e.g. transformational leadership, instructional leadership; Robinson *et al.* 2008). Leadership styles represent combinations of practices that are geared towards a specific goal. Principals with an instructional leadership style, for instance, design clear and meaningful instructional goals for their school, supervise and evaluate the quality of teaching and promote teachers' professional development (Leithwood *et al.* 2004, Hallinger 2005). More recent studies on principal leadership, however, criticise concepts such as transformational and instructional leadership due to their alleged lack of distinct theoretical foundations (Kwan 2020). Many scholars recommend that, instead, research should investigate actual leadership practices that focus on teachers' and students' learning in schools (Hallinger 2003, Day *et al.* 2016, Daniëls *et al.* 2019, Robinson and Gray 2019).

In the present study, we investigate principal leadership in terms of specific practices that focus on improving the instructional quality in schools (i.e. instructional improvement) and on developing the professional competencies of teachers (i.e. staff development). We thus address leadership practices that principals in some countries feel that they are not adequately prepared to carry out (Johnson *et al.* 2021). For example, a large proportion of German principals lack skills in encouraging teachers to reflect on their instruction, in planning and implementing instructional innovations in their schools and in conducting classroom observation to evaluate instructional quality (Klein and Schwanenberg 2022). Moreover, principals report not feeling prepared to establish and implement staff development programs or motivate teachers to participate in PD (Klein and Schwanenberg 2022). Few studies so far have examined whether in-service PD programs address the needs of principals and effectively prepare them for their job. The present study helps to narrow

this research gap. In the following, we present empirical findings on the design of formal PD programs for principals and the effects of these programs on principals' leadership practices, teachers' professional practices and students' achievement.

Features of effective PD programs for principals

While extensive evidence exists on effective formal PD programs for teachers (Garet *et al.* 2001, Desimone *et al.* 2002, Desimone 2009, Borko *et al.* 2010, Darling-Hammond *et al.* 2017), research on principals' professional learning is limited. Findings overall indicate a set of features that appear to be equally important for the professional learning of both teachers and principals (e.g. sustained duration, active learning, collaboration among participants, design and facilitation by experts in the field; Garet *et al.* 2001, Desimone 2009, Orr 2011, Ni *et al.* 2019, Sahlin 2023). Findings also emphasise features of PD programs specific to principals (Davis and Darling-Hammond 2012). Principals' learning should focus on clear leadership practices or school improvement (Davis and Darling-Hammond 2012, Master *et al.* 2022, Steinberg and Yang 2022). PD themes should connect with principals' prior knowledge and align with their roles and needs (e.g. as instructional leaders or administrators; Eadens and Ceballos 2022). Effective PD should encourage principals to apply new skills and reflect on their impact (Davis and Darling-Hammond 2012, Daniëls *et al.* 2019).

Principal PD programs in the United States, such as the *Executive Development Program* (Nunnery *et al.* 2011, 2011, Master *et al.* 2022), *McREL's Balanced Leadership* PD program (Jacob *et al.* 2015, Miller *et al.* 2016) and *Pennsylvania's Inspired Leadership* induction program (Steinberg and Yang 2022), all contain key features of effective PD. Each of these programs consists of multiple PD activities over a period of 18 to 24 months. For example, the *Executive Development Program* includes 27 days of face-to-face sessions, 40 hours of online PD, and 50 to 100 hours of self-study over about 18 months (Nunnery *et al.* 2011, 2011). It aims to enhance principals' leadership in instructional improvement and data-driven organisational change, with additional coaching to help apply new skills (Master *et al.* 2022). Similarly, *McREL's Balanced Leadership* program consists of 10 two-day PD sessions for cohorts of 20 to 30 principals over 24 months. Principals discuss real-life cases, share experiences, and receive feedback from peers and trainers. The program aims to improve 21 empirically validated leadership practices, such as monitoring instruction and involvement in curriculum development, instruction, and assessment (Jacob *et al.* 2015, Miller *et al.* 2016). Evaluation studies of principal PD programs have produced mixed results on their impact on principals' and teachers' professional practices, as well as on changes in students' learning outcomes. In the following, we present a summary of the most relevant findings on the effects of principal PD programs.

Empirical findings on the effects of Principal PD programs

To date, few studies have investigated the effects of principals' professional learning in formal PD programs. These studies have generated mixed findings on changes in (1) principals' leadership practices, (2) teachers' practices and (3) students' achievement.

At the principal level, findings generally show positive changes in leadership practices due to formal PD activities. Andreoli *et al.* (2020) found that a three-year leadership program for 10 US principals led to improvements in their practices and school outcomes. In Turkey, Gümüő and Bellibaő (2020) observed a positive link between the frequency of PD participation and instructional leadership practices. Miller *et al.* (2016) conducted a randomised controlled trial with 126 Michigan principals and found significant improvements in leadership practices such as leading change and monitoring instructional quality. Herrmann *et al.* (2019) found minimal changes in leadership practices related to classroom observations and feedback from a two-year PD program involving 100 US principals. Also, Master *et al.* (2022) reported positive changes in leadership practices,

including increased classroom observations, in their randomised controlled trial of the 2016–2018 Executive Development Program cohorts.

Barnes *et al.* (2010) used a mixed-methods approach to evaluate a district-wide PD program for US principals, focusing on instructional leadership and school improvement. They found a shift from managerial to more instructional leadership, including strategic planning and data-driven decision-making. However, Camburn *et al.* (2016) found no significant impact from the same PD program in their experimental study, highlighting challenges in evaluating principal PD programs. Similarly, Gümüş and Bellibaş (2016) used TALIS 2013 data from 38 countries and found no significant relationship between formal PD activities and changes in leadership practices. Instead, professional networking, mentoring, and research showed weak but significant relationships with instructional leadership. These findings, based on cross-sectional data, suggest a need for more sophisticated and longitudinal study designs to better assess principals' professional learning.

At the teacher level, Steinberg and Yang (2022) found that principals' participation in Pennsylvania's Inspired Leadership program was linked to improved maths teacher effectiveness. Daniëls *et al.* (2021) conducted a quasi-experimental study with 190 teachers in Belgium and reported positive changes in the schools' organisational learning climate, such as increased teacher support, following their principals' participation in a seven-month PD program. Conversely, Herrmann *et al.* (2019) found no changes in teachers' practices from their principals' PD participation. Overall, there is limited evidence on the impact of principal PD programs on teachers' perceptions and behaviours.

At the student level, results are mixed. Nunnery *et al.* (2011, 2011) found positive effects of the Executive Development Program for principals on student achievement. Similarly, Steinberg and Yang (2022) reported improvements in students' reading achievement linked to principals' participation. However, randomised controlled trials by Herrmann *et al.* (2019), Jacob *et al.* (2015), and Master *et al.* (2022) found no significant improvements in student achievement. Fryer (2017) observed significant gains in student achievement after the first year of a two-year PD program but no effects in the second year.

In summary, few studies have investigated the effects of principals' professional learning through formal PD activities, and results have been mixed. The effectiveness of such PD is likely linked to specific design features like sustained duration, participant collaboration, and expert facilitation (Orr 2011, Davis and Darling-Hammond 2012, Darling-Hammond *et al.* 2017, Ni *et al.* 2019). The PD program studied here incorporates several of these features, which will be detailed in the next section. Most research has been US-based, with the exceptions of Daniëls *et al.* (2021) and Gümüş and Bellibaş (2016). This study explores the effects of a PD program for principals in Germany.

Principals training and development in Germany

In Germany, school principals are initially trained as teachers and typically gain experience in schools before entering principal preparation programs. However, these programs lack systematic approaches due to the decentralised nature of educational policymaking, leading to significant differences in requirements, duration, and content across federal states (Tulowitzki *et al.* 2019). Generally, prospective principals participate in formal training lasting one to two years before taking on the principalship, followed by on-the-job PD during their initial years. They are expected, but not required, to engage in regular PD thereafter. PD activities, offered mainly by state-level institutions, cover a range of topics such as instructional improvement, staff development, and school law (Tulowitzki *et al.* 2019). Recently, more universities have introduced Master's programs in school management and improvement. While there is limited international knowledge on the effects of principal training and PD, understanding of its impact within Germany is even scarcer. This study seeks to fill this gap by examining the impact of a PD program embedded in formal on-the-job PD activities in one German federal state.

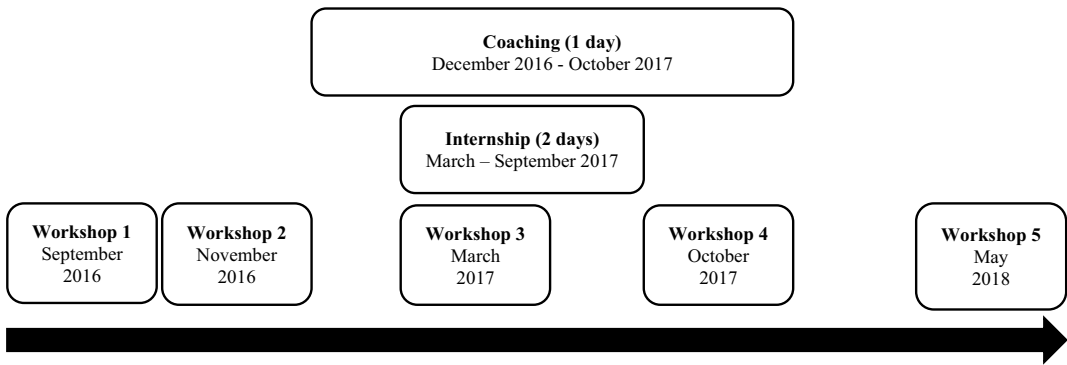


Figure 1. Elements and timetable of the 18-month PD program 'Werkstatt Schule leiten' displayed for the first cohort.

The present study

We investigate the effects of an 18-month PD program for principals in Germany, conducted with three cohorts from 2016 to 2021. The program, 'Werkstatt Schule leiten' ['Workshop School Leadership'], aimed to enhance principals' motivation (e.g. self-efficacy) and leadership practices to support school improvement, including instructional enhancement, teacher development, and fostering teacher collaboration. Principals learned to identify their schools' needs and goals for improvement, such as revising curricula to promote self-directed learning. They also learned to plan and implement a school improvement initiative, working in pairs with another leadership team member to design and execute initiatives tailored to their school's needs.

The program included five workshops, each lasting two to three days (see Figure 1). Activities comprised presentations by experienced principals and coaches, group discussions, self-study, reflection, and feedback. In *Workshop 1*, principals assessed their leadership strengths and weaknesses, created a vision for their school's long-term improvement, and identified short-term goals. *Workshop 2* focused on setting clear goals based on this vision. In *Workshop 3*, leadership pairs developed an improvement initiative and timeline, addressing areas like teacher collaboration, curriculum revision, or digital media integration. *Workshop 4* involved reflecting on the initiative's progress, learning feedback techniques, and discussing issues with peers and trainers. In *Workshop 5*, leadership pairs evaluated their initiative's outcomes and planned further steps for continued improvement after the program.

Apart from the face-to-face workshops, leadership pairs each received a full day of coaching and completed a two-day internship in a school that had previously been awarded the German school prize. This took place sometime after *Workshop 3*. The entire PD program was designed and facilitated by in-service principals and coaches with extensive experience in school leadership.

As the present PD program incorporated several features that previous studies found to be associated with participants' learning, we investigated the following research questions:

- (1) How do principals rate the overall quality of the PD after completing the 18-month program? How do principals evaluate the usefulness of specific design features of the PD program?
- (2) How do principals' self-reported leadership practices to support instructional improvement and staff development change over the course of the PD program?
- (3) How do principals' leadership practices to support instructional improvement and staff development change as reported by the teachers at the school?

Our research questions are structured around the evaluation framework of Kirkpatrick and Kirkpatrick (2006), who differentiate the outcomes of formal PD on four levels with increasing

complexity: (1) principals' satisfaction, (2) changes in principals' motivation and knowledge, (3) changes in principals' practices and (4) results of improved leadership practices. We focus on principals' satisfaction with the PD program (level 1) and changes in their leadership practices (level 3).

Methods

Study design

Over the 18-month PD program, we conducted two surveys with principals. The first survey, administered about two months after the program began (T_1), asked principals to report on their leadership practices. The second survey, conducted approximately two months after the program ended (T_2), again solicited reports on leadership practices and ratings of the PD program's quality and specific design features. To obtain objective measures, we also surveyed teachers about their principals' leadership practices. Teachers completed their surveys simultaneously with the principal surveys, with an additional post-survey conducted about one year after the program ended (T_3), assuming teachers might observe changes later. Both principals and teachers used individual codes to match their responses longitudinally.

Sample

To address Research Questions 1 and 2, we analysed panel data from $n = 62$ principals from $n = 31$ schools who participated as leadership pairs in two cohorts of the PD program ($n_1 = 32$, $n_2 = 30$). Thirty-five principals in our sample are women. Thirteen principals work at primary schools, 46 at secondary schools. We investigated Research Question 3 using a panel of $n = 1112$ teachers from $n = 29$ schools (69% female, 31% male). The mean response rates were 60% of teachers per school for the baseline survey at T_1 and 67% of teachers per school for the survey at T_2 . Fifty-four percent of teachers per school responded on average to the follow-up survey at T_3 .

Instruments

To investigate Research Question 1, we asked principals to rate the overall quality of the PD program after its completion using a scale that ranged from 1 (*very good*) to 6 (*not good at all*). We further asked principals to evaluate the usefulness of specific design features of the PD program using a four-point scale that ranged from 1 (*not useful at all*) to 4 (*very useful*). These design features included joint work of leadership pairs, written self-reflection, group discussions, lectures, collaboration in small groups, internships and coaching. To investigate Research Question 2, we used three items to measure principals' self-reported leadership practices to improve the instructional quality at their school (instructional improvement; e.g. 'I make sure that teachers have a common understanding of high-quality instruction'). Moreover, we asked principals to report on their leadership practices to support teacher development at their school using three items (staff development; e.g. 'I regularly provide feedback to teachers on their individual performance'). Items in both scales were rated on a four-point Likert scale from 1 (*strongly disagree*) to 4 (*strongly agree*). To address Research Question 3, we asked teachers to assess the leadership practices of their principals in terms of instructional improvement and staff development. We used the same list of items here that were used in the principals' survey and adapted the subject to the teachers' perspective (e.g. 'The principal at my school makes sure that teachers have a common understanding of high-quality instruction'). Again, teachers rated all items on a four-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). See [Table A1](#) in the Appendix for a full list of items.

Table 1. Information on scale reliability and intra-class correlations.

	McDonald's ω			ICC(1)			ICC(2)		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
Principal panel									
Instructional improvement	.73	.76	-	-	-	-	-	-	-
Staff development	.83	.79	-	-	-	-	-	-	-
Teacher panel									
Instructional improvement	.73	.76	.79	.15	.22	.17	.79	.87	.84
Staff development	.82	.81	.86	.22	.32	.22	.85	.91	.88

Analyses

We used McDonald's ω to evaluate all scales for internal consistency using the software Mplus 8.3 (Zinbarg *et al.* 2005, Hayes and Coutts 2020). We chose McDonald's ω for estimating scale reliability and used cut-off values for satisfactory reliability of $\omega \geq .65$ as suggested by Nájera Catalán (2019). As can be seen in Table 1, all scales showed satisfactory internal consistency.

As the data structure of the teacher panel was nested within schools – that is, several teachers from one school assessed the leadership practices of their principal – we calculated intra-class correlation coefficients ICC(1) and ICC(2) to estimate group-level correlations (James *et al.* 1984, Bliese 1998). ICC(1) indicates the amount of variance that can be attributed to differences between groups (Bliese 1998, LeBreton and Senter 2008). ICC(2) indicates the reliability of the group mean on the school level (Bliese 1998). We used cut-off values of ICC(1) $\geq .10$ and ICC(2) $\geq .70$ as recommended by Bliese (1998) and LeBreton and Senter (2008), respectively. As can be seen in Table 1, a sufficient amount of variance can be attributed to differences between groups (ICC(1)). Also, we find reliable group estimates for all scales (ICC(2)). As we modelled latent factors for each scale at each time point, we first conducted a confirmatory factor analysis to evaluate construct validity, estimating measurement models for each construct at each time point (Hair *et al.* 2014). For all measurement occasions, we find standardised factor loadings for instructional improvement and staff development – measured from the perspectives of principals and the teachers, respectively – greater than .46. We further evaluated measurement invariance to assess if all factors have the same meaning across repeated measurements (Putnick and Bornstein 2016). To adequately examine longitudinal mean differences of factor means, strong factorial invariance needs to be given in the data. That is, factors need to be equal in terms of factor structure, factor loadings and item intercepts (van de Schoot *et al.* 2012, Putnick and Bornstein 2016). We estimated relative fit indices such as AIC, BIC, aBIC and χ^2 , which allowed us to compare models with increasing restrictiveness, as well as absolute fit indices such as RMSEA and CFI that can be interpreted using cut-off values (Meade *et al.* 2008, Putnick and Bornstein 2016). When comparing models with regard to relative fit indices AIC, BIC and aBIC, smaller values indicate a better fit of the model to the data. In terms of χ^2 , non-significant values indicate good model fit. We further used well-established cut-off values of RMSEA $\leq .06$ and CFI $\geq .95$ to identify good model fit (Hu and Bentler 1999, Kenny *et al.* 2015). With regard to RMSEA, however, Chen *et al.* (2008) and Kenny *et al.* (2015) report that the RMSEA is too sensitive for models with small degrees of freedom (*df*) and small sample size ($N < 100$), which might lead to over-rejection of correct models. In this case, Meade *et al.* (2008) and Putnick and Bornstein (2016) recommend using $\Delta\chi^2$ and Δ CFI to compare models with increasing restriction. If $\Delta\chi^2$ is significant, the more restrictive model is to be rejected. We used cut-off values of Δ CFI $\leq -.02$ for evaluating weak factorial invariance and Δ CFI $\leq -.01$ for strong factorial invariance (Putnick and Bornstein 2016). As can be seen in Table A2 and Table A3 (see Appendix), strong factorial invariance was given for both scales in the principal panel and the teacher panel, respectively.

To address Research Question 1, we conducted descriptive analyses using IBM SPSS Statistics to estimate the mean ratings of the PD program's overall quality and the relative frequencies of principals' evaluations of specific design features. For Research Question 2, we used the principal panel and estimated two latent baseline change score models with Mplus 8.3 (Klopack and Wickrama 2020). Latent change score models are a specialised type of structural equation modelling (SEM) used to investigate changes in measures taken at multiple time points (McArdle 2009). This approach is more advanced than methods like repeated measures ANOVA because it accounts for measurement errors and latent difference scores, leading to less biased estimates of mean differences over time (Geiser 2020). We thus estimated latent factor means for the constructs of instructional improvement and staff development at T_1 and T_2 . We then calculated a latent change score that indicates whether there was significant change in latent factor means over time. For Research Question 3, we used a similar approach in analysing data from the teacher panel. As we collected data from teachers on three occasions, we estimated a baseline latent change score model (Geiser 2020). That is, we estimated latent change scores that refer to the baseline measure at T_1 . These change scores indicate whether there was a significant change in leadership practices from T_1 to T_2 and from T_1 to T_3 . For each model, we estimated unstandardised change scores that resemble actual mean differences. We further estimated standardised change scores that can be interpreted as effect sizes, as suggested by Little (1988). Values of Cohen's $d \geq .20$ indicate small effects, $d \geq .50$ indicate medium effects and $d \geq .80$ indicate large effects (Cohen 1988). We accounted for the nested structure of the data from the teacher panel by using the school ID as a cluster variable in the model. Moreover, the algorithm TYPE=COMPLEX was used.

Missing data

We evaluated the data for missing values using the MCAR test (missing completely at random) in the software IBM SPSS Statistics (Little 1988). Non-significant results of the MCAR test for the principal and the teacher panels indicated that data were missing completely at random (principals: $\chi^2 = 240.21$, $df = 235$, $p > .05$; teachers: $\chi^2 = 3196.77$, $df = 3169$, $p > .05$). All further analyses were conducted in Mplus 8.3 using FIML estimation (full information maximum likelihood), which takes all available information into account. This creates unbiased estimates without imputing missing values (Schafer and Graham 2002, Graham 2003).

Results

In Research Question 1, we were interested in how principals rated the overall quality of the PD program on a scale from 1 (*very good*) to 6 (*not good at all*). Results of descriptive analyses show that principals gave the PD program very high ratings ($M = 1.56$, $SD = .67$). Moreover, we asked principals to evaluate the usefulness of specific design features of the PD program (see Figure 2). The vast majority of principals perceived the work of the leadership pairs – that is, their work with the second member of their leadership team – as being the most useful activity in the PD program (very useful: 86%). Moreover, principals evaluated group discussions, presentations, collaboration in small groups and the two-day internship to be of equal usefulness (very useful: $\approx 70\%$). Finally, principals perceived the coaching they received and phases of written self-reflection to be the least useful to them.

Research Question 2 addressed changes in principals' self-reported leadership practices to support instructional improvement and staff development over the course of the PD program. The results show a significant change in principals' leadership practices to support instructional improvement from T_1 to T_2 ($\Delta M = .21$, $SD = .34$, $p < .05$), with a medium effect size of Cohen's $d = .61$ (see Figure 3). This means that principals reported improving their leadership practices aimed at improving the instructional quality at their school over the course of the PD program. Our results further indicate a significant change in principals' leadership practices aimed at staff development ($\Delta M = .21$, $SD = .57$,

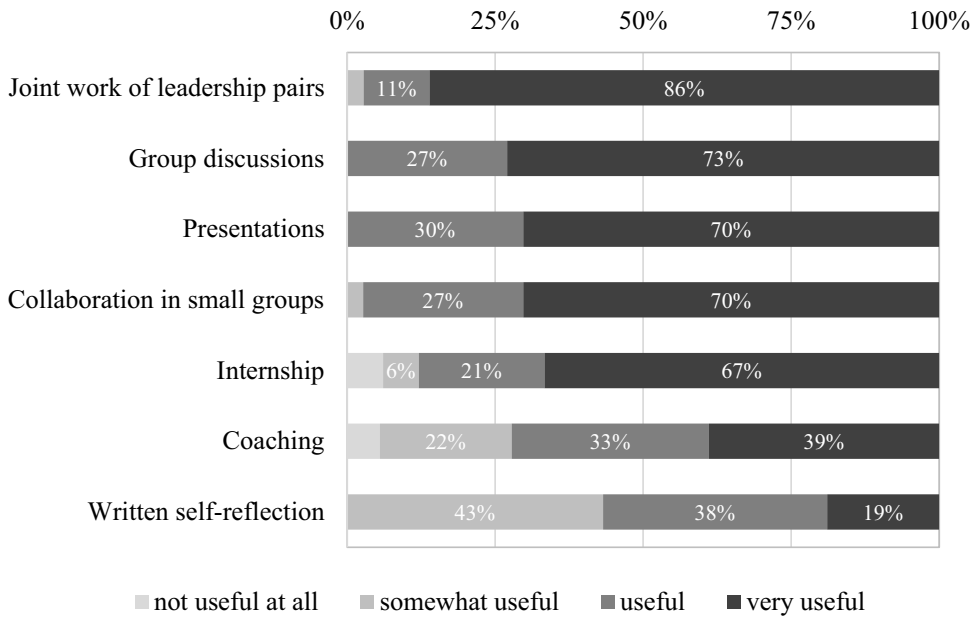


Figure 2. Relative frequency of perceived usefulness of specific components of the PD program as reported by principals (n = 58).

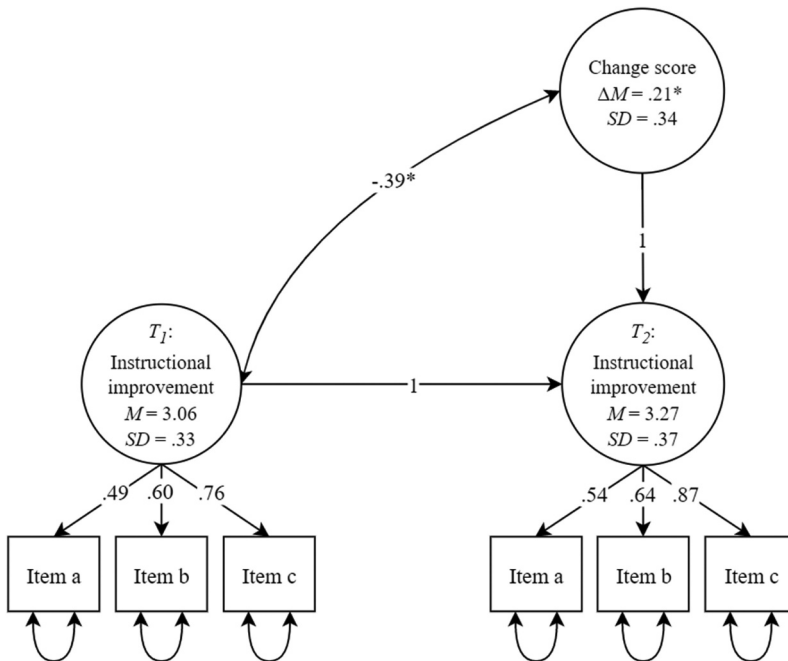


Figure 3. Latent change score model of principals' self-reported leadership practices aimed at instructional improvement. Note: * $p < .05$, ** $p < .01$, *** $p < .001$

$p < .05$), with small effect size of Cohen's $d = .37$ (see Figure 4). Principals' self-reported staff development practices, such as providing feedback to teachers on their performance, increased significantly over the 18-month PD program. Despite similar change scores for instructional improvement and staff development, effect sizes differ due to differences in standard deviation. For both measures,

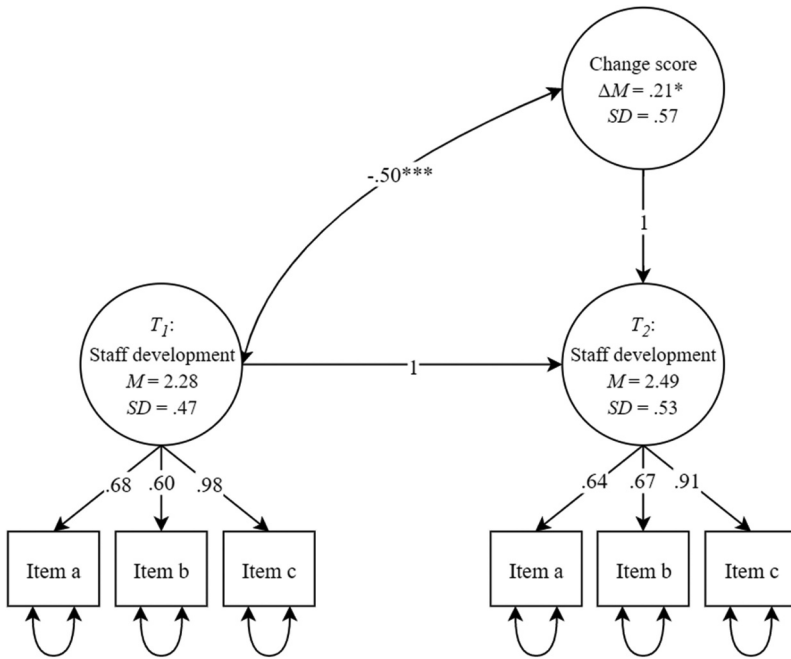


Figure 4. Latent change score model of principals' self-reported leadership practices aimed at staff development. Note: * $p < .05$, ** $p < .01$, *** $p < .001$

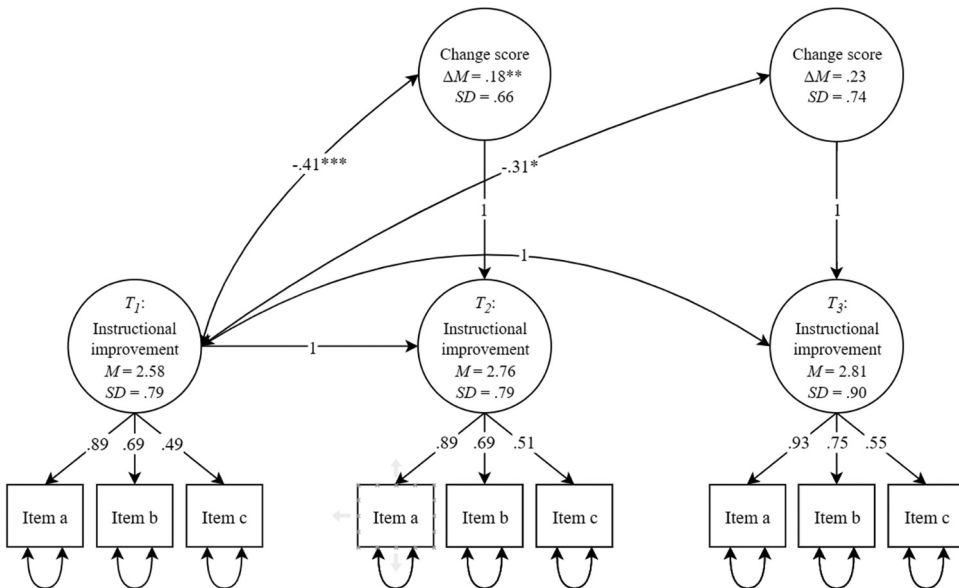


Figure 5. Baseline latent change score model of principals' leadership practices aimed at instructional improvement as reported by teachers. Note: * $p < .05$, ** $p < .01$, *** $p < .001$

correlation coefficients between the baseline measurement (T_1) and the change score indicate that principals who rated their leadership practices at T_1 more positively reported smaller changes over time (instructional improvement: $r = -.39, p < .05$; staff development: $r = -.50, p < .001$).

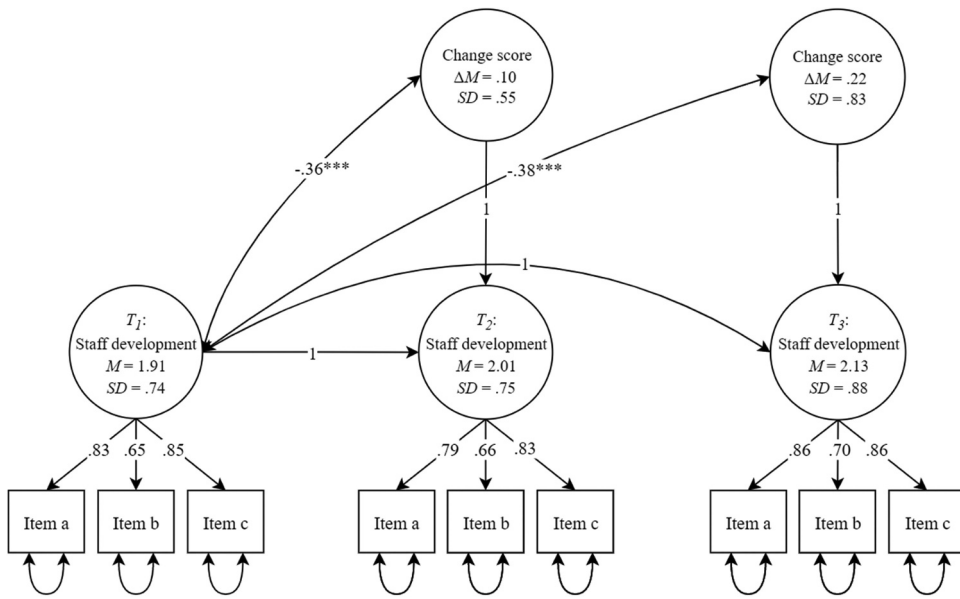


Figure 6. Baseline latent change score model of principals' leadership practices aimed at staff development as reported by teachers. Note: * $p < .05$, ** $p < .01$, *** $p < .001$

In Research Question 3, we investigated changes in principals' leadership practices from the teachers' perspective. Again, the results show significant changes in principals' practices to support instructional improvement from T_1 to T_2 ($\Delta M = .18$, $SD = .66$, $p < .01$) with a small effect size of Cohen's $d = .27$ (see Figure 5). Results also show a non-significant change from T_1 to T_3 ($\Delta M = .23$, $SD = .74$, $p > .05$), with a small effect size of Cohen's $d = .31$. This finding indicates that teachers also perceive positive changes in principals' practices aimed at improving the instructional quality at their school. With regard to principals' practices aimed at staff development, we find positive changes from T_1 to T_2 ($\Delta M = .10$, $SD = .55$, $p > .05$), with an effect size of Cohen's $d = .18$ and from T_1 to T_3 ($\Delta M = .22$, $SD = .83$, $p > .05$), with a small effect size of Cohen's $d = .27$ (see Figure 6). However, none of these changes was statistically significant. Finally, our results indicate that teachers who assessed their principals' leadership practices more positively at the baseline measurement (T_1) reported smaller changes over time (instructional improvement: $r_{T_1-T_2} = -.41$, $p < .001$; $r_{T_1-T_3} = -.31$, $p < .05$; staff development: $r_{T_1-T_2} = -.36$, $p < .001$; $r_{T_1-T_3} = -.38$, $p < .001$).

Discussion

This paper examined the quality and longitudinal effects of an 18-month PD program on principals' leadership practices for school improvement, focusing on instructional improvement and staff development. We used Kirkpatrick's and Kirkpatrick (2006) evaluation framework, which assesses (1) participant satisfaction, (2) changes in motivation and knowledge, (3) changes in professional practices, and (4) results of improved practices. We anticipated finding improvements in principals' practices, given that the PD program included several effective PD features identified by previous research (Davis and Darling-Hammond 2012, Darling-Hammond *et al.* 2017). The following sections address our research questions and discuss the results, considering previous findings, methodological limitations, and implications for researchers, practitioners, and policymakers.

How do principals rate the overall quality of the PD program? How useful do they perceive specific components of the PD program to be?

Our findings indicate that principals rated the overall quality of the PD program very positively, especially appreciating the collaborative elements such as joint work in leadership pairs, group discussions, and small-group collaborations. Leadership pairs found their joint work to be the most beneficial aspect of the PD program for their professional development. This may be because principals often lack time for such collaboration during their regular workday (Brauckmann *et al.* 2015). Research on distributed leadership shows that principals typically delegate leadership responsibilities, such as supporting teachers, to assistant principals or teacher leaders (Tian *et al.* 2016). The PD program's focus on collaborative school improvement required principals to collectively reflect on their school's needs, develop and implement improvement initiatives, and evaluate their outcomes (Meyer *et al.* 2022). The positive assessment of the PD program likely stems from this structured opportunity for collaboration, aligning with evidence that collaboration is a significant motivator for principals and teachers to engage in PD activities (Davis and Darling-Hammond 2012, Richter *et al.* 2022, Daniëls *et al.* 2023, Sahlin 2023). Additionally, principals highlighted the usefulness of group discussions within their PD cohort, supporting findings from Barnes *et al.* (2010) and Sahlin (2023) that suggest high-quality interactions among principals foster the sharing of knowledge, experiences, and ideas for school improvement.

Did the PD program support changes in principals' leadership practices?

We asked principals twice to assess their leadership practices in instructional improvement and staff development. Principals' self-reports indicated small to medium positive changes during the PD program. Teachers' reports partially validated this, showing short- and long-term improvements in principals' practices. However, only the short-term change in instructional improvement was statistically significant. Despite small effect size differences, the change scores from T_1 to T_3 were non-significant.

This finding aligns with research by Andreoli *et al.* (2020), Gümüş and Bellibaş (2020), Herrmann *et al.* (2019), Master *et al.* (2022), and Miller *et al.* (2016), all of whom reported positive effects of principals' formal professional development (PD) on their leadership practices, such as increased classroom observations (Master *et al.* 2022). However, our results differ from Gümüş and Bellibaş (2016), who found no relationship between principals' PD participation and leadership practices, likely due to their focus on PD quantity rather than quality. Research emphasises that the quality of PD activities is crucial for professional learning (Desimone *et al.* 2002, Borko *et al.* 2010, Davis and Darling-Hammond 2012, Daniëls *et al.* 2019). In our study, the PD program incorporated key features of effective PD, including workshops, group discussions, coaching, and the opportunity for principals to apply their skills in a school improvement initiative. Unlike previous studies (Jacob *et al.* 2015, Herrmann *et al.* 2019, Master *et al.* 2022), this PD program also paired principals with a leadership team member, fostering joint reflection on their school's unique needs.

Limitations

The findings of our study should be considered with a few methodological limitations. First, principals' retrospective assessments of the PD program's quality may be biased by recall uncertainties (Coghlin 1990). Second, without a control group, we cannot establish a causal link between the PD program and changes in principals' practices, as we cannot control for confounding variables. Third, we lack information on how specific PD components relate to principals' professional learning. Fourth, this study focused on one PD program for principals in Germany, so the results should be interpreted within that national context. Despite these limitations, the study offers valuable insights into the effects of formal PD on principals' learning and the design features that support it.

Implications

The implications of this study are important for researchers, educators, and policymakers. Future research should explore the design features of formal PD programs for principals in greater detail, ideally using a control group to provide causal evidence and account for confounding variables. However, Camburn *et al.* (2016) highlight the challenges of conducting experiments in education research, including the confounding effects of dynamic social and policy contexts, treatment fidelity, and compliance issues, particularly within control groups. A possible solution could be using a waiting group of potential PD participants as a control.

While we know that formal PD can influence principals' leadership, teachers' instruction, and student outcomes, we still lack understanding of which PD features contribute most to principals' learning. Future studies should examine the impact of collaboration in leadership pairs and across PD cohorts on knowledge gains, motivation, and practice changes (Barnes *et al.* 2010). Additionally, research should focus on how principals transfer new knowledge into daily practice, as this is crucial for achieving meaningful changes at the school, teacher, and student levels (Sims *et al.* 2023).

Increases in teacher-reported change scores suggest that teachers may perceive changes in their principals' practices with some delay. This aligns with Wang *et al.* (2019), who found that innovations from principals' professional learning can take up to three years to fully implement, with teachers noticing changes only in the later stages. Future studies on principal PD programs should consider that teacher reports may be an unreliable measure of changes in principals' practices if the pre- and post-survey intervals are too short. Follow-up surveys with teachers may provide more accurate assessments of such changes over time.

Based on our results, we recommend that educational practitioners and policymakers design principal PD programs that offer a range of continuous learning activities, such as presentations, internships, coaching, and reflection phases. Collaborative activities, like working in leadership pairs, group discussions, and small group work, should be included. Leadership pairs, in particular, allow principals to collaboratively reflect on their school's needs, develop ideas, and implement improvements – tasks often overlooked in daily routines. Group discussions and mutual feedback within the PD cohort also promote knowledge sharing and idea exchange.

We further recommend that PD activities be of extended duration, with workshops lasting at least two days. While potentially costly, this allows principals to step away from daily routines, engage fully in learning, and adopt an objective perspective on their school's structure and needs (Sahlin 2023). Complex PD programs involving multiple activities – such as input sessions, discussions, coaching, and feedback – are especially important for addressing challenging areas like instructional improvement, staff development, and organisational learning, considering the school context (Andreoli *et al.* 2020, Brauckmann *et al.* 2023).

In sum, our study advances research on the effects of principals' professional learning in formal PD programs and provides new insights into effective PD design. The 18-month program significantly improved principals' leadership practices, especially in instructional improvement and staff development, as shown by both principals' self-reports and teacher feedback. Collaborative elements, such as leadership pair work and group discussions, were particularly beneficial. The study highlights that effective PD should include sustained, varied activities and structured collaboration to promote meaningful professional growth. Future research should examine which PD features have the greatest impact on leadership development and incorporate control groups to investigate causality.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is supported by Robert Bosch Stiftung, formerly Die Deutsche Schulakademie. The views contained in this article are those of the authors and not their institutions or the Robert Bosch Stiftung.

ORCID

André Meyer  <http://orcid.org/0000-0003-3198-5191>

Dirk Richter  <http://orcid.org/0000-0002-2384-1588>

References

- Andreoli, P.M., *et al.* 2020. Learning to lead school improvement: an analysis of rural school leadership development. *Journal of educational change*, 21 (4), 517–542. doi:10.1007/s10833-019-09357-z.
- Barnes, C.A., *et al.* 2010. Developing instructional leaders: using mixed methods to explore the Black Box of planned change in principals' professional practice. *Educational administration quarterly*, 46 (2), 241–279. doi:10.1177/1094670510361748.
- Bellibaş, M.Ş., Gümüş, S., and 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. doi:10.1080/09243453.2020.1858119.
- Bliese, P.D., 1998. Group size, ICC values, and group-level correlations: a simulation. *Organizational research methods*, 4 (1), 355–373. doi:10.1177/109442819814001.
- Borko, H., Jacobs, J., and Koellner, K., 2010. Contemporary approaches to Teacher professional development. *International encyclopedia of Education*, 7 (2), 548–556. doi:10.1016/B978-0-08-044894-7.00654-0.
- Brauckmann, S., *et al.* 2015. No time to manage? The trade-off between relevant tasks and actual priorities of school leaders in Germany. *International journal of educational management*, 29 (6), 749–765. doi:10.1108/IJEM-10-2014-0138.
- Brauckmann, S., Pashiardis, P., and Ärlestig, H., 2023. Bringing context and educational leadership together: fostering the professional development of school principals. *Professional development in education*, 49 (1), 4–15. doi:10.1080/19415257.2020.1747105.
- Camburn, E.M., *et al.* 2016. An examination of the benefits, limitations, and challenges of conducting randomized experiments with principals. *Educational administration quarterly*, 52 (2), 187–220. doi:10.1177/0013161X15617808.
- Chen, F., *et al.* 2008. An empirical evaluation of the Use of fixed cutoff points in RMSEA Test statistic in structural equation models. *Sociological methods & research*, 36 (4), 462–494. doi:10.1177/0049124108314720.
- Coghlin, S.S., 1990. Recall bias in epidemiologic studies. *Journal of clinical epidemiology*, 43 (12), 1431–1432. doi:10.1016/0895-4356(90)90113-4.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. New York, NY: Lawrence Erlbaum Associates.
- Daniëls, E., Hondeghem, A., and Dochy, F., 2019. A review on leadership and leadership development in educational settings. *Educational research review*, 27, 110–125. doi:10.1016/j.edurev.2019.02.003.
- Daniëls, E., Hondeghem, A., and Heystek, J., 2023. Developing school leaders: responses of school leaders to group reflective learning. *Professional development in education*, 49 (1), 135–149. doi:10.1080/19415257.2020.1766543.
- Daniëls, E., Muyters, G., and Hondeghem, A., 2021. Leadership training and organizational learning climate: measuring influences based on a field experiment in education. *International journal of training and development*, 25 (1), 43–59. doi:10.1111/ijtd.12206.
- Darling-Hammond, L., Hyler, M.E., and Gardner, M., 2017. *Effective Teacher professional development*. Palo Alto, CA: Learning Policy Institute.
- Davis, S.H. and Darling-Hammond, L., 2012. Innovative Principal preparation programs: what works and how we know. *Planning and changing*, 43 (1), 25–45.
- Day, C., Gu, Q., and Sammons, P., 2016. The impact of leadership on Student outcomes. *Educational administration quarterly*, 52 (2), 221–258. doi:10.1177/0013161X15616863.
- Desimone, L.M., *et al.* 2002. Effects of professional development on teachers' instruction: results from a three-year longitudinal study. *Educational evaluation and policy analysis*, 24 (2), 81–112. doi:10.3102/01623737024002081.
- Desimone, L.M., 2009. Improving impact studies of teachers' professional development: toward better conceptualizations and measures. *Educational Researcher*, 38 (3), 181–199. doi:10.3102/0013189X08331140.
- Eadens, D.W. and Ceballos, M., 2022. Educational leadership preparation and professional roles: are we serving the needs of leadership roles along the leadership continuum? *Journal of research on leadership education*. 18 (2), 277–300. doi:10.1177/19427751221076416. Advance online publication.

- Fryer, R.G., 2017. *Management and Student achievement: evidence from a randomized field experiment*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w23437.
- Garet, M.S., et al. 2001. What makes professional development effective? Results from a national sample of teachers. *American educational research journal*, 38 (4), 915–945. doi:10.3102/00028312038004915.
- Geiser, C., 2020. *Longitudinal structural equation modeling with Mplus. A latent state-trait perspective*. New York, NY: Guilford Press.
- Graham, J.W., 2003. Adding missing-data-relevant variables to FIML-Based structural equation models. *Structural equation modeling: A multidisciplinary journal*, 10 (1), 80–100. doi:10.1207/S15328007SEM1001_4.
- Grissom, J.A., Egalite, A., and Lindsay, C.A., 2021. *How principals affect students and schools: a systematic synthesis of two decades of research*. New York: The Wallace Foundation.
- Gümüş, E. and Bellibaş, 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. doi:10.1080/03055698.2016.1172958.
- Gümüş, S. and Bellibaş, M.Ş., 2020. The relationship between professional development and school principals' leadership practices: the mediating role of self-efficacy. *International journal of educational management*, 34 (7), 1155–1170. doi:10.1108/IJEM-10-2019-0380.
- Hair, J.F., et al. 2014. *Multivariate data analysis*. 7th ed. ed. Harlow: Pearson Education.
- Hallinger, P., 2003. Leading educational change: reflections on the practice of instructional and transformational leadership. *Cambridge journal of education*, 33 (3), 329–352. doi: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. doi:10.1080/15700760500244793.
- Hayes, A.F. and Coutts, J.J., 2020. Use omega rather than Cronbach's alpha for estimating reliability. *But ... communication methods and measures*, 14 (1), 1–24. doi:10.1080/19312458.2020.1718629.
- Herrmann, M., et al. 2019. *The effects of a Principal professional development program focused on instructional leadership*. Washington D.C.: National Center for Education Evaluation and Regional Assistance.
- Hu, L. and 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. doi:10.1080/10705519909540118.
- Jacob, R., et al. 2015. Exploring the causal Impact of the McREL balanced leadership program on leadership, Principal efficacy, instructional climate, educator turnover, and Student achievement. *Educational evaluation and policy analysis*, 37 (3), 314–332. doi:10.3102/0162373714549620.
- James, L.R., Demaree, R.G., and Wolf, G., 1984. Estimating within-group interrater reliability with and without response bias. *Journal of applied psychology*, 69 (1), 85–98. doi:10.1037/0021-9010.69.1.85.
- Johnson, A.D., et al. 2021. The professional learning needs of school principals. *Journal of research on leadership education*, 16 (4), 305–338. doi:10.1177/1942775120933933.
- Kenny, D.A., Kaniskan, B., and McCoach, D.B., 2015. The performance of RMSEA in models with small degrees of freedom. *Sociological methods & research*, 44 (3), 486–507. doi:10.1177/0049124114543236.
- Kirkpatrick, D.L. and Kirkpatrick, J.D., 2006. *Evaluating training programs: the four levels*. San Francisco, CA: Berrett-Koehler Publishers, Inc.
- Klein, E.D. and Schwanenber, J., 2022. Ready to lead school improvement? Perceived professional development needs of principals in Germany. *Educational management administration & leadership*, 50 (3), 371–391. doi:10.1177/1741143220933901.
- Klopach, E.T. and Wickrama, K.K.A.S., 2020. Modeling latent change score analysis and extensions in Mplus: a practical Guide for researchers. *Structural equation modeling: A multidisciplinary journal*, 27 (1), 97–110. doi:10.1080/10705511.2018.1562929.
- Kwan, P., 2020. Is transformational leadership theory passé? Revisiting the integrative effect of instructional leadership and transformational leadership on Student outcomes. *Educational administration quarterly*, 56 (2), 321–349. doi:10.1177/0013161X19861137.
- LeBreton, J.M. and Senter, J.L., 2008. Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, 11 (4), 815–852. doi:10.1177/1094428106296642 .
- Leithwood, K., et al. 2004. *How leadership influences student learning*. New York: The Wallace Foundation.
- Little, R.J.A., 1988. A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical association*, 83 (404), 1198–1202. doi:10.1080/01621459.1988.10478722.
- Master, B.K., et al. 2022. Developing school leaders: findings from a randomized control trial study of the executive development program and paired coaching. *Educational evaluation and policy analysis*, 44 (2), 257–282. doi:10.3102/01623737211047256.
- McArdle, J.J., 2009. Latent variable modeling of differences and changes with longitudinal data. *Annual review of psychology*, 60 (1), 577–605. doi:10.1146/annurev.psych.60.110707.163612.
- Meade, A.W., Johnson, E.C., and Braddy, P.W., 2008. Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of applied psychology*, 93 (3), 568–592. doi:10.1037/0021-9010.93.3.568.

- Meyer, A., *et al.* 2022. How can principal leadership practices promote teacher collaboration and organizational change? A longitudinal multiple case study of three school improvement initiatives. *Journal of educational change*, 24 (3), 425–455. doi:10.1007/s10833-022-09451-9. Online first
- Miller, R.J., *et al.* 2016. Can professional development improve school leadership? Results from a randomized control trial assessing the impact of McREL's balanced leadership program on principals in Rural Michigan schools. *Educational administration quarterly*, 52 (4), 531–566. doi:10.1177/0013161X16651926.
- Nájera Catalán, H.E., 2019. Reliability, population classification and weighting in multidimensional poverty measurement: a monte carlo study. *Social indicators research*, 142 (3), 887–910. doi:10.1007/s11205-018-1950-z.
- Ni, Y., *et al.* 2019. Leadership matters: preparation program quality and learning outcomes. *Journal of educational administration*, 57 (2), 185–206. doi:10.1108/JEA-05-2018-0093.
- Nunnery, J.A., *et al.* 2011. *The impact of the NISL executive development program*. Norfolk, VA: The Center for Educational Partnerships, Old Dominion University.
- Nunnery, J.A., Yen, C.-, and Ross, S.M., 2011. *Effects of the national institute for school Leadership's executive development program on school performance in Pennsylvania: 2006-2010 Pilot cohort results*. Norfolk, VA: The Center for Educational Partnerships, Old Dominion University.
- Orr, M.T., 2011. Pipeline to preparation to advancement: graduates' experiences In, through, and beyond leadership preparation. *Educational administration quarterly*, 47 (1), 114–172. doi:10.1177/0011000010378612.
- Putnick, D.L. and Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental review*, 41, 71–90. doi:10.1016/j.dr.2016.06.004.
- Richter, E., *et al.* 2022. Teacher collaboration and professional learning. Examining professional development during a National Education Reform. *Zeitschrift für Pädagogik*, 68 (6), 798–819. doi:10.3262/ZP2206798.
- Robinson, V.M.J. and Gray, E., 2019. What difference does school leadership make to student outcomes? *Journal of the royal society of New Zealand*, 49 (2), 171–187. doi:10.1080/03036758.2019.1582075.
- Robinson, V.M.J., Lloyd, C.A., and 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. doi:10.4135/9781473915152.n6.
- Sahlin, S., 2023. Professional development of school principals – how do experienced school leaders make sense of their professional learning? *Educational management administration & leadership*, 174114322311682. doi:10.1177/17411432231168235.
- Schafer, J.L. and Graham, J.W., 2002. Missing data: our view of the state of the art. *Psychological methods*, 7 (2), 147–177. doi:10.1037/1082-989X.7.2.147.
- Sebastian, J., Huang, H., and Allensworth, E., 2017. Examining integrated leadership systems in high schools: connecting principal and teacher leadership to organizational processes and student outcomes. *School effectiveness and school improvement*, 28 (3), 463–488. doi:10.1080/09243453.2017.1319392.
- Sims, S., *et al.* 2023. Effective Teacher professional development: new theory and a Meta-analytic Test. *Review of educational research*. doi:10.3102/00346543231217480.
- Steinberg, M.P. and Yang, H., 2022. Does Principal professional development improve schooling outcomes? Evidence from Pennsylvania's inspired leadership induction program. *Journal of research on educational effectiveness*, 15 (4), 799–847. doi:10.1080/19345747.2022.2052386.
- Tan, C.Y., Dimmock, C., and 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. doi:10.1177/17411432211061445. Advance online publication.
- Thillmann, K., *et al.* 2015. School-based staff development in two federal states in Germany. *International journal of educational management*, 29 (6), 714–734. doi:10.1108/IJEM-07-2014-0094.
- Tian, M., Risku, M., and Collin, K., 2016. A meta-analysis of distributed leadership from 2002 to 2013. *Educational management administration & leadership*, 44 (1), 146–164. doi:10.1177/1741143214558576.
- Tulowitzki, P., Hinzen, I., and Roller, M., 2019. Die Qualifizierung von Schulleiter*innen in Deutschland - ein bundesweiter Überblick [The Qualification of School Leaders in Germany: A National Overview]. *Die Deutsche Schule*, 111 (2), 149–170. doi:10.31244/ds.2019.02.04.
- Tuytens, M. and Devos, G., 2017. The role of feedback from the school leader during teacher evaluation for teacher and school improvement. *Teachers & teaching*, 23 (1), 6–24. doi:10.1080/13540602.2016.1203770.
- van de Schoot, R., Lugtig, P., and Hox, J., 2012. A checklist for testing measurement invariance. *The European journal of developmental psychology*, 9 (4), 486–492. doi:10.1080/17405629.2012.686740.
- Wang, E.L., *et al.* 2019. *Putting professional learning to work: what principals do with their Executive development program learning*. Santa Monica, CA: RAND Corporation.
- Wu, H. and Shen, J., 2022. The association between principal leadership and student achievement: a multivariate meta-meta-analysis. *Educational research review*, 35, 100423. doi:10.1016/j.edurev.2021.100423.
- Zinbarg, R.E., *et al.* 2005. Cronbach's α , Revelle's β , and McDonald's ω : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70 (1), 123–133. doi:10.1007/s11336-003-0974-7.

Appendix

Table A1. Full list of items used in the Principal and Teacher surveys.

Principal survey	
<i>Instructional improvement</i>	
I make sure that teachers have a common understanding of high-quality instruction.	
I urge teachers to make sure that all students achieve the instructional goals.	
I use data from student assessments for instructional improvement.	
<i>Staff development</i>	
I regularly provide feedback to teachers on their individual performance.	
Staff development is an essential part of my leadership practice.	
I regularly conduct appraisal interviews with teachers.	
Teacher survey	
<i>Instructional improvement</i>	
The principal at my school makes sure that teachers have a common understanding of high-quality instruction.	
The principal at my school urges teachers to make sure that all students achieve the instructional goals.	
The principal at my school uses data from student assessments for instructional improvement.	
<i>Staff development</i>	
The principal at my school regularly provides feedback to teachers on their individual performance.	
Staff development is an essential part of our principal's leadership practice.	
The principal at my school regularly conducts appraisal interviews with teachers.	

Table A2. Measurement invariance across time for principals' self-reported leadership practices.

Model	AIC	BIC	aBIC	χ^2 (df)	RMSEA	CFI	$\Delta\chi^2$ (Δdf)	ΔCFI
Instructional improvement (n = 58)								
M1a: configural invariance	621.82	671.27	595.81	2.67 (3)	.00	1.00	-	-
M1b: weak factorial invariance	620.27	663.54	597.51	7.13 (6)	.06	.99	4.45 (3)	-.01
M1c: strong factorial invariance	618.20	657.35	597.61	9.06 (8)	.05	.99	1.94 (2)	.00
M1d: strict factorial invariance	613.47	646.44	596.13	10.33 (11)	.00	1.00	1.27 (3)	.01
Staff development (n = 57)								
M2a: configural invariance	666.76	715.82	640.37	7.19 (3)	.16	.96	-	-
M2b: weak factorial invariance	666.03	710.98	641.82	10.433 (5)	.14	.95	3.25 (2)	-.01
M2c: strong factorial invariance	663.94	704.81	641.93	12.34 (7)	.12	.95	1.91 (2)	.00
M2d: strict factorial invariance	660.79	695.53	642.08	15.19 (10)	.10	.96	2.85 (2)	.01

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table A3. Measurement invariance across time for principals' leadership practices as reported by teachers.

Model	AIC	BIC	aBIC	χ^2 (<i>df</i>)	RMSEA	CFI	$\Delta\chi^2$ (Δdf)	ΔCFI
Instructional improvement (<i>n</i> = 1088)								
M3a: configural invariance	9882.66	10067.36	9949.84	26.28 (17)	.02	.99	-	-
M3b: weak factorial invariance	9880.60	10035.35	9936.89	38.89* (23)	.03	.99	12.62 (6)	.00
M3c: strong factorial invariance	9878.74	10013.53	9927.77	44.75* (27)	.03	.99	5.85 (4)	.00
M3d: strict factorial invariance	9870.69	9975.53	9908.82	46.77 (33)	.02	.99	2.02 (6)	.00
Staff development (<i>n</i> = 1068)								
M4a: configural invariance	9703.02	9887.04	9769.52	22.40 (17)	.02	1.00	-	-
M4b: weak factorial invariance	9695.51	9849.69	9751.22	26.03 (23)	.01	1.00	3.63 (6)	.00
M4c: strong factorial invariance	9690.23	9824.49	9738.73	27.91 (27)	.01	1.00	1.88 (4)	.00
M4d: strict factorial invariance	9691.56	9796.00	9729.30	38.29 (33)	.01	1.00	10.38 (6)	.00

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.