



Regular Article

Understanding teacher leadership: Development and testing of the survey of teacher leadership

Peter D. Wiens^{a,*}, Vanessa Vongkulluksn^b, Jori Beck^c, Brenda Pearson^d

^a Teacher Education, University of Nevada, Las Vegas, USA

^b Learning Sciences, University of Nevada, Las Vegas, USA

^c Teacher Education, Old Dominion University, USA

^d Strategic Initiatives, Clark County Education Association, USA

ARTICLE INFO

Keywords:

Teacher leadership

Survey development

Equity

Shared leadership

Teacher professional learning

ABSTRACT

Teacher leadership (TL) has been identified as key component of effective schools. While research has grown in this area, there is a dearth of empirically and theoretically supported measures for understanding TL. This study examines the validity and reliability of the Survey of Teacher Leadership (STL). The STL builds on previous theory and research in TL by creating an assessment of teacher TL practices and their associated self-efficacy in these practices. In addition to more established domains of TL, the STL also examines teachers' leadership practices and self-efficacy in diversity and equity. The STL was sent to all non-charter public school teachers in the U.S. state of Nevada. Over 1200 teachers responded and their data provided a firm empirical foundation for examining the psychometric qualities of the STL. Evidence indicates a five-factor structure with acceptable psychometric qualities.

Understanding teacher leadership: development and testing of the Teacher Leadership Survey

Educational decentralization has been a trend in the United States since the mid-1980s with a particular aim to move away from a focus on individual leaders and hierarchical structures that too often result in teachers feeling overburdened, disenchanted, and alienated (Evans, 1996; Frymier, 1987). Shared governance has instead become more prevalent, whereby leadership is shared among a distributed group of professionals (Hallinger & Kovačević, 2021; Harris, 2003). One manifestation of this trend is the growing focus on teacher leadership (TL). Interest in TL as an academic field has grown in the last few decades (Pan et al., 2023). Even with this growth in research in the field of TL, there remains a great deal of uncertainty around TL and the nature of this important construct.

The benefit of TL to schools has become increasingly clear through improvement in job satisfaction, teacher buy-in for school change, and, notably, student achievement (Nguyen et al., 2020; Schott et al., 2020; Tsai, 2015; York-Barr & Duke, 2004). However, there continues to be a need to further understand the nature of TL. There is a need for an instrument that can systematically collect data on both teachers'

experiences with TL and their feelings of self-efficacy towards TL. This paper reports on the creation and validation of the Teacher Leadership Survey. It seeks to answer the following research questions.

- (1) How well does the Survey Teacher Leadership represent five aspects of teacher leadership: personal growth, professional development leadership, collaboration, shared leadership, and equity leadership?
- (2) What are the psychometric properties of the Survey of Teacher Leadership?

1. Review of literature

1.1. Teacher leadership

While TL has been shown to be important to school, teacher, and student outcomes, it remains an area ill-defined by literature. Definitions of TL have evolved over time encompassing everything from the first wave of traditional classroom roles like department chair, to the second wave of roles leading out of the classroom (such as team leaders, curriculum developers, reading specialists, etc.), and now to the third

* Corresponding author. 4505 S. Maryland Parkway Box 453005 Las Vegas, NV, 89135, USA.

E-mail addresses: peter.wiens@unlv.edu (P.D. Wiens), vanessa.vongkulluksn@unlv.edu (V. Vongkulluksn), jori.beck@odu.edu (J. Beck), bpearson@ccea-nv.org (B. Pearson).

<https://doi.org/10.1016/j.ssaho.2025.101738>

Received 31 July 2024; Received in revised form 16 June 2025; Accepted 25 June 2025

Available online 28 June 2025

2590-2911/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

wave of TL focused on teachers as agents of school change in and out of the classroom (Silva et al., 2000) with a movement towards TL as transformational classroom leadership (Pounder, 2006). One popular definition of TL was summarized by Wenner and Campbell (2017), who defined teacher leaders as, “teachers who maintain K-12 classroom-based teaching responsibilities, while also taking on leadership responsibilities outside of the classroom” (p. 140). However, this article takes a broader view of TL in agreement with York-Barr and Duke (2004) and recognizes that TL can be conceptualized and practiced in a variety of different ways. Teacher leaders may no longer have assigned classroom duties such as full-time instructional coaches or mentors. Instead, we define teacher leaders as those teachers who provide leadership in various capacities, but are not evaluative or supervisory. Therefore, a teacher leader can provide instructional support and leadership, but would not be responsible for evaluating another teacher’s performance for the purposes of annual evaluation. Overall, successful TL entails a distribution of leadership roles within the expertise areas of experienced teachers that disestablishes administrative hierarchy in order to model collaboration, create cross-curricular integration, promote collegial encouragement, form consensus among faculty, and display vigilant professionalism (Lambert, 2002).

Individuals enact TL in both formal and informal ways (Berg & Zoellick, 2019; Nguyen et al., 2020). For example, teachers can be formally assigned as professional learning committee leaders or mentors (Cooper et al., 2016). On the other hand, teachers can take on leadership responsibilities even where no formal roles have been assigned (Fairman & MacKenzie, 2015). Whether formal or informal, TL includes a variety of responsibilities, such as mentoring for improved teacher practice, influencing the school for learning effectiveness, bridging gaps between administration and faculty, and contributing to a broader community of teacher leaders (Schott et al., 2020; Tsai, 2015).

1.2. TL behaviors

Researchers have identified many behaviors that constitute TL. The current survey was created based on the work of previous literature in the field of teacher leadership (Cheung et al., 2018; Harrison & Killion, 2007; Poekert et al., 2016; York-Barr & Duke, 2004). TL is a complex concept that can be examined in different ways. One common way to understand TL is through the examination of practices or behaviors in which teacher engage as leader. York-Barr and Duke (2004), in their seminal literature review on teacher leadership, identified 20 practices in teacher leadership that they grouped into six categories including: coordination and management; school or district curriculum work; professional development of colleagues; participation in school change/improvement; parent and community involvement; contributions to the profession; and preservice teacher education. Meanwhile, Cheung et al. (2018), developed a framework of practices in the categories of collaborating with others to improve instruction, providing resources for effective instruction, advocating in service of effective instruction, and modeling effective instruction. On the other hand, Poekert et al. (2016) developed a framework of teacher leadership development that focused on teacher growth including as a teacher, as a researcher, as a leader, and personal growth. Finally, Harrison and Killion (2007) outline ten roles for teacher leaders that summarize the primary responsibilities individuals take on in leadership roles while remaining primarily focused on teaching. While teachers take on a wide variety of roles required differing skills, TL behaviors can be coalesced into specific categories including: professional growth, professional development leadership, collaboration, shared leadership, and equity leadership (Cheung et al., 2018).

1.2.1. Professional growth

Teacher leaders must first be excellent teachers (Wiens & Beck, 2022). Effective teachers rely on a constellation of teaching strategies, dispositions, relational approaches, and curriculum understanding to

engage their students and promote learning. teachers should adopt and implement innovative teaching strategies that improve student learning (Ainley & Carstens, 2018). Effective teachers also use evidence informed instructional practices (Arends & Kilcher, 2010; Rosenshine, 2012). Meanwhile, the best teachers continually reflect on their own teaching (Korthagen & Vasalos, 2005; Zeichner & Liston, 2013) and use these reflective practices in systematic ways to improve classroom practices (Ginns et al., 2001; Jay & Johnson, 2002). An international study of TL found that teachers demonstrate a “strong self-awareness that facilitated meaningful reflection about their work” (Webber, 2023, p. 346). However, teacher leaders do not only focus on their own instruction, but also continually look for ways to support colleagues in their professional learning (Harrison & Killion, 2007).

1.2.2. Professional development leadership

In addition to being effective teachers and engaging in personal growth and development, teacher leaders engage with colleagues in professional learning (Cheung et al., 2018). Many scholars have identified supporting colleagues’ professional learning as a central component of TL (ex. Chen, 2022; Harrison & Killion, 2007; York-Barr & Duke, 2004). Teacher leaders promote colleague professional learning in a variety of formal and informal roles including mentoring, coaching, and modeling instruction. Teacher mentoring is one-on-one support of a teacher in a way that supports their mentee’s development as a teacher and professional (Hobson et al., 2009). Mentoring has primarily been conducted with pairings of a novice teacher mentee and a more experienced mentor (ex. Hobson et al., 2009; Shanks et al., 2022) or a pairing of a preservice teachers with an experienced teacher (ex. Ellis et al., 2020; Hobson & Maxwell, 2020). However, mentoring can occur with teachers across the experience spectrum (Jewell, 2007). While mentoring is fairly broadly defined as supporting teachers in all areas of professionalism, instructional coaching is more focused on developing classroom practices. It has been used as a “mechanism for new teacher induction, ongoing teacher learning, assisting in the implementation of new initiatives, and, most recently in helping teachers understand and adapt their instruction” (Desimone & Pak, 2017, p. 4). Similar to mentoring, coaching focuses on a one-to-one relationship between coachee and coach whereby the coach assists the coachee in building instructional skills (Knight & van Nieuwerburgh, 2012). While this can happen informally, there has also been the growth of formal instructional coach positions in schools in the United States whereby the coach supports all the teachers in the school through one-on-one relationships and through supporting school-wide professional learning (Galey, 2016; Kane & Rosenquist, 2018).

1.2.3. Collaboration

Collaboration among teachers is a foundation aspect of teacher leadership. Teacher leaders actively collaborate with their colleagues in many ways including the sharing of resources and knowledge (Mangin & Ross, 2022). Teachers serve as leaders when they create learning communities with their colleagues offering all teachers better opportunities to improve their practice and create better outcomes for students (Berg et al., 2014). Teacher leaders also collaborate with their colleagues to collect and understand data. This can happen within structured professional learning communities (Farley-Ripple & Burtram, 2014; Muñoz & Branham, 2016), through formal roles as instructional coaches (Marsh et al., 2010, 2015), or informally (Mangin & Ross, 2022). TL is relational (Hunzicker, 2017) and therefore requires that teacher leaders work towards creating a culture of collaboration in their schools (Heikka et al., 2021; Muijs & Harris, 2007).

1.2.4. Shared leadership

Shared leadership moves away from a focus on an individual leader (Fletcher & Käufer, 2003; Lindahl, 2008) and instead posits that schools operate best when administrators and teachers lead collaboratively (Lindahl, 2008; Lovely, 2005). Shared leadership and the building of TL

within school communities has many benefits for principals, teachers, and students. Innovation and positive school change are more likely to occur in schools with institutionalized shared leadership and TL structures and even occurs more readily in schools where TL is informal but present (Printy & Marks, 2006). While teachers' development of leadership identity varies (Hunzicker, 2017; Sinha & Hanuscin, 2017), in schools where TL is flourishing, teachers will begin to see themselves as leaders (Poekert et al., 2016). In a meta-analysis of TL literature, Shen and colleagues (2020) found that teachers collaborating to facilitate improvements in curriculum, instruction, and assessment as well as engaging in policy making were two of the highest components of TL that were associated with student achievement. Beyond the school itself, teacher leaders also create partnerships and develop relationships with families, community organizations and businesses (Crowther et al., 2002).

1.2.5. Equity leadership

Schools are inherently inequitable and prone to hierarchical structures that privilege some while neglecting others. Equitable schools are those where every student as the "unique supports they need to succeed" (Atchison et al., 2017, p. 1). However, schools are incredibly complex organizations and it is impossible for one person to understand and meet the needs of all individuals within the school. The traditional, hierarchical bureaucratic model of schooling reproduces inequity; however, schools that focus on individual student needs and support teachers working together can be a model for challenging the status quo (Darling-Hammond & Friedlander, 2008). To make schools more equitable, power needs to be distributed and teacher voices need to be supported on a systems level (Berg, 2021). Research in TL points to the ability to develop a social justice stance in teacher leaders (Milner et al., 2022) and that teacher leaders can positively influence socioeconomic disparities (Hunzicker, 2022). Teacher leaders can explore and challenge inequity and work collaboratively with others to advocate for socially just schools and classrooms (Bradley-Levine, 2022).

1.3. Self-efficacy

In addition to the behaviors identified above for TL, this survey also assessed teachers' self-efficacy towards completing these actions. A teacher's sense of self-efficacy can be generally described as a summation of a teacher's beliefs towards their own abilities as a teacher, and to what extent they are capable of executing various instructional, management, and engagement practices for the purpose of student achievement (Bandura, 1997). High teacher self-efficacy is correlated with increased teacher resiliency to setbacks, increased student support, greater job satisfaction, higher teacher retention, and overall improved student outcomes (Ainley & Carstens, 2018; Tschannen-Moran & Hoy, 2001). As such, teacher self-efficacy can provide a reasonable prediction of what extent a given individual succeeds professionally as a teacher in terms of both the position itself and student achievement. Recent research has found that TL is strongly correlated with teacher self-efficacy (Akman, 2021; Gümüş et al., 2022; Li & Liu, 2022; Luo et al., 2024; Zarate). While empirical literature has shown an association between teachers' self-efficacy and TL, this study moves the field forward by examining teachers' TL self-efficacy.

1.4. Teacher leadership instruments

The past two decades have seen researchers make progress towards assessing TL in different ways through surveys. There is some confusion in the terminology of TL whereby some researchers have created instruments measuring TL within classrooms (Batista et al., 2016; Tsai, 2017); however, following the definition of TL as described above, TL instruments should be designed to reflect the conception of TL as pertaining to teachers influencing schools beyond their classroom. Previous surveys in TL have adopted various lenses to examine TL. One lens used

to understand TL is to focus on teachers' perceptions of the culture of TL in their school (Chen, 2022; Demir, 2014; Katzenmeyer & Moller, 2009; Wan et al., 2020). In the United States, Katzenmeyer and Moller (2009) developed a survey focusing on teachers' perceptions of school TL climate based on developmental focus, recognition, autonomy, collegiality, participation, open communication, and positive environment. The authors reported strong internal consistency across these factors (0.83–0.93) and the survey has also been used in other cultures and contexts (Acton, 2022; Aliakbari & Sadeghi, 2014; Mack, 2014). In China, Wan and colleagues (2020) developed a survey that consisted of two parts. The first part asked teachers to respond to items related to their identity as leaders and the second part elicited responses about their perceptions of TL at the department level including reflective dialogue, shared leadership and continuous improvement. Demir (2014) developed a survey examining the culture of TL in the school including teacher collaboration, managerial support, and supportive work environment. While the Demir (2014) survey reported strong internal consistency (greater than 0.93) it lacked strong explanation of validity. Meanwhile, Chen (2022) presented perhaps the most rigorously tested instrument in the Teacher Leadership Inventory (TLI). Over five empirical studies the TLI was developed as a five-factor survey including: promoting professional learning, focusing on learning process, encouraging collegial collaboration, engaging in decision-making, and liaising with external affiliations (Chen, 2022).

Other researchers have created instruments focused on the TL practices of teachers (Ghamrawi, 2013; Kilinc et al., 2021; Wan et al., 2020). Ghamrawi (2013) used Harrison and Killion's (2007) ten roles of teacher leaders as a framework for a ten-item survey asking participants to indicate the extent to which they are encouraged to act in those roles. They did not conduct any factor analysis and instead reported the descriptive statistics of each item separately. Meanwhile, Kilinc and colleagues (2021) implemented the Teacher Leadership Scale (TLS: Beycioglu & Aslan, 2010). The TLS was developed and published in Turkish and measures "the leadership behaviors of teachers' colleagues in their schools" (Kilinc et al., 2021, p. 1561). The TLS contains three subscales: institutional improvement, professional improvement, and collaboration among colleagues (Kilinc et al., 2021). As mentioned previously, the second part of Wan et al.'s (2020) survey assess teachers' perceptions of TL at the department level. Finally, Yuet and colleagues (2016) created the Teacher Leadership Competency Scale (TLCS) to assess teachers' perceptions the competency of teachers to enact TL practices. Exploratory factor analysis found four factors of TL competency including: fostering a collaborative culture, facilitating improvement and establishing standards, modeling leadership attributes and skills, and performing as a referral leader (Yuet et al., 2016). Finally, Zarate and colleagues (2023) developed a survey to assess special education teacher leadership. Through exploratory factor analysis, they found that a four-factor model fit their data best including the factors of pedagogical confidence, ethical considerations, collaboration, and assessment practices.

Previously published surveys in TL have made great strides towards understanding the concept. There is consensus that TL is a complex and multi-factor concept. However, there is still need to better understand what behaviors teachers are engaging in *vis a vis* TL. A theoretically and empirically supported instrument would allow the field to better understand how TL is enacted in schools and provide the capacity for comparison across locations and situations. The Survey of Teacher Leadership (STL), presented in this study sought to assess the frequency of teachers' enactment of TL practices as well as their self-efficacy to perform these behaviors. The STL complements existing scales by directly assessing teachers' TL behaviors and associated self-efficacy. Furthermore, the added explicit focus on equity leadership as one component of TL that reflects the need for teachers to respond to increasingly diverse classrooms and develop a social justice stance.

The STL was developed following established measurement creation procedures including survey item creation, expert review, pilot study

(Wiens et al., 2021), and full-scale survey implementation (Groves et al., 2011). The process towards the final STL was iterative with continual revision of items. First, members of the research team reviewed the literature (Cheung et al., 2018; Harrison & Killion, 2007; Poekert et al., 2016; York-Barr & Duke, 2004) and developed a list of TL practices. Then these items were developed into statements for a Likert-type survey. After careful review for clarity among the research team, a scale with 33 items was distributed to prominent TL scholars. Two scholars provided feedback on the STL and items were revised and two items were removed. The new 31-item was pilot tested with 111 undergraduate and graduate students enrolled in university coursework in teacher education (Wiens et al., 2021). At this stage unidimensional analysis was employed whereby all items in each category (ie. frequency and self-efficacy) were combined into one factor ($\alpha = 0.961$). However, a unidimensional analysis is not consistent with the literature on TL. Therefore, further validation was required. The STL was then distributed to a small group of four public-school administrators responsible to developing TL in their districts in Nevada. They provided further feedback on the survey and questions were revised and 29-item instrument used in this study was created.

2. Methods

This paper reports on the results of the survey of 1274 Nevada public-school teachers. Participants self-reported demographics are illustrated in Table 1. Teachers reported an average of 15.75 years of teaching in total and 6.74 years of teaching at their current school. The largest group of teachers were White (72 %). Meanwhile, the largest group of teachers were female (76.2 %). Teachers taught across all grade levels and the largest category for highest degree was a Master’s degree (69.2 %). Over half of teachers reported working in urban schools (51.6 %). Teachers also reported working in low income (72.0 %) and diverse with the largest group being the 0–25 % White (51.7 %). These data compare to the entire Nevada public school teacher workforce with the following: African American (5.6 %), Asian (4.3 %), White (72.7 %), Latinx (12.1 %), Native American (0.9 %), and two or more races (3.1 %) (National Center for Education Statistics, 2024). Research described in this paper was approved by the University Institutional Review Board (#2022-437).

2.1. Instrument

The STL follows a similar survey construction to the Teaching and Learning International Survey (TALIS) 2018 section on Initial Teacher Education (ITE: Ainley & Carstens, 2018). Following the TALIS 2018 example, participants respond to both their self-efficacy and frequency of employing the items. This allows for collecting two types of data for each item. Participants were provided with 29 TL behaviors and asked to (1) indicate how well they can do it (on a 4-point Likert scale, 1 = not at all, and 4 = a lot), and (2) how frequently they engage in these behaviors (on a 6-point Likert scale, 1 = never to 6 = once a week or more). Items covered five subtypes of TL, including professional growth, professional development leadership, collaboration, shared leadership, and equity leadership (Appendix 1). Participants were also asked to provide demographic information about themselves and their schools.

2.2. Data collection

Instead of using a sampling strategy for this study, we defined the target population (Groves et al., 2011) as all public-school teachers in Nevada who weren’t teaching in charter schools. Non-charter public school teachers were selected as the population because they constitute the vast majority of teachers in most school systems. Additionally, the overwhelming majority of research conducted in TL has focuses on non-charter public school teachers. Therefore, this was the most appropriate population to begin full scale testing of the STL. Teacher

Table 1
Participant demographics.

Characteristic	% Affirmative Response
Individual	
Race/Ethnicity	
African American	5.3
Asian	5.3
White	72.0
Latinx	7.4
Native American	0.7
Other	1.7
More than one Race	7.7
Gender	
Female	76.2
Male	22.6
Gender Non-Conforming	0.8
Transgender	0.2
Gender-other	0.7
Teaching Level	
Pre-School	7.1
Elementary	54.1
Middle	29.3
High	32.9
Post-Secondary	4.3
Not Currently Teaching	4.7
Highest Degree	
High School Diploma	0.2
Associate’s Degree	0.1
Bachelor’s Degree	23.0
Master’s Degree	69.2
Educational Specialist Degree	3.2
Terminal Degree	4.3
School/Community	
Urbanicity	
Rural	14.0
Suburban	34.4
Urban	51.6
School SES	
Low Income	72.0
Middle Income	23.9
High Income	4.2
% Students are White/Caucasian	
0–25 %	51.7
26–50 %	26.7
51–75 %	16.3
76–100 %	5.3

emails were retrieved from districts and from school websites. A total of 20,168 emails addresses were collected and survey invitations were sent. Teachers were recruited to participate via email including Qualtrics survey link in the fall of 2022. After the initial invitation, teachers were sent two follow up reminders. Teachers were provided with information about the study and the informed consent form. Afterwards, teachers were asked to take the survey electronically. In all, 1274 teachers completed all or a portion of the survey for a response rate of 6.32 %.

2.3. Analysis approach

To evaluate construct validity and psychometric properties, model-level, scale-level, and item-level evaluations were conducted with confirmatory factor analysis (CFA) and item-response theory (IRT) analyses (Beau & Bowen, 2021; Embretson & Reise, 2000). First, confirmatory factor analysis was performed for each TL subtype, which was appropriate given the theoretically-driven nature of the survey construction. The weighted least squares mean and variance adjusted (WLSMV) estimator featuring mean- and variance-adjusted chi-squared statistic and scaled standard errors was selected to account for the use of Likert scales with 4–6 response categories, which may not sufficiently approximate a continuous scale (Kline, 2015). Compared to the maximum likelihood estimator, the WLSMV estimator provides more

accurate fit indices, parameter estimates, and standard errors when used with metric data with a limited number of response categories (Beauducel & Herzberg, 2006; Finney & DiStefano, 2006; Myers et al., 2011). Item-level missingness range from 2 to 6 % across items and listwise deletion was used. Because the amount of missing data was small, bias and loss of power from the use of partial data is likely inconsequential (Graham, 2009). While the use of Full Information Maximum Likelihood (FIML) was considered, the benefits of using the WLSMV estimator for this type of Likert scale data outweigh those that might be gained from handling missing data through FIML, especially given the small percentage of missingness.

Model fit indices, factor loadings, item communalities, and residual correlations were used to determine construct validity and psychometric properties. Model fit is assessed with non-significant χ^2 , CFI and $TLI \geq 0.95$, as well as $RMSEA \leq 0.05$ for excellent fit and ≤ 0.08 for adequate fit (Browne & Cudeck, 1992; Hu & Bentler, 1999; Marsh et al., 2004). Items with residual correlation $r > | +/ - 0.10 |$ may be more closely related above and beyond their relationship through the common latent factor, which may contribute to poor model fit (Kline, 2015). While there is no strict cutoff for item communality and standardized loadings, $R^2 > 0.50$ and standardized loadings > 0.30 is generally preferred (Beau & Bowen, 2021). Additionally, items with lower item communality and standardized loadings compared to other items in the same scale may indicate poor contribution to the definition of the focal latent factor.

Resulting unidimensional model(s) were then fitted with graded response models under the IRT framework. Overall model fit was assessed with the C^2 RMSEA and SRMR indices (Cai & Monroe, 2014; Maydeu-Olivares & Joe, 2014). Additionally, item-level functioning was examined with the $S-\chi^2$ and RMSEA item fit indices, as well as the resulting slope (a) and threshold (b) parameters (Bean & Bowen, 2021; Embretson & Reise, 2000). Scale-level properties were assessed with the Fisher information function to determine the range of theta scores for which the scale is informative. CFA and IRT procedures were carried out according to the goals of each phase.

The analyses followed two phases. Data was split into two

subsamples, with even-numbered observation designated for phase 1 analysis and odd-numbered observation for phase 2. In phase 1 scale refinement, the goal was to derive at a unidimensional and well-fitting measurement model by identifying poor-fitting and ill-functioning items. Item-level indices as described above were used to flag items that may contribute to model misfit, which were then compared with the overall fit of each measurement model. Our research team consisting of experts in teacher leadership and measurement then qualitatively assessed the flagged items to discern potential qualitative sources of misfit. Subsequently, items were removed from phase 1 models when deemed appropriate, with updated fit indices serving as an initial test of whether item removal resulted in better fitting measurement models. In phase 2 scale validation, the goal is to confirm the fit of the previously derived measurement model, as well as report the resulting item- and scale-properties. As the revised models were fitted to an independent dataset, the final models with items removed were tested without the confirmation bias potentially present in phase 1. Fig. 1 provides an overview of our data analyses across 2 phases.

3. Results

3.1. Phase I: scale refinement

For TL self-efficacy, CFA results demonstrated that some items had low communality estimates and/or high residual correlations (Appendix 2). We reexamined these items to understand their qualitative characteristics. In the professional growth self-efficacy subscale, item pg1 (select and develop curriculum) have very low communality with the latent factor at $R^2 = 0.156$. This means that only about 16 % of the variance in the item response is explained by the latent factor, compared to 46–62 % for the other items. Item pg1 also had the lowest standardized loading at 0.39 compared to 0.68–0.79 for other items, indicating that the relationship between the latent factor and this item was much weaker. Examining the item revealed that the ability for teachers to select and develop curriculum may be due to other external factors such as administrative processes or professional climate, such as how

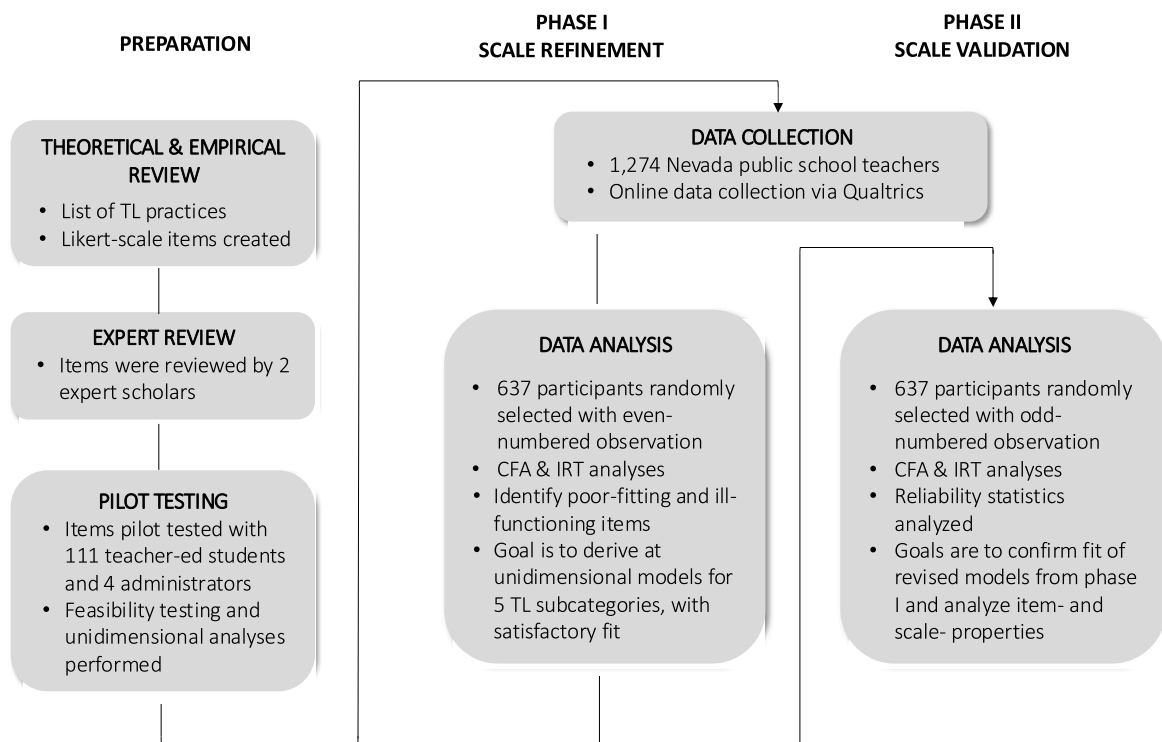


Fig. 1. Scale refinement and validation approach overview.

stringently their school expects teachers to adhere to a set curriculum. Thus, this item may not fit with others which describes teacher-initiated strategies to improve their own instructional practice. This item was removed and the resulting fit indices showed excellent fit ($\chi^2(5) = 18.45$, CFI = 1.00, TLI = 1.00, RMSEA = 0.034). Similar results were found for personal growth frequency. Fit indices supported a well-fitting model, with $\chi^2(5) = 14.53$, CFI = 1.00, TLI = 1.00, RMSEA = 0.034.

For collaboration self-efficacy, col3 (Invite other teachers to observe your class/participate in educational rounds) was found to have high residual correlations with col4 (Go in other classrooms to support colleagues) at $r = 0.10$. Typically, residual correlation $r > | \pm 0.10 |$ suggest that the two items are more closely related above and beyond their relationship through the common latent factor. Inspecting the two items showed that they indeed probe the same collaborative activity teachers may engage in. In subsequent modeling, col3 was selected to be removed since col4 refers to the same collaborative action while maintaining additional flexibility. For example, teachers may go into colleagues' classroom not only to observe, but perhaps also to provide support in other manners. Removal of col3 resulted in improved model fit, with decreased RMSEA and SRMR ($\chi^2(5) = 10.95$, CFI = 1.00, TLI = 1.00, RMSEA = 0.025). The exact fit hypothesis was also supported, with the lower bound of the RMSEA 90 % confidence interval at 0.00. Similarly, for collaboration frequency, the residual correlation between col3 and col4 was $r = 0.116$. The removal of col3 from the collaboration frequency CFA model resulted in a better fitting model ($\chi^2(5) = 7.62$, CFI = 1.00, TLI = 1.00, RMSEA = 0.019).

For shared leadership self-efficacy, sl1 and sl2 have quite high residual correlation at 0.83. Item inspection revealed that the two items are related, with sl1 as "Take part in school-wide decisions" and sl2 as "Work with colleagues for school change". Item sl1 can be seen as a rudimentary form of shared leadership, and reflects less active engagement towards change compared to sl2. Further, whether teachers take part in school-wide decisions may depend heavily on administrative policies. With these considerations, sl1 was removed during model re-specification. Resulting fit indices showed improved model fit, with decreased RMSEA, decreased SRMR, and the exact fit hypothesis now supported ($\chi^2(5) = 8.83$, CFI = 1.00, TLI = 1.00, RMSEA = 0.027). A similar trend was found for shared leadership frequency. Items sl1 and sl2 had residual correlation at $r = 0.106$. Removal of sl1 from the measurement model of shared leadership frequency also resulted in improved fit ($\chi^2(5) = 10.44$, CFI = 1.00, TLI = 1.00, RMSEA = 0.027).

For equity self-efficacy, items eq5 (Identify challenges faced by historically marginalized students in your school and devise solutions) and eq6 (Devise solutions to challenges faced by historically marginalized students in your school) showed high residual correlation at $r = 0.174$. Indeed, these two items tapped the same teacher action to devise solutions to challenges faced by historically marginalized students. Item eq5 can be seen as a double-pronged item, asking teachers regarding their perceived ability to both *identify challenges* and *devise solutions*. Thus, item eq5 was removed from the CFA model during re-specification, resulting in significantly improved model fit ($\chi^2(5) = 10.44$, CFI = 1.00, TLI = 1.00, RMSEA = 0.027). Similar results were found for equity leadership frequency, with eq5 and eq6 having high residual correlation at $r = 0.164$. Without the item, the measurement model for equity leadership frequency displayed excellent fit ($\chi^2(5) = 11.16$, CFI = 1.00, TLI = 1.00, RMSEA = 0.030).

Based on the resulting unidimensional CFA models, IRT graded response models were fitted. All IRT models showed satisfactory overall fit, with C^2 RMSEA < 0.08 and SRMR < 0.04 (Maydeu-Olivares & Joe, 2014). Item-level level analysis also showed RMSEA < 0.06 and most items with non-significant p values associated with the item fit index $S-\chi^2$, demonstrating that the model-based response patterns had good fit to the data (Appendix 3). Items with significant p-values had acceptably small RMSEA value, and could be a result of χ^2 statistics being over-sensitive to minor model misspecifications even in moderately-sized samples (i.e., Schermelleh-Engel et al., 2003). Overall, these results

suggested satisfactory fit between observed responses and expected model-based responses, corroborating CFA results.

3.2. Phase 2: scale validation

Final CFA models fit the data satisfactorily, with RMSEA ≤ 0.07, CFI ≥ 0.97, TLI ≥ 0.98, and SRMR < 0.03 (Table 2). The close fit hypothesis was also retained for all final CFA models for TL self-efficacy and behavioral frequency (lower RMSEA 90 % CI < 0.05). Likewise, all IRT models showed adequate fit, with C^2 RMSEA < 0.08 and SRMR < 0.04.

For professional growth self-efficacy, all standardized loadings were ≥ 0.70, suggesting a strong relationship between the latent factor and item response (Table 3). IRT slope parameters (a) ranged from 2.12 to 2.82, which according to Baker and Kim's classification (2017) demonstrate "very high" capabilities to differentiate levels of the latent trait. Item RMSEA were all sufficiently low, showing good fit between model-based and observed item response. The strongest functioning item was pg3 (*use evidence-informed instructional and teaching practices to improve child outcomes*), with the highest standardized loading and highest slope parameter. Examination of the threshold parameters also reveals the scale's ability to measure a range of personal growth self-efficacy latent trait scores. Each threshold represents the latent trait score (theta) at which the respondent would have 0.50 probability of endorsing above a specific response (Edwards, 2009). For example, item pg2 showed all three thresholds at negative theta scores. That is, it does not take a teacher with very high professional growth self-efficacy to respond "a lot" to the item pg2 – "regularly reflect on teaching practices." A teacher who has a latent trait score of 0.34 standard deviations lower than the mean on professional growth self-efficacy would be predicted to have 0.50 probability to select the highest response category for this item. On the other hand, the threshold parameters for pg4 are shifted to the right on the latent score continuum. A teacher would need to have a professional growth self-efficacy score of 0.52 standard deviations above the mean to be predicted to have 0.50 probability to endorse "a lot" on item pg4 – "develop a systematic and iterative approach to improving classroom practices." Threshold parameters are also sometimes called "difficulty" parameters. In our example, item pg4 might be regarded as the "more difficult" item, requiring higher latent score to endorse the same response category compared to pg2. This mix of threshold values means that there are items on the scale which could discriminate between teachers with lower professional growth self-efficacy scores as well as those with high scores.

For professional growth behavioral frequency, standardized loadings range from 0.61 to 0.77, showing moderate to strong relationship between the item and the latent factor (Table 4). Similar to professional growth self-efficacy, slope parameters displayed "high" to "very high" capability to differentiate levels of the latent trait. Items pg4 and pg6 were the most difficult, whereas pg2 and pg5 were the least difficult items.

For PD leadership self-efficacy, standardized loadings range from 0.68 to 0.86 and communalities from 0.47 to 0.74, indicating strong relationship between the latent factor and item response (Table 3). IRT slope parameters ranged from 1.95 to 4.01, showing "very high" capability to differentiate levels of the latent trait. The strongest item which has the closest connection with the latent trait appeared to be pdl3 (*Serve as a peer or instructional coach either formally or informally*), with the highest standardized loading and communality. The most difficult item with the highest thresholds was pdl2 (*lead professional development initiatives*), whereas the least difficult item was pdl1 (*Mentor other teachers either formally or informally*). For PD leadership behavior, similar results were found with pdl3 as the strongest item, pdl2 as the most difficult item, and pdl1 as the least difficult item (Table 4). Notably, communalities (R^2) estimates were notably lower for PD leadership behavior items compared to corresponding self-efficacy items. This may be because PD leadership behaviors may be somewhat determined by external factors (e.g., leadership opportunities, structural barriers)

Table 2
Validation round CFA fit indices.

Models	χ^2	df	χ^2 p-value	Robust CFI	Robust TLI	Robust RMSEA	RMSEA 90 %CI	SRMR	Communality
Self-Efficacy									
PG: Prof Growth	29.35	5	0.000	0.99	0.99	0.043	[0.028, 0.058]	0.029	0.414–0.592
PDL: PD Leader	9.33	5	0.010	1.00	1.00	0.019	[0.000, 0.037]	0.013	0.468–0.742
COL: Collaboration	42.87	5	0.000	0.99	0.98	0.069	[0.051, 0.089]	0.039	0.345–0.532
SL: Shared Leadership	13.19	5	0.022	1.00	1.00	0.032	[0.011, 0.054]	0.021	0.291–0.652
EQ: Equity	5.91	5	0.315	1.00	1.00	0.010	[0.000, 0.035]	0.012	0.482–0.630
Frequency									
PG: Prof Growth	19.66	5	0.001	1.00	1.00	0.033	[0.019, 0.050]	0.025	0.376–0.591
PDL: PD Leader	12.16	5	0.033	1.00	1.00	0.030	[0.008, 0.052]	0.019	0.259–0.710
COL: Collaboration	26.47	5	0.000	0.99	0.98	0.056	[0.036, 0.078]	0.033	0.290–0.522
SL: Shared Leadership	12.60	5	0.027	1.00	1.00	0.034	[0.010, 0.057]	0.023	0.272–0.613
EQ: Equity	14.89	5	0.011	1.00	1.00	0.035	[0.015, 0.056]	0.022	0.383–0.560

*p < 0.5, *p < 0.01, ***p < 0.001.

Notes. WLSMV estimation was used. χ^2 p-value refers to the result of the χ^2 test which compares the current specified model versus its saturated model. AIC refers to Akaike Information Criterion. SABIC refers to sample-size adjusted Bayesian Information Criterion. CFI refers to Comparative Fit Index. TLI refers to Tucker-Lewis Index. RMSEA refers to Root Mean Square Error of Approximation. SRMR refers to Standardized Root Mean Square Residual.

Table 3
Validation round CFA and IRT-Based item parameters for self-efficacy.

Item	CFA			IRT			
	Std. Loading	R ²	RMSEA	a	b ₁	b ₂	b ₃
Professional Growth							
pg2	0.72	0.52	0.018	2.71	-2.79	-1.47	-0.34
pg3	0.77	0.59	0.000	2.82	-2.30	-1.01	0.06
pg4	0.75	0.56	0.043	2.52	-1.84	-0.52	0.52
pg5	0.64	0.41	0.031	1.94	-3.06	-1.56	-0.31
pg6	0.70	0.50	0.017	2.12	-2.26	-0.75	0.45
PD Leadership							
pd11	0.78	0.61	0.036	2.93	-1.43	-0.44	0.47
pd12	0.70	0.50	0.019	1.98	-0.72	0.30	1.20
pd13	0.86	0.74	0.000	4.01	-0.87	-0.01	0.71
pd14	0.82	0.67	0.028	3.03	-0.89	0.22	0.98
pd15	0.68	0.47	0.014	1.95	-0.61	0.30	1.02
Collaboration							
col1	0.69	0.47	0.000	2.48	-2.45	-1.42	-0.29
col2	0.73	0.53	0.025	2.41	-1.41	-0.32	0.59
col4	0.59	0.35	0.000	1.49	-1.47	-0.19	0.87
col5	0.73	0.53	0.003	2.10	-1.89	-0.73	0.22
col6	0.59	0.35	0.019	1.47	-2.23	-0.77	0.63
Shared Leadership							
sl2	0.74	0.54	0.041	2.27	-1.12	0.10	1.01
sl3	0.70	0.50	0.000	1.97	-0.92	0.62	1.69
sl4	0.54	0.29	0.000	1.31	-0.37	1.25	2.45
sl5	0.65	0.42	0.000	1.81	-1.54	-0.34	0.72
sl6	0.81	0.65	0.021	3.00	-1.15	0.12	0.99
Equity Leadership							
eq1	0.69	0.48	0.023	2.05	-0.63	0.51	1.56
eq2	0.79	0.63	0.014	2.84	-0.92	0.18	1.14
eq3	0.74	0.55	0.036	2.48	-1.15	-0.10	0.70
eq4	0.70	0.49	0.000	2.09	-1.22	-0.03	1.03
eq6	0.71	0.51	0.035	2.08	-1.09	0.34	1.41

above and beyond what is determined by the latent trait.

For collaboration self-efficacy, standardized loadings range from 0.59 to 0.73 and communalities from 0.35 to 0.53, suggesting moderately strong relationship between the latent factor and item response (Table 3). IRT slope parameters ranged from 1.47 to 2.48, which indicate “high” to “very high” capability for trait level differentiation. Two items col2 (*Share research-based practices with colleagues*) and col5 (*Collaborate with colleagues on understanding and use of data*) were the items with the highest standardized loadings and communalities. Item col1 (*Share resources with colleagues*) appeared to be the least difficult item with all thresholds below the mean latent trait score, whereas item col4 (*Go into other classrooms to support colleagues*) was the most difficult item. For collaboration behavioral frequency, col2 was still the strongest item with the highest standardized loading and communality (Table 4). Similar to collaboration self-efficacy, items col1 was the least difficult item and col4 the most difficult. Overall, standardized loadings and communalities were similar to collaboration self-efficacy, suggesting that the strength of association between the latent factor and each item were similar across constructs.

For shared leadership self-efficacy, standardized loadings range from 0.54 to 0.81 and communalities from 0.29 to 0.65, suggesting moderate to strong relationship between the latent factor and item response (Table 3). IRT slope parameters range from 1.31 to 3.00, which reflect “moderate” to “very high” level differentiation capability. The strongest item was sl6 (*Identify school challenges and initiate a process to improve them*), the most difficult item was sl4 (*Create partnerships with community businesses and organizations*), and the least difficult item was sl5 (*Adopt a leadership stance to advocate for self and others*). Similarly, for shared leadership behavioral frequency, the standardized loadings were between 0.52 and 0.78, and the communalities were between 0.27 and 0.61 (Table 4). The IRT slope parameters also range similarly from 1.26 to 2.57. The same pattern of strongest, most difficult, and least difficult item manifested as shared leadership self-efficacy. Interestingly, all items appeared to have a high threshold for the higher response categories. For all but one item, teachers would need to have about 2–3 standard deviations above the mean for a 0.50 probability in endorsing “once a week” about the behavioral frequency referenced in each item.

Table 4
Validation round CFA and IRT-Based item parameters for behavioral frequency.

Item	CFA			IRT					
	Std. Loading	R ²	RMSEA	a	b ₁	b ₂	b ₃	b ₄	b ₅
Professional Growth									
pg2	0.66	0.44	0.016	2.24	-2.93	-2.22	-1.72	-1.29	-0.55
pg3	0.77	0.59	0.016	2.59	-2.26	-1.66	-0.98	-0.52	0.07
pg4	0.76	0.58	0.031	2.43	-1.71	-1.15	-0.55	-0.02	0.77
pg5	0.61	0.38	0.025	1.63	-3.26	-2.65	-1.64	-0.94	0.00
pg6	0.72	0.52	0.027	2.02	-2.12	-1.48	-0.61	0.10	1.19
PD Leadership									
pdl1	0.73	0.53	0.000	2.40	-1.30	-0.84	-0.44	-0.13	0.51
pdl2	0.51	0.26	0.010	1.09	-0.68	0.20	0.96	1.46	2.46
pdl3	0.84	0.71	0.025	3.44	-0.66	-0.26	0.16	0.40	0.95
pdl4	0.71	0.50	0.000	2.00	-0.70	-0.08	0.61	1.14	1.92
pdl5	0.58	0.33	0.000	1.30	-0.30	0.60	1.08	1.46	2.00
Collaboration									
col1	0.71	0.50	0.012	2.51	-2.28	-1.80	-1.23	-0.74	0.05
col2	0.72	0.52	0.019	2.25	-1.28	-0.79	-0.19	0.29	1.10
col4	0.54	0.29	0.000	1.07	-1.46	-0.84	-0.04	0.62	1.65
col5	0.69	0.47	0.000	1.58	-2.07	-1.47	-0.76	-0.21	0.95
col6	0.56	0.31	0.026	1.25	-2.08	-1.35	-0.61	0.10	1.13
Shared Leadership									
sl2	0.69	0.48	0.000	1.85	-0.93	-0.30	0.58	1.09	1.97
sl3	0.64	0.41	0.019	1.58	-0.81	-0.04	0.78	1.38	2.41
sl4	0.52	0.27	0.000	1.26	-0.07	0.89	1.80	2.59	3.29
sl5	0.66	0.43	0.016	1.66	-1.47	-0.85	-0.22	0.25	1.32
sl6	0.78	0.61	0.000	2.57	-1.00	-0.36	0.42	0.94	1.79
Equity									
eq1	0.62	0.38	0.000	1.57	-0.43	0.22	1.09	1.73	2.56
eq2	0.75	0.56	0.032	2.38	-0.75	-0.18	0.37	0.82	1.51
eq3	0.70	0.49	0.026	2.12	-1.08	-0.42	0.10	0.51	1.08
eq4	0.69	0.48	0.018	1.97	-1.13	-0.45	0.17	0.65	1.25
eq6	0.70	0.49	0.000	1.94	-0.94	-0.23	0.52	1.13	1.93

One interpretation of this result may be that teachers may not have sufficient opportunity or time to engage in most shared leadership activities on a frequent basis.

For equity leadership self-efficacy, standardized loadings were between 0.69 and 0.71 and communalities were between 0.48 and 0.63, indicating strong relationship between the latent factor and items on the subscale (Table 3). IRT slope parameters were all above 2, suggesting “very high” capability for trait level differentiation. All items were similar in terms of the strength of relationship with the latent factor, yet item eq2 (*Build alliances that advocate for best practice in socially just classrooms*) had the highest factor loading and communality. The most difficult item was eq1 (*Explore and challenge inequity in the educational system*) and the least difficult was eq3 (*Collaborate purposefully with individuals and groups from other cultures*). For equity leadership behavioral frequency, standardized loadings and communality estimates were generally in the same order of magnitude as corresponding self-efficacy items (Table 4). IRT slope parameters range from 1.57 to 2.38, suggesting “high” to “very high” differentiation capability. The same items were the strongest, and most and least difficult as equity leadership self-efficacy. Item eq1 had a similar threshold profile as shared leadership behavioral frequency items in that high latent trait score would be needed for teachers to endorse “once a month” or “once a week.” Other equity leadership behavioral frequency items required much lower latent trait scores for endorsement of the most frequent response categories.

The resulting STL had satisfactory reliability characteristics, with Cronbach’s α across subscales ranging from 0.77 to 0.88 and McDonald’s ω from 0.81 to 0.89 (Table 5).

After establishing satisfactory psychometric properties of the STL scale, data for the two phases were combined to arrive at descriptive

Table 5
Reliability statistics.

	Cronbach’s α	McDonald’s ω
Self-Efficacy		
Professional Growth	0.84	0.87
PD Leadership	0.88	0.89
Collaboration	0.80	0.84
Shared Leadership	0.82	0.84
Equity Leadership	0.85	0.86
Behavioral Frequency		
Professional Growth	0.83	0.85
PD Leadership	0.80	0.84
Collaboration	0.77	0.81
Shared Leadership	0.79	0.81
Equity Leadership	0.82	0.84

statistics for the whole sample using the validated scale. Mean raw scores showed that sample teachers had the highest TL self-efficacy for professional growth, and lowest self-efficacy for shared leadership (Table 6). Relatedly, the highest TL average behavioral frequency score was also for professional growth and the lowest for shared leadership. All subtypes of TL self-efficacy and behavioral frequency have significant and positive factor correlations at $p < 0.001$. For both self-efficacy and behavioral frequency, the highest factor correlation among subtypes was between shared leadership and equity leadership ($r = 0.74$ and 0.69 , respectively), whereas the lowest was between professional growth and PD leadership ($r = 0.43$ and 0.26 , respectively). The factor correlations between corresponding self-efficacy and behavioral frequency (e.g., professional growth self-efficacy and behavioral frequency) were high

Table 6
Descriptive statistics.

	M _{RAW} (SD)	M _{FAC} (SD)	Pearson's Correlations of Latent Factor Scores										
			1.	2.	3.	4.	5.	6.	7.	8.	9.		
1. pg.se	3.24 (0.64)	0.00 (0.91)	1.00										
2. pdl.se	2.48 (0.88)	0.00 (0.93)	0.43	1.00									
3. col.se	2.96 (0.72)	0.00 (0.90)	0.66	0.63	1.00								
4. sl.se	2.35 (0.76)	0.00 (0.91)	0.55	0.64	0.66	1.00							
5. eql.se	2.39 (0.81)	0.00 (0.92)	0.56	0.55	0.65	0.74	1.00						
6. pg.bf	4.76 (1.04)	0.00 (0.90)	0.77	0.24	0.47	0.36	0.39	1.00					
7. pdl.bf	3.09 (1.37)	0.00 (0.91)	0.26	0.63	0.42	0.34	0.29	0.29	1.00				
8. col.bf	4.12 (1.18)	0.00 (0.89)	0.48	0.37	0.72	0.39	0.41	0.56	0.53	1.00			
9. sl.bf	2.92 (1.18)	0.00 (0.90)	0.36	0.34	0.42	0.67	0.50	0.45	0.50	0.56	1.00		
10. eql.bf	3.01 (1.29)	0.00 (0.91)	0.42	0.30	0.45	0.49	0.73	0.49	0.42	0.59	0.69	1.00	

Notes. All correlations significant at $p < 0.001$. M_{RAW} = mean raw score. M_{FAC} = mean latent factor score. pg = professional growth. pdl = professional development leadership. col = collaboration; sl = shared leadership. eql = equity leadership. se refers to self-efficacy items. bf refers to behavioral frequency items.

across all five subtypes, ranging from $r = 0.63$ for PD leadership to $r = 0.77$ for professional growth. The final suggested items are presented in Table 7.

4. Discussion

Encouraging TL is an important step forward in improving schools and ensuring positive outcomes for both teachers and students (Nguyen et al., 2020; Schott et al., 2020; Tsai, 2015; York-Barr & Duke, 2004). However, TL is an incredibly complex phenomenon that has been criticized as ill-defined (Wenner & Campbell, 2017) and under-studied (Berg & Zoellick, 2019). There continues to be a need to further understand the nature, properties, and structure of TL.

The field of research in TL would benefit from additional measures of various aspects of TL that are psychometrically and theoretically sound. This study reported on the development and implementation of the Survey of Teacher Leadership (STL). The STL was distributed to every public-school teacher in Nevada and data collected provide strong evidence on the nature of TL and the qualities of the STL.

In the creation of the STL, we attempted to capture the complexity of TL by organizing teacher behaviors into five categories. The literature clearly indicates that teachers practice TL in both formal and informal ways (Berg & Zoellick, 2019; Cooper et al., 2016; Fairman & MacKenzie, 2015; Nguyen et al., 2020). Researchers have used various lenses to assess TL in schools. Previous surveys have focused on the culture of TL in a school (Chen, 2022; Demir, 2014; Katzenmeyer & Moller, 2009;

Table 7
Final suggested items.

Self-Efficacy Directions: In your current role, how well can you do the following? (1 = not at all to 4 = a lot)
Behavior Frequency Directions: In your current role, how frequently do you do the following? (1 = never to 6 = once a month)
PROFESSIONAL GROWTH
Regularly reflect on teaching practices
Use evidence informed instructional and teaching practices to improve child outcomes
Develop a systematic and iterative approach to improving classroom practices
Engage in continuous self-improvement
Actively seek out and adopt new and innovative teaching strategies
PD LEADER
Mentor other teachers either formally or informally
Lead professional development initiatives (e.g., workshops, training, PLCs, etc.)
Serve as a peer or instructional coach either formally or informally
Model instruction for colleagues/others
Prepare future teachers (e.g., mentoring field experiences, teaching university-level classes, etc.)
COLLABORATION
Share resources with colleagues
Share research-based practices with colleagues
Go into other classrooms to support colleagues
Collaborate with colleagues on understanding and use of data
Develop a collaborative culture in your school
SHARED LEADERSHIP
Work with colleagues for school change
Challenge the status quo in your school's culture and structures
Create partnerships with community businesses and organizations
Adopt a leadership stance to advocate for self and others
Identify school challenges and initiate a process to improve them
EQUITY
Explore and challenge inequity in the educational system
Build alliances that advocate for best practice in socially just classrooms
Collaborate purposefully with individuals and groups from other cultures
Explore my implicit biases in working with others
Devise solutions to challenges faced by historically marginalized students in your school

Wan et al., 2020), identity as teacher leaders (Wan et al., 2020), the practices of teachers (Ghamrawi, 2013; Kilinc et al., 2021; Wan et al., 2020), competency in TL (Yuet et al., 2016), and special education teacher leadership (Zarate et al., 2023). These studies grouped TL behaviors into different categories (Chen, 2022; Cheung et al., 2018; Poekert et al., 2016). The STL builds on this previous work and organized TL behaviors into five categories: professional growth, professional development leadership, collaboration, shared leadership, and equity leadership. Results from the analysis presented in this study support this structure for TL behaviors as well as teachers' self-efficacy in TL. In the scale validation round, model-level results showed that the specified measurement models for all five TL categories showed excellent fit to the data, with each model's RMSEA satisfying the close-fit hypothesis. The STL provides an empirically and theoretically supported way to organize TL practices. We note that our scale refinement efforts relied on interpretations of our research team comprising of experts in teacher leadership and measurement. However, we recognize that our interpretations may be biased according to our own previous experiences related to the focal teacher leadership constructs. Future research should confirm these interpretations through different data sources, such as through cognitive interviews.

Additionally, this study sought to understand the psychometric properties of the STL. Essentially, we sought to understand if the STL provides a valid and reliable tool for understanding TL. Following established survey development procedures (Groves et al., 2011), the STL went through multiple stages of development including expert review and pilot testing. This paper reports on the collection of a large sample of preK-12 public school teachers. The analysis presented in this paper established strong empirical support for the future use of this measure. IRT results supported satisfactory item- and scale-functioning. The overwhelming majority of items exhibited high to very high capability to different levels of the latent trait. IRT item functioning analysis showed that the items have diverse capabilities to distinguish teachers with both low and high scores on the TL latent traits. The reliability statistics also showed strong evidence of unidimensionality, with all Cronbach's α and McDonald's ω exceeding 0.80, despite the brevity of each subscale (5 items each). Pearson's correlations of latent factor scores also showed positive associations among TL categories, with all correlations <0.80 indicating no evidence of multicollinearity.

Findings from the STL indicate that educational administrators and policy makers need to attend to the multiple facets of TL. Teachers need opportunities and support in professional growth, leading professional development and learning, collaboration with each other, being involved in shared leadership and decision-making, and becoming equity leaders. As TL has been associated with positive outcomes for teachers and students (Nguyen et al., 2020; Schott et al., 2020; Tsai, 2015; York-Barr & Duke, 2004) ensuring that teachers have these opportunities is crucial. Understanding the complex nature of TL also will assist policy makers, professional development leaders, and high education institutions to create training opportunities for teachers that align with the five factors of TL practices and self-efficacy. Building leadership capacity in this well will ensure better outcomes for both teachers and students.

5. Limitations and future research

The purpose of this study was limited to the validation of the STL. In such a tightly focused study, there are several limitations that must be acknowledged and, consequently, associated future research opportunity. The first limitation is the selected population for this study. This study was limited to non-charter public school teachers. While this constitutes the vast majority of teachers in the state, it does need to be considered that the generalizability to other populations has not been tested. Additionally, the population was limited to one state in the U.S. which may not be representative of other locations. Future research needs to extend the use of the STL to different populations both in terms

of school type and geographic location.

Secondly, we acknowledge that school geographic location and student and teacher population may make a difference in teacher experiences with TL (Wiens et al., 2023; Wiens et al., in press). However, this study did not include such variables into the analysis. Additionally, teacher roles within the school may shape their views on TL (Beck et al., 2023). Future research needs to include more information about the teachers, schools, communities, and teacher roles to further understand the experiences of teachers with TL and to further validation the use of the STL. Specifically future research should focus on the experiences of historically marginalized teachers and students *vis a vis* their experiences with TL.

6. Conclusion

In a relatively nascent research field such as the study of TL, empirically and theoretically supported tools are essential for the further development of knowledge. The STL was developed and tested to provide an instrument that can assess teachers practices in TL and their self-efficacy to implement these practices. Data presented in this study indicate that the STL may be well-suited for this task, but more research in a variety of contexts will be required. Much of the literature in TL relies on smaller, qualitative data to understand how TL is enacted by teachers. The STL provides the potential for research in TL to better understand teacher's experiences with TL in a more standardized way and across contexts. TL has gained attention as vital to school success across the globe. The STL provides a measure that can be used to understand how teachers practice TL in various contexts. Additional research studies can implement the STL in additional teacher populations in other geographic locations or different types of schools. As the field of research in TL continues to develop and broaden, the STL may provide a helpful tool to researchers.

CRedit authorship contribution statement

Peter D. Wiens: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Vanessa Vongkulluksn:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis. **Jori Beck:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Brenda Pearson:** Writing – review & editing, Resources.

Data availability statement

Data is available through contacting the authors.

Ethics statement

All research described in this manuscript adheres to the Declaration of Helsinki that informed consent was obtained from all participants. This research project was approved by the Institutional Review Board of the University of Nevada, Las Vegas for Human Subjects Research. No children or vulnerable populations were involved.

Declaration of the use of AI

AI tools were not used in the preparation of this paper.

Funding statement

There is no funding to report for this research.

Declaration of competing interest

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Survey Items

Professional Growth	
pg1	Select and develop curriculum
pg2	Regularly reflect on teaching practices
pg3	Use evidence informed instructional and teaching practices to improve child outcomes
pg4	Develop a systematic and iterative approach to improving classroom practices
pg5	Engage in continuous self-improvement
pg6	Actively seek out and adopt new and innovative teaching strategies
Professional Development Leadership	
pd11	Mentor other teachers either formally or informally
pd12	Lead professional development initiatives (e.g., workshops, training, PLCs, etc.)
pd13	Serve as a peer or instructional coach either formally or informally
pd14	Model instruction for colleagues/others
pd15	Prepare future teachers (e.g., mentoring field experiences, teaching university-level classes, etc.)
Collaboration	
col1	Share resources with colleagues
col2	Share research-based practices with colleagues
col3	Invite other teachers to observe your class/participate in educational rounds
col4	Go into other classrooms to support colleagues
col5	Collaborate with colleagues on understanding and use of data
col6	Develop a collaborative culture in your school
Shared Leadership	
sl1	Take part in school-wide decisions
sl2	Work with colleagues for school change
sl3	Challenge the status quo in your school's culture and structures
sl4	Create partnerships with community businesses and organizations
sl5	Adopt a leadership stance to advocate for self and others
sl6	Identify school challenges and initiate a process to improve them
Equity Leadership	
eq1	Explore and challenge inequity in the educational system
eq2	Build alliances that advocate for best practice in socially just classrooms
eq3	Collaborate purposefully with individuals and groups from other cultures
eq4	Explore my implicit biases in working with others
eq5	Identify challenges faced by historically marginalized students in your school and devise solutions
eq6	Devise solutions to challenges faced by historically marginalized students in your school

Note. All items were used to measure both self-efficacy and behavioral frequency. Teachers were asked to consider each item relative to self-efficacy with the stem: "In your current role, how well can you do the following?", from 1 = not at all to 4 = a lot. They also considered each item relative to behavioral frequency with the stem: "In your current role, how frequently do you do the following?", from 1 = never to 6 = once a month. Highlighted cells are items that were removed in the scale refinement phase.

Appendix 2. Refinement Round CFA Fit Indices

Models	χ^2	df	χ^2 p-value	Robust CFI	Robust TLI	Robust RMSEA	RMSEA 90 %CI	SRMR	Communality
Self-Efficacy									
PG: all items	22.28	9	0.008	1.00	1.00	0.028	[0.013,0.042]	0.025	0.156–0.623
PG: no pg1	18.45	5	0.002	1.00	1.00	0.034	[0.018,0.051]	0.025	0.463–0.637
PDL: all items	8.702	5	0.122	1.00	1.00	0.018	[0.000,0.017]	0.013	0.537–0.687
COL: all items	45.6	9	0.000	0.99	0.99	0.050	[0.036, 0.065]	0.032	0.385–0.523
COL: no col3	10.95	5	0.052	1.00	1.00	0.025	[0.000, 0.045]	0.017	0.435–0.477
SL: all items	46.83	9	0.000	1.00	1.00	0.044	[0.032, 0.057]	0.028	0.336–0.668
SL: no sl1	8.83	5	0.116	1.00	1.00	0.020	[0.000, 0.041]	0.016	0.333–0.639
EQ: all items	115.31	9	0.000	0.99	0.98	0.081	[0.068, 0.094]	0.049	0.419–0.695
EQ: no eq5	12.55	5	0.028	1.00	1.00	0.027	[0.008, 0.046]	0.018	0.419–0.568
Frequency									
PG: all items	12.05	9	0.210	1.00	1.00	0.016	[0.000, 0.038]	0.026	0.064–0.587
PG: no pg1	14.53	5	0.013	1.00	1.00	0.034	[0.014, 0.055]	0.030	0.321–0.584
PDL: all items	17.29	5	0.004	1.00	1.00	0.040	[0.020, 0.061]	0.022	0.366–0.607
COL: all items	51.25	9	0.000	0.99	0.98	0.062	[0.046, 0.079]	0.040	0.275–0.498
COL: no col3	7.62	5	0.178	1.00	1.00	0.019	[0.000, 0.045]	0.017	0.291–0.538
SL: all items	53.10	9	0.000	0.99	0.98	0.057	[0.043, 0.073]	0.037	0.242–0.556
SL: no sl1	10.44	5	0.064	1.00	1.00	0.027	[0.000, 0.050]	0.020	0.224–0.595
EQ: all items	114.49	9	0.000	0.98	0.97	0.095	[0.080, 0.110]	0.056	0.321–0.730
EQ: no eq5	11.16	5	0.048	1.00	1.00	0.030	[0.002, 0.054]	0.021	0.335–0.554

*p < 0.5, *p < 0.01, ***p < 0.001.

Notes. WLSMV estimation was used. χ^2 p-value refers to the result of the χ^2 test which compares the current specified model versus its saturated model. AIC refers to Akaike Information Criterion. SABIC refers to sample-size adjusted Bayesian Information Criterion. CFI refers to Comparative Fit Index. TLI refers to Tucker-Lewis Index. RMSEA refers to Root Mean Square Error of Approximation. SRMR refers to Standardized Root Mean Square Residual.

Appendix 3. Refinement Round IRT-Based Item Fit Parameters

Item	Self-Efficacy				Frequency			
	S-X ²	df	p (S-X ²)	RMSEA	S-X ²	df	p (S-X ²)	RMSEA
pg2	16.77	15	0.333	0.014	80.10	42	0.000	0.040
pg3	10.88	13	0.621	0.000	61.50	40	0.016	0.031
pg4	23.05	14	0.059	0.033	62.26	40	0.014	0.031
pg5	17.98	13	0.158	0.025	49.72	41	0.165	0.019
pg6	16.22	16	0.438	0.005	36.99	43	0.729	0.000
pd11	21.58	17	0.201	0.021	70.09	48	0.020	0.029
pd12	30.55	21	0.081	0.028	78.52	65	0.121	0.019
pd13	9.45	18	0.948	0.000	83.88	55	0.007	0.030
pd14	31.03	19	0.040	0.033	57.88	56	0.406	0.008
pd15	37.70	22	0.020	0.035	66.52	66	0.459	0.004
col1	37.77	15	0.001	0.051	49.85	45	0.287	0.014
col2	16.27	17	0.505	0.000	47.68	51	0.606	0.000
col4	21.42	20	0.373	0.011	55.05	59	0.622	0.000
col5	28.06	20	0.108	0.026	72.95	57	0.076	0.022
col6	23.88	19	0.201	0.021	65.08	62	0.370	0.009
sl2	18.04	18	0.453	0.002	49.76	57	0.741	0.000
sl3	27.56	21	0.153	0.023	52.40	57	0.648	0.000
sl4	24.19	22	0.337	0.013	54.54	53	0.416	0.007
sl5	23.50	18	0.172	0.023	90.77	55	0.002	0.034
sl6	22.54	17	0.165	0.023	56.35	49	0.219	0.016
eq1	33.33	21	0.043	0.032	59.51	62	0.566	0.000
eq2	27.17	20	0.130	0.025	53.84	55	0.519	0.000
eq3	26.03	18	0.099	0.028	66.09	56	0.168	0.018
eq4	26.61	19	0.114	0.026	51.00	60	0.790	0.000
eq6	14.82	20	0.787	0.000	54.60	58	0.602	0.000

References

Acton, K. S. (2022). Environmental teacher leadership: Overcoming barriers posed by school culture, school structure, and the principal. *International Journal of Leadership in Education*. <https://doi.org/10.1080/13603124.2022.2032369>

Ainley, J., & Carstens, R. (2018). Teaching and learning international survey (TALIS) 2018 conceptual framework. *OECD education working papers no. 187*. Organisation for Economic Co-operation and Development. <https://doi.org/10.1787/799337c2-en>

Akman, Y. (2021). The relationships among teacher leadership, teacher self-efficacy and teacher performance. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science]*, 14(4), 720–744.

Aliakbari, M., & Sadeghi, A. (2014). Iranian teachers' perceptions of teacher leadership practices in schools. *Educational Management Administration & Leadership*, 42(4), 576–592. <https://doi.org/10.1177/1741143213510500>

Arends, D., & Kilcher, A. (2010). *Teaching for student learning: Becoming an accomplished teacher*. Routledge.

Atchison, B., Diffey, L., Rafa, A., & Sarubbi, M. (2017). Equity in education: Key questions to consider. *Education Commission of the states*.

Baker, F. B., & Kim, S.-H. (2017). *The basics of item response theory using R*. Springer.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Batista, A., Weber, L., & de Toni, P. (2016). Unidimensionalidade das Escalas do Inventário de Estilos de Liderança de Professores – IELP (Unidimensionality of the Scales of the Teacher Leadership Styles Inventory – TLSI). *Trends in Psychology*, 24(3), 843–853.

Bean, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, 18(6), 597–618. <https://doi.org/10.1080/26408066.2021.1906813>

Beau, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, 18(6), 597–618.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, 13(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Beck, J. S., Hinton, K., Wiens, P., & Butler, B. (2023). A mixed methods investigation of administrators' and teachers' perceptions of teacher leadership. *School Leadership & Management*, 44(3), 271–291. <https://doi.org/10.1080/13632434.2023.2298811>

Berg, J. H. (2021). To lead for equity, learn from teachers. *The Learning Professional*. <https://learningforward.org/journal/leading-for-equity/to-lead-for-equity-learn-from-teachers/>.

Berg, J. H., Carver, C. L., & Mangin, M. M. (2014). Teacher leader model standards: Implications for preparation, policy, and practice. *Journal of Research on Leadership in Education*, 9(2), 195–217. <https://doi.org/10.1177/1942775113507714>

Berg, J. H., & Zoellick, B. (2019). Teacher leadership: Toward a new conceptual framework. *Journal of Professional Capital and Community*, 4(1), 2–14. <https://doi.org/10.1108/JPCCC-06-2018-0017>

Beycioglu, K., & Aslan, B. (2010). Öğretmen liderliği ölççeği: Geçerlik ve güvenilirlik çalışması. *İlköğretim Online*, 9(2), 2–13.

Bradley-Levine, J. (2022). *Promoting teacher advocacy as critical teacher leadership*. Routledge.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESST Report 839)*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chen, J. (2022). Understanding teacher leaders' behaviours: Development and validation of the teacher leadership Inventory. *Educational Management Administration & Leadership*, 50(4), 630–648.

Cheung, R., Reinhardt, T., Stone, E., & Warren Little, J. (2018). Defining teacher leadership: A framework. *Phi Delta Kappan*, 100(3), 38–44.

Cooper, K. S., Stanulis, R. N., Brondyk, S. K., Hamilton, E. R., Macaluso, M., & Meier, J. A. (2016). The teacher leadership process: Attempting change within embedded systems. *Journal of Educational Change*, 17, 85–113. <https://doi.org/10.1007/s10833-015-9262-4>

Crowther, F., Kaagen, S. S., Ferguson, M., & Hann, L. (2002). *Developing teacher leaders: How teacher leadership enhances school success*. Corwin Press.

Darling-Hammond, L., & Friedlander, D. (2008). Creating excellent and equitable schools. *Educational Leadership*, 65(8), 14–21.

Demir, K. (2014). Teacher leadership culture scale: A validity and reliability study. *İlköğretim Online*, 13(2), 334–344.

Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory into Practice*, 56(1), 3–12. <https://doi.org/10.1080/00405841.2016.1241947>

- Edwards, M. C. (2009). An introduction to item response theory using the need for cognition scale. *Social and Personality Psychology Compass*, 3(4), 507–529. <https://doi.org/10.1111/j.1751-9004.2009.00194.x>
- Ellis, N. J., Alonzo, D., & Nguyen, H. T. M. (2020). Elements of a quality pre-service teacher mentor: A literature review. *Teaching and Teacher Education*, 92, Article 103072. <https://doi.org/10.1016/j.tate.2020.103072>
- Embretson, S., & Reise. (2000). *Item response theory for psychologists*. Erlbaum Associates.
- Evans, R. (1996). *The human side of school change*. Jossey-Bass.
- Fairman, J. C., & MacKenzie, S. V. (2015). How teacher leaders influence others and understand their leadership. *International Journal of Leadership in Education*, 18(1), 61–87. <https://doi.org/10.1080/13603124.2014.904002>
- Farley-Ripple, E. N., & Burrtram, J. L. (2014). Developing collaborative data use through professional learning communities. *Studies In Educational Evaluation*, 42, 41–53. <https://doi.org/10.1016/j.stueduc.2013.09.006>
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. *Structural Equation Modeling: A Second Course*, 10(6), 269–314.
- Fletcher, J. K., & Käufer, K. (2003). Shared leadership: Paradox and possibility. In C. L. Pearce, & J. A. Conger (Eds.), *Shared leadership: Reframing the hows and whys of leadership* (pp. 21–47). Sage.
- Frymier, J. (1987). Bureaucracy and the neutering of teachers. *Phi Delta Kappan*, 69(1), 8–14. <https://www.jstor.org/stable/20403522>.
- Galey, S. (2016). The evolving role of instructional coaches in U.S. policy contexts. *The William & Mary Educational Review*, 4(2), 54–71. <https://scholarworks.wm.edu/wmer/vol4/iss2/11>.
- Ghamrawi, N. (2013). In principle, it is not only the principal! Teacher leadership architecture in schools. *International Education Studies*, 6(2), 148–159.
- Ginns, I., Heirdsfield, A., Atweh, B., & Watters, J. J. (2001). Beginning teachers becoming professionals through action research. *Educational Action Research*, 9(1), 111–133. <https://doi.org/10.1080/09650790100200140>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.
- Gümüş, S., Kılınc, A. C., & Bellibas, M. S. (2022). The relationship between teacher leadership capacity at school and teacher self-efficacy: The mediating role of teacher professional learning. *School Leadership & Management*, 42(5), 478–497. <https://doi.org/10.1080/13632434.2022.2123791>
- Hallinger, P., & Kovačević, J. (2021). Science mapping the knowledge base in educational leadership and management: A longitudinal bibliometric analysis, 1960 to 2018. *Educational Management Administration & Leadership*, 49(1), 5–30. <https://doi.org/10.1177/1741143219859002>
- Harris, A. (2003). Distributed leadership in schools: Leading or misleading? *Management in Education*, 16(5), 10–13. <https://doi.org/10.1177/089202060301600504>
- Harrison, C., & Killion, J. (2007). Ten roles for teacher leaders. *Educational Leadership*, 65(1), 74–77.
- Heikka, J., Pitkaniemi, H., Kettukangas, T., & Hyttinen, T. (2021). Distributed pedagogical leadership and teacher leadership in early childhood education contexts. *International Journal of Leadership in Education*, 24(3), 333–348. <https://doi.org/10.1080/13603124.2019.1623923>
- Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education*, 25(1), 207–216. <https://doi.org/10.1016/j.tate.2008.09.001>
- Hobson, A. J., & Maxwell, B. (2020). Mentoring substructures and superstructures: An extension and reconceptualization of the architecture for teacher mentoring. *Journal of Education for Teaching*, 46(2), 184–206. <https://doi.org/10.1080/02607476.2020.1724653>
- 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.
- Hunzicker, J. (2017). From teacher to teacher leader: A conceptual model. *International Journal of Teacher Leadership*, 8(2), 1–26.
- Hunzicker, J. (2022). Equity-focused teacher leadership: Using equity audits to positively influence socioeconomic disparities. In N. Bond (Ed.), *The power of teacher leaders: Their roles, influence, and impact* (pp. 197–212). Routledge.
- Jay, J. K., & Johnson, K. L. (2002). Capturing complexity: A typology of reflective practice for teacher education. *Teaching and Teacher Education*, 18(1), 73–85. [https://doi.org/10.1016/S0742-051X\(01\)00051-8](https://doi.org/10.1016/S0742-051X(01)00051-8)
- Jewell, M. J. (2007). What does mentoring mean to experienced teachers? A phenomenological interview study. *The Teacher Educator*, 42(4), 289–303. <https://doi.org/10.1080/08878730709555408>
- Kane, B. D., & Rosenquist, B. (2018). Making the most of instructional coaches. *Kappan*, 99(7), 21–25.
- Katzenmeyer, M., & Moller, G. (2009). *Awakening a sleeping giant: Leadership development for teachers* (3rd ed.). Corwin.
- Kilinc, A. C., Bellibas, M. S., & Bektas, F. (2021). Antecedents and outcomes of teacher leadership: The role of teacher trust, teacher self-efficacy and instructional practice. *International Journal of Educational Management*, 35(7), 1556–1571.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Knight, J., & van Nieuwerburgh, C. (2012). Instructional coaching: A focus on practice. *Coaching: An International Journal of Theory, Research and Practice*, 5(2), 100–112. <https://doi.org/10.1080/17521882.2012.707668>
- Korthagen, F., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice*, 11(1), 47–71. <https://doi.org/10.1080/1354060042000337093>
- Lambert, L. (2002). A framework for shared leadership. *Educational Leadership*, 59(8), 37–40.
- Li, L., & Liu, Y. (2022). An integrated model of principal transformational leadership and teacher leadership that is related to teacher self-efficacy and student academic performance. *Asia Pacific Journal of Education*, 42(4), 661–678. <https://doi.org/10.1080/02188791.2020.1806036>
- Lindahl, R. (2008). Shared leadership: Can it work in schools? *The Educational Forum*, 72(4), 298–307. <https://doi.org/10.1080/00131720802361894>
- Lovely, S. D. (2005). Making the leap to shared leadership. *Journal of Staff Development*, 26(2), 16–21.
- Luo, X., Alias, B. S., & Adnan, N. H. (2024). Exploring the interplay between teacher leadership and self-efficacy: A systematic literature review (2013–2024). *Education Sciences*, 14, 990. <https://doi.org/10.3390/educsci14090990>
- Mack, J. (2014). *An examination of teacher leadership in public schools*. University of Southern California [Unpublished doctoral dissertation/master's thesis].
- Mangin, M., & Ross, C. (2022). Informal teacher leadership: Reculturing schools and the profession. In N. Bond (Ed.), *The power of teacher leaders: Their roles, influence, and impact* (2nd ed., pp. 7–21). Routledge. <https://doi.org/10.4324/9781003123972-3>.
- Marsh, J. A., Bertrand, M., & Huguet, A. (2015). Using data to alter instructional practice: The mediating role of coaches and professional learning communities. *Teachers College Record*, 117(4), 1–40. <https://doi.org/10.1177/016146811511700411>
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. In *Structural Equation modeling* (Vol. 11, pp. 320–341). Routledge. <https://doi.org/10.1207/s15328007sem1103.2>
- Issue 3.
- Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making. *Educational Policy*, 24(6), 872–907. <https://doi.org/10.1177/0895904809341467>
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305–328.
- Milner, H., McCardle, T., Zuccaro, E., & Worlds, M. (2022). Beyond managerial and bureaucratic: Developing teacher leaders for social justice. In N. Bond (Ed.), *The power of teacher leaders: Their roles, influence, and impact* (pp. 144–154). Routledge.
- Muijs, D., & Harris, A. (2007). Teacher leadership in (in) action: Three case studies of contrasting schools. *Educational Management Administration & Leadership*, 35(1), 111–134. <https://doi.org/10.1177/1741143207071387>
- Muñoz, M. A., & Branham, K. E. (2016). Professional learning communities focusing on results and data-use to improve student learning: The right implementation matters. *Planning and Changing*, 47(1/2), 37–46.
- Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in Exercise and Sport: A Monte Carlo approach. *Research Quarterly for Exercise & Sport*, 82(3), 412–423.
- National Center for Education Statistics. (2024). *NTPS state Dashboard, 2020-21 Nevada*. Author <https://nces.ed.gov/surveys/ntps/ntpsdashboard/Dashboard/NV>.
- Nguyen, D., Harris, A., & Ng, D. (2020). A review of the empirical research on teacher leadership (2003-2017). *Journal of Educational Administration*, 58(1), 60–80. <https://doi.org/10.1108/JEA-02-2018-0023>
- Pan, H.-L. W., Wiens, P. D., & Moyal, A. (2023). A bibliometric analysis of the teacher leadership scholarship. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2022.103936>
- Poekert, P., Alexandrou, A., & Shannon, D. (2016). How teachers become leaders: An internationally validated theoretical model of teacher leadership development. *Research in Post-Compulsory Education*, 21(4), 307–329. <https://doi.org/10.1080/13596748.2016.1226559>
- Pounder, P. (2006). Transformational classroom leadership: The fourth wave of teacher leadership? *Educational Management Administration & Leadership*, 34(4), 533–545.
- Printy, S. M., & Marks, H. M. (2006). Shared leadership for teacher and student learning. *Theory Into Practice*, 45(2), 125–132. https://doi.org/10.1207/s15430421tip4502_4
- Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator*, 36(1), 12–39.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schott, C., van Roekel, H., & Tummers, L. G. (2020). Teacher leadership: A systematic review, methodological quality assessment and conceptual framework. *Educational Research Review*, 31, 100–352. <https://doi.org/10.1016/j.edurev.2020.100352>
- Shanks, R., Tonna, M. A., Krøjgaard, F., Paaske, K. A., Robson, D., & Bjerkholt, E. (2022). A comparative study of mentoring new teachers. *Professional Development in Education*, 48(5), 751–765. <https://doi.org/10.1080/19415257.2020.1744684>
- 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. <https://doi.org/10.1016/j.edurev.2020.100357>
- Silva, D. Y., Gimbert, B., & Nolan, J. (2000). Sliding the doors: Locking and unlocking possibilities for teacher leadership. *Teachers College Record*, 102(4), 779–804.
- Sinha, S., & Hanuscin, D. L. (2017). Development of teacher leadership identity: A multiple case study. *Teaching and Teacher Education*, 63, 356–371. <https://doi.org/10.1016/j.tate.2017.01.004>
- Tsai, K. C. (2015). A preliminary meta-analysis of teacher leadership. *Journal of Education and Literature*, 3(3), 131–137.
- Tsai, K. (2017). Development of the teacher leadership style scale. *Social Behaviour and Personality: International Journal*, 45(3), 477–490.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)

- Wan, S. W.-Y., Tse, Y.-S., Tsang, W.-K., Lee, T. W.-K., Wong, Y. Y., Wan, A. W.-T., & Want, W.-P. E. (2020). 'Who can support me?': Studying teacher leadership in a Hong Kong primary school. *Educational Management Administration & Leadership*, 48(1), 133–163.
- Webber, C. F. (2023). Teacher leaders: Cautions and questions. In C. F. Webber (Ed.), *Teacher leadership in international contexts* (pp. 345–348). Springer.
- Wenner, J. A., & Campbell, T. (2017). The theoretical and empirical basis of teacher leadership: A review of the literature. *Review of Educational Research*, 87(1), 134–171. <https://doi.org/10.3102/0034654316653478>
- Wiens, P. D., & Beck, J. S. (2022). But are they good teachers? Examining who takes up teacher leadership and how their instruction differs from their peers. *School Leadership & Management*, 42(4), 381–401. <https://doi.org/10.1080/13632434.2022.2111413>
- Wiens, P., Beck, J., & Schrader, P. G. (2021). Developing and testing a survey to assess teachers' thoughts and practices regarding teacher leadership. *Paper presented at the annual meeting of the Association of Teacher Educators. Conference Online due to COVID-19 pandemic.*
- Wiens, P. D., Calkins, L., & Skousen, J. (2023). Examining teacher leadership in the United States: How do BIPOC teachers and white teachers in urban environments experience shared leadership? *Urban Education*. <https://doi.org/10.1177/00420859231198171>
- Wiens, P.D., Hinton, K., & Pritchard, A. (in press). Examining the practice of teacher leadership across Nevada: How is context associated with practice? *International Journal of Education Policy and Leadership*.
- York-Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research*, 74(3), 255–316. <https://doi.org/10.3102/00346543074003255>
- Yuet, F., Yusof, H., & Mohamad, S. (2016). Development and validation of the teacher leadership competency scale. *Malaysian Journal of Learning and Instruction*, 13(2), 43–69.
- Zarate, K., Hughes, M. T., & Maggin, D. M. (2023). Special education teacher leadership: Toward a teacher leadership framework for professional development. *Exceptionality*, 31(2), 149–163. <https://doi.org/10.1080/09362835.2022.2100393>
- Zeichner, K. M., & Liston, D. P. (2013). *Reflective teaching: An introduction*. Routledge.