Review of Educational Research
October 2019, Vol. 89, No. 5, pp. 785–827
DOI: 10.3102/0034654319866133
Article reuse guidelines: sagepub.com/journals-permissions

The Effect of Principal Behaviors on Student, Teacher, and School Outcomes: A Systematic Review and Meta-Analysis of the Empirical Literature

David D. Liebowitz and Lorna Porter
University of Oregon

Principals are understood to be critical actors in improving teaching and learning conditions in schools; however, relatively little is known about the leadership strategies to which principals should dedicate their time and effort to improve outcomes. We review the empirical literature from 51 studies of principal behaviors and student, teacher, and school outcomes and conduct a meta-analysis of these relationships. Our analysis has three central findings: (1) we find direct evidence of the relationship between principal behaviors and student achievement (0.08–0.16 SD), teacher well-being (0.34–0.38 SD), teacher instructional practices (0.35 SD), and school organizational health (0.72–0.81 SD); (2) we highlight the importance of principal behaviors beyond instructional management as potential tools to improve student achievement outcomes; and (3) the preceding findings are based almost entirely on observational studies because the causal evidence base on school leadership behaviors is nonexistent. We argue our findings suggest value in investing in school leadership capacities. We conclude by discussing opportunities to improve the quality of future research examining the relationship between principal behaviors and student, teacher, and school outcomes.

KEYWORDS: meta-analysis, principals, school leaders, principal behaviors

School principals play a central role in ensuring a positive environment for teaching and learning in schools. Recent educational policy developments, including high-stakes teacher evaluation systems and increased levels of external accountability, have further raised expectations for principals to improve school climate, instructional practice, and student outcomes. Though there are methodologic difficulties in measuring principals' impact on student learning (Austin et al., 2019; Chiang, Lipscomb, & Gill, 2016; Grissom, Kalogrides, & Loeb, 2015), strong evidence suggests that principals are important factors in the variability of student learning gains across schools and over time (Branch, Hanushek, & Rivkin, 2012; Coelli & Green, 2012; Dhuey & Smith, 2014). Despite the perceived importance of the role of the principal, relatively little is known about how principals might best achieve these aims. Principals (defined operationally in our study as school principals and assistant principals, to which we refer hereafter collectively as "principals") are responsible for a dizzying set of responsibilities (e.g., Kraft & Gilmour, 2016; Spillane & Hunt, 2010; Spillane & Lee, 2014), which range from ensuring that hallways are clear of disruption, family members engage in the school's improvement strategy, compliance documents are completed on time, and students demonstrate mastery of complex academic and social skills. Furthermore, principals come to the profession with a range of prior educational and professional experiences (Grissom & Loeb, 2011; U.S. Department of Education, National Center for Education Statistics, 2017) and with widely varying skill sets (Grissom, Bartanen, & Mitani, 2018; Grissom & Loeb, 2011, 2017).

Given the time-constrained nature of their role, principals—and those charged with their preparation and professional growth—would benefit from knowing whether devoting more time to particular tasks or acquiring new skills in specific domains would make them more effective in their jobs. In this systematic review and meta-analysis, we synthesize the empirical literature documenting the relationship between five broad categories of principal behaviors and a set of student, teacher, and school outcomes. We seek to understand whether principals' time allocation and skill development in these categories of behavior relate to improved outcomes and whether relative efficiencies might be gained by focusing on one or more of these behaviors.

Several well-identified studies suggest that variation in principal quality has meaningful effects on student learning outcomes. The strongest identification strategies on principal quality rely on instances in which principals transition from one school to another. Branch et al. (2012) estimate as a lower bound that a 1 standard deviation unit more effective principal predicts a 0.05 standard deviation unit increase in students' rate of achievement growth *per year* in Texas; their upper-bound estimates suggest effects on the order of 0.20 *SD* per year. The Canadian province of British Columbia has offered a particularly fruitful context in which to isolate the causal effects of principals on student learning, as many school districts there implement a policy in which system administrators rotate principals across schools on a periodic basis to spread knowledge and leadership practices. Dhuey and Smith (2014) estimate that a 1 *SD* increase in principals' effectiveness results in between 0.29 to 0.41 *SD* increases in student achievement gains. Coelli and Green (2012) attribute even larger effects, finding that a 1 standard deviation more effective principal will generate an approximately one third standard deviation higher graduation rate and a 1 standard deviation higher English score in British Columbia high schools. In addition to these direct effects on student learning outcomes, there is wide variability in principals' effects on intermediate process outcomes such as teachers' reported satisfaction with their working conditions (Burkhauser, 2017). In fact, principals' ability to attract and retain highly effective teachers is surely an important mechanism for the observed variation in principal effects. Thus, ensuring schools are staffed with effective leaders is a critical tool to improve learning conditions and address educational inequalities.

As policy and practice interest grows in the potential of principals, however, the mechanisms by which their behaviors directly affect outcomes remain unclear. Although the field of education leadership has been widely studied, much of the early empirical research literature sought to design and test the effects of broad leadership styles, for example, *transformative* (e.g., Leithwood & Jantzi, 2000, 2005), *distributed* (e.g., Heck & Hallinger, 2009), and *collaborative* (e.g., Hallinger & Heck, 2010; Heck & Hallinger, 2010) leadership, or to model the general pathways of influence, such as teacher quality or professional climate, through which principals affect school and student outcomes (e.g., Hallinger & Heck, 1996). These studies generally found transformative or collaborative leadership were positively related to process outcomes such as teachers' satisfaction, skills, or sense of efficacy. They frequently did not detect direct effects of leadership styles on student learning outcomes, but modeled small, indirect effects of leadership strategies on student achievement.

More recent work has attempted to estimate the correlational and causal relationship between types of principal preparation pathways or principal characteristics and various school and student outcomes. The most rigorous principal training studies have focused on nontraditional pathways to the principalship, outside of university preparation programs. A multiyear RAND Corporation evaluation of New Leaders, a highly selective, practice-oriented alternative certification pathway, finds that New Leaders graduates produce between 1- and 3-percentile point greater gains on state mathematics and reading assessments for students in their schools compared with time periods when they were not principal of these same schools (Gates, Baird, Doss, et al., 2019; Gates et al., 2014). Though statistically significant, the magnitude of these effects are relatively small and vary substantially in magnitude and direction across district settings. Two studies (Clark, Martorell, & Rockoff, 2009; Corcoran, Schwartz, & Weinstein, 2012) find less promising outcomes for a similar highly selective, alternative principal preparation program in New York, the Aspiring Principals Program (APP). Though the results across the two studies differ somewhat, both find that APP graduates led schools where students experienced no different, and possibly weaker, outcomes than comparison schools whose new principals were trained in non-APP programs. Similarly, Grissom, Mitani, and Woo (2019) find that principal preparation programs' graduates vary in their effectiveness depending on the particular outcome studied. Thus, the contribution of principal training programs to principals' influence on student outcomes appears to be modest at best and presents measurement challenges.

Results of efforts to expand concepts of principal preservice activities to comprehensive approaches to prepare, recruit, select, place, develop, evaluate, and retain principals have shown promise, but such efforts are still in early phases of programmatic development and evaluation. In conjunction with the Wallace Foundation, six large, urban U.S. school districts participated in a series of reforms to improve the principal "pipeline" by developing strong leadership standards, providing preservice training (either internally in alternative programs or through higher education partnerships) aligned to these standards, recruiting widely and hiring selectively, and evaluating and supporting principals according to these standards. A RAND evaluation found that students of school principals hired through this pipeline improved their math and reading scores by 6- and 3-percentile points, respectively, compared with matched comparison principals, and were also more likely to remain in their roles for at least 3 years (Gates, Baird, Master, & Chavez-Herrerias, 2019).

Separate research documents the influence of principal experience or personal demographic characteristics on student and school outcomes. Several studies document that principals' contributions to student learning outcomes increase with their tenure at a school. Béteille, Kalogrides, and Loeb (2012) find that in Miami-Dade County principal turnover negatively affects school performance, and is both more frequent and most detrimental in high-poverty, low-achieving schools. Coelli and Green (2012) find that principals influence graduation rates and English test score outcomes more the longer they remain in their schools. Goldhaber, Holden, and Chen (2019) find that teachers who become principals are more likely to have higher levels of education and to be male, but have no higher licensure scores than those teachers who do not become principals. They find suggestive evidence that principals with higher school-value-added scores are also those who were more effective at raising students' reading and math scores when they were teachers. Recently, Husain, Matsa, and Miller (2018) document that male teachers are more likely to leave schools led by female principals.

Though these lines of research can be instructive, they do not provide direct guidance on the particular actions a current principal might take to improve school and student outcomes. The framework for our analysis is most similar to recent studies that examine whether the ways in which principals allocate their time across school leadership tasks relates to student learning gains (Grissom, Loeb, & Master, 2013; Horng, Klasik, & Loeb, 2010), or how teachers perceive their principals' effectiveness in particular leadership behaviors predicts educational outcomes (Supovitz, Sirinides, & May, 2010). These studies, which examine how particular principal behaviors influence educational outcomes, may provide guidance more directly relevant to practitioners, those who train them, and those who coach and evaluate them.

In this study, we conduct a meta-analysis of the findings of 51 studies identified in a systematic review by estimating the effects of five categories of principal behaviors on student, teacher, and school outcomes. There have been previous efforts to synthesize the effects of principals on schools (Hitt & Tucker, 2016; Leithwood & Sun, 2012; Marzano, Waters, & McNulty, 2005; Osborne-Lampkin, Folsom Sidler, & Herrington, 2015; Robinson, Lloyd, & Rowe, 2008; Sun & Leithwood, 2012, 2015; Witziers, Bosker, & Krüger, 2003). Prior meta-analyses on school leadership have drawn the overwhelming majority of their sample from studies conducted prior to 2001 and use vote counting or effect-size summation estimation techniques (Marzano et al., 2005; Robinson et al., 2008; Witziers et al., 2003). Syntheses of more recent literature have either taken a qualitative approach (Hitt & Tucker, 2016; Osborne-Lampkin et al., 2015) or examined broad leadership styles and practices (Leithwood & Sun, 2012; Sun & Leithwood, 2012, 2015). Our systematic review and meta-analysis extends these previous efforts by applying modern methods of quantitative meta-analysis to studies drawn from the current era of school accountability, with a particular focus on the direct effects of principals' skill and time allocated to particular leadership behaviors.

To preview our findings, we reach three broad conclusions about the relationship between principal behaviors and student, teacher, and school outcomes: First, we find direct evidence of the relationship between principal behaviors and student achievement, teacher well-being, teacher instructional practices, and school organizational health. Second, despite the recent primacy in the educational management field of theories connected to instructional leadership (e.g., Blase & Blase, 2004; Zepeda, 2013), we find evidence that principal behaviors other than instructional management may be equally important mechanisms to improve student outcomes. Third, the preceding findings are based almost entirely on observational studies because the causal evidence base on school leadership behaviors is, to our knowledge, nonexistent. Despite the dearth of research permitting causal inference, we argue that the insights generated through our meta-analysis are suggestive of the value of investing in principal capacity and helpful in guiding future research agendas interested in estimating the causal relationship between principal behaviors and important educational outcomes.

In the sections that follow, we articulate a framework for how principal behaviors influence educational outcomes, describe our systematic review procedures and meta-analytic measures and methods, present our results, and discuss the implications of these for future research, policy, and practice. We argue that our findings highlight the critical importance of expanding the knowledge base about strategies principals can take to improve learning in schools, and the value of investing in school leadership capacity.

Framework for Principal Behaviors' Effects on Student, Teacher, and School Outcomes

In Figure 1, we present a theory of action of the role that principals play in contributing to student outcomes. We note the situated nature of our theory of action in which principals' prior skills, characteristics, and preparation interact with school context (including demographics, social capital, school readiness and more) to influence principals' development of skill and devotion of time to particular leadership areas. Though not explicitly modeled in our diagram for the sake of clarity in focusing on the role of the principal, we recognize the direct effects of school context on our outcomes of interest as well. For our analysis, we focus on the ways in which principals' behaviors influence student, teacher, and school outcomes.

To be sure, there is value in shaping the nature of the principal labor market to improve the recruitment and retention of high-capacity principals, as well as in improving the preservice preparation of principals. Nevertheless, there were over 90,000 public school principals employed in the 2015–2016 school year, 40% of whom had at least 4 years of experience in their current schools, and for whom there was an exit rate from the principalship of only 10% in 2016–2017 (U.S. Department of Education, National Center for Education Statistics, 2017). Thus,

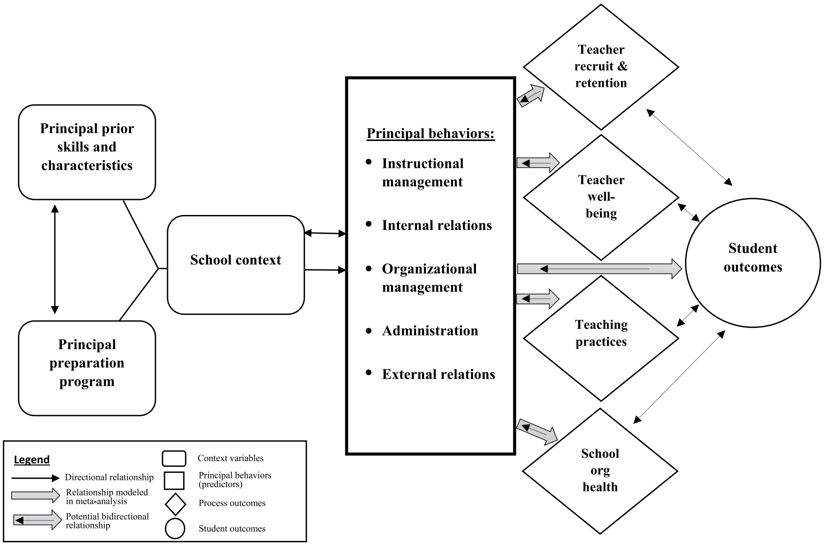

FIGURE 1. *Framework for relationship between principal behaviors and student, teacher, and school outcomes.*

efforts to leverage the potential of principals to improve student outcomes must consider the importance of improving the effectiveness of current principals, in addition to shaping the future labor market.

We draw on the work of Grissom and Loeb (2011) to construct five distinct categories of principal behavior: (1) instructional management, (2) internal relations, (3) organizational management, (4) administration, and (5) external relations. We define each of these categories of principal behaviors more thoroughly in our Measures section below. We theorize that principals' dedication of time and their skill levels in these five categories of leadership behaviors are both directly related to student outcomes, as well as influence student outcomes via their effects on teacher recruitment and retention, teacher well-being, teaching practices, and overall school organizational health (defined below in greater detail). Thus, our theory of action posits both process outcomes at the teacher and school level, as well as final outcomes at the student level. Our framework explicitly notes the bidirectional nature of these relationships. Even in a world in which studies of the relationship between principal behaviors and outcomes are able to isolate the unidirectional, causal component of the relationship, we anticipate that (particularly over time) as the levels and trends in teacher, school, and student outcomes vary so too will principal behaviors.

An important dimension of the framework we present is the relative importance which we assign to the five categories of principal behavior. Recent, widely read practitioner-focused guides to the principalship have focused on the importance of minimizing time spent on managerial and administrative tasks, in favor of instructional leadership activities such as classroom observation, facilitating professional development on curriculum and instruction, and other similar areas. Blase and Blase (2004) argue, simply put, that effective leaders are skilled at and spend time on instructional supervision, and ineffective ones do not. Similarly, Zepeda (2013) argues that "there is a need to elevate the work of principal as an instructional leader" (p. 3). Though most of these guides acknowledge the many competing managerial demands of the principalship, they start from a theory of action that instructional leadership is the linchpin. This has rarely been tested empirically. Grissom and Loeb (2011) find that principals' time on organizational management is more consistently predictive of student learning gains than instructional management. In an intriguing finding, Grissom et al. (2013) find that some instructional leadership tasks such as observation and feedback sessions and managing the school's education program predict improved student learning gains, whereas others such as classroom walk-throughs predict comparatively worse student learning growth.

Our theory of action accords each of the five categories of principal behaviors an equal footing, and we test whether there is a difference in the relative contributions of the behaviors relative to instructional management. The behavioral constructs in the original Grissom and Loeb (2011) study have relatively high levels of internal consistency[1]; thus, we feel reasonably confident that these categories represent underlying dimensions of principal behaviors. However, we recognize that these are broad constructs that may represent quite different activities across studies. Relationships between these five behavioral categories and our outcomes of interest may be sensitive to alternative constructions of school leadership behaviors. We return to these topics in our discussion in an effort to synthesize what conclusions can and cannot be drawn about which behaviors to prioritize.

We also recognize that our models conflate principals' allocation of time to particular behaviors and their skill in effectively executing those leadership actions. It is difficult to disentangle the interrelated nature of these two dimensions of principal behaviors: Do principals spend more time in areas in which they have relative strength, or must they dedicate more time to areas in which their skills are the weakest? For the most part, the studies in our meta-analysis do not permit any efforts to test these two propositions separately. Studies that explicitly examine these two leadership constructs may permit future estimates that examine the interactions and nonlinearity in these relationships (e.g., diminishing returns in increased time allocation with low skills). However, we have no reason to believe that these separate constructs violate principles of monotonicity. In other words, more time spent or more skill in an area should both result in a consistently signed direction of the outcome. Thus, our analysis will look at the joint, linear contributions of skill and time.

Previous Syntheses of Principals' Contributions to Educational Outcomes

Our study is not the first to attempt to synthesize the effects of principals across multiple empirical studies. The first modern synthesis of the relationship between principals and student achievement is Witziers et al.'s (2003) meta-analysis. Emerging from a debate that dominated much of the 1990s research literature on whether principals had a "direct" or "indirect" effect on student achievement (Hallinger & Heck, 1996; Leithwood & Jantzi, 2000), Witziers and colleagues document a small but significant direct effect (0.02–0.04 *SD*) of principals on student achievement across 37 studies published between 1986 and 1996. Witziers's study focuses on *whether* principals influence student achievement, but not on *how* they do so.

The most extensively referenced meta-analysis of principal behaviors is Marzano et al.'s (2005) book: *School Leadership that Works*. Covering studies published between 1980 and 2001, Marzano and colleagues simultaneously seek to document the relationship between school leadership and student outcomes as well as to capture the separate effects of various behaviors by principals. Marzano and colleagues document an average relationship between school leader effectiveness and student learning of 0.25 *SD*. Despite the important contributions to the field, Marzano, Waters, and McNulty's review is both different in nature than ours and limited in its methodology. First, although Marzano and coauthors characterize the 31 constructs examined in the meta-analysis as behaviors, many are associated with attributes, styles of leadership, or overall conditions in the school (e.g., focus, change agent, optimizer, situational awareness) rather than particular leadership behaviors. Additionally, Marzano and colleagues select only one effect size from each study, leaving out of their analysis other potentially important relationships between principals and student outcomes modeled in the primary studies. This has the potential of overstating the strength of the modeled relationship, as the selected effect size may not be representative of all relationships estimated in the study.

Robinson et al. (2008) compare the effects of two different types of leadership styles on student outcomes. Reviewing 22 studies between 1978 and 2006, they find that the average effect size in five studies analyzing "transformative" leadership was 0.11 *SD*, whereas the average effect size in the 12 studies analyzing "instructional" leadership was 0.42 *SD*. An additional five studies included other conceptions of leadership style. They next identify a subset of 12 of the 22 studies in which they are able to analyze particular behaviors within the overall style. They identify positive associations between establishing goals, strategic resourcing, supervising and supporting teaching and the curriculum, and ensuring order with student achievement. Our results align broadly with these findings.

Three recent meta-analyses by Leithwood and Sun examine the effects of transformational leadership. All three meta-analyses, Sun and Leithwood (2012, 2015) and Leithwood and Sun (2012), synthesize 79 doctoral dissertations from various contexts between 1996 and 2008 in a narrative review, a directional count of results, and a meta-analysis. They find, across 45 effect sizes in 20 studies, a mean correlation of 0.12 between overall transformational leadership and student achievement along with stronger relationships with teacher and school process outcomes. They also note slightly stronger relationships to improvements in student achievement for a small group of studies that examine individual transformational leadership practices, such as *Building Collaborative Structures* (three studies) or *Providing Individualized Support* (six studies). Sun and Leithwood (2015) focuses on the relationship between direction-setting leadership practices, an amalgam of eight different leadership practices, and two subcategories, *Developing a Shared Vision* and *Holding High Performance Expectations*. They find strong correlations between the use of direction-setting leadership practices and school and teacher process outcomes but no relationship with student achievement outcomes. All three meta-analyses conclude that future research should focus more on specific leadership practices rather than overarching styles or models.

This encouragement motivates our work, but it is important to note that our review is fundamentally different in substantive and methodological approach.[2] First, these syntheses combine empirical relationships between school leadership practices and outcomes in contexts as different as England, Tanzania, and the Philippines. Second, we use meta-analytic techniques that permit us to account for the nesting of effect sizes within studies and the potential for result correlation within author groups. Most important, we rely on a different conception of school leadership behaviors. Leithwood and Sun (2012) demonstrate that transformational leadership practices such as *Holding High Expectations, Achieving Consensus among Staff*, or *Giving Staff Purpose for their Work* may be valuable techniques in school leadership, but the mechanisms by which principals can accomplish these ends remain unspecified. We distinguish the *practices* (Leithwood and Sun's term) of the transformational leadership framework from principal *behaviors* in which time spent or skill in, for example, instructional management is directly actionable and observable. We contend that this concept of principal behaviors as mechanisms to improve outcomes has benefits for both practitioners and researchers, though it does have limits to which we return in our discussion.

Hitt and Tucker (2016) conduct a qualitative systematic review of frameworks within the empirical research literature that has been used to estimate the relationship between principal behaviors and student outcomes. They identify five domains of principal leadership: establishing and conveying the mission, building professional capacity, creating a supportive organization for learning, facilitating high-quality learning experiences for students, and connecting with external partners, within each of which they specify various actions that principals might take to further these ends. Though we employ a different framework for organizing leadership behaviors tied more directly to specific actions, we find potential overlaps across these categories and the Grissom and Loeb (2011) framework (e.g., instructional management, internal relations, and external relations).

Our framework is most similar to Osborne et al.'s (2015) systematic review of the relationship between principal behaviors and student achievement. Osborne-Lampkin and coauthors use Grissom and Loeb's (2011) principal behavioral definitions and categorize leadership behaviors across 18 quantitative and qualitative studies in a U.S.-based sample between 2001 and 2012. Based on their qualitative review of these studies, they conclude there is evidence indicative of the value of instructional and organizational management and external relations, but not administrative duties, on student achievement. Our study attempts to extend their analysis to a broader set of studies and formalize it in a meta-analysis framework, which allows us to quantify the effects of these behaviors.

This prior analytic work on principals' behaviors as reviewed above and the framework articulated in Figure 1 motivate the following two overarching research questions:

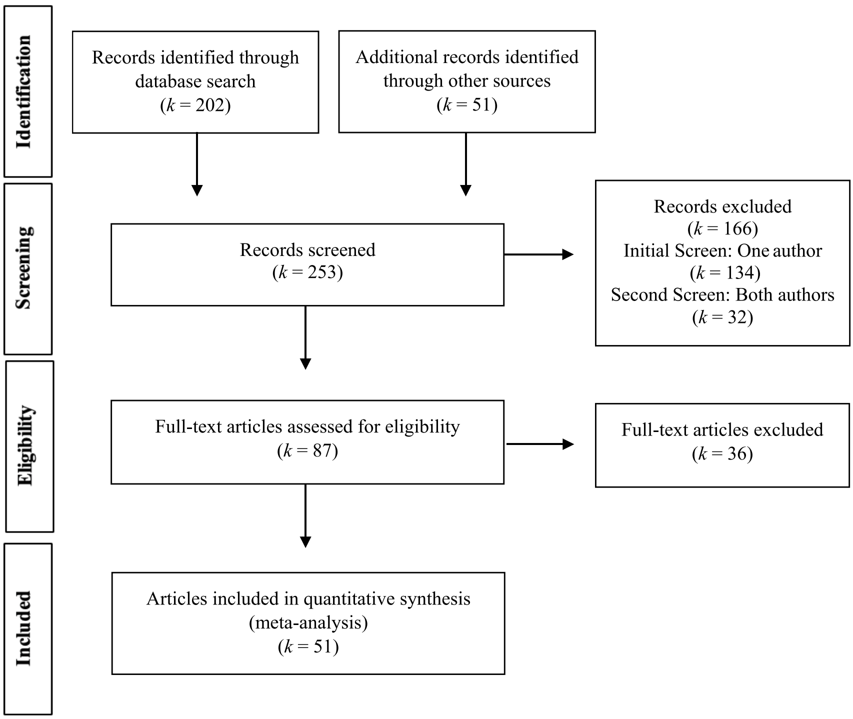

FIGURE 2. *PRISMA flow chart of literature identification, screening, eligibility, and inclusion process.*
Note. Derived from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Moher, Liberati, Tetzlaff, Altman, & the Prisma Group, 2009).

Research Question 1: Are there direct relationships between principal behaviors and their students' achievement, their teachers' well-being and instructional practices, and their school's organizational health outcomes?
Research Question 2: Are principals' instructional management actions more strongly related to differences in student, teacher, and school outcomes than other principal behaviors?

Method

Literature Search Procedures

We outline our literature search process in Figure 2, which follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Moher et al., 2009) guidelines for reporting the identification, screening, eligibility, and inclusion steps of a systematic review. To begin, one author conducted a systematic search and identification of relevant literature using the ERIC, Academic Search Premier, EconLit, ProQuest, and PsycINFO databases, including unpublished and published works. We used the terms ("school leader*" OR "principal") AND ("student achievement" OR "student outcomes" OR "teach* quality" OR "teach* effectiveness" OR "teach* capacity" OR "teach* turnover") to conduct a search of article titles. The search yielded 101 results from ERIC, 120 from Academic Search Premier, 146 from PsycINFO, 94 from ProQuest, and 1 from EconLit. After reviewing the resulting citations for duplicates across databases, we were left with 202 unique articles. The same author then conducted a supplementary title search using Google Scholar with the same terms. As is typical, this search returned tens of thousands of articles, so the author stopped the review in Google Scholar once results were no longer qualitatively relevant. The last search date for all databases occurred on March 28, 2019. Once the database search was complete, we reviewed reference lists from the studies to identify other relevant articles that had not been located in the database search. We also sent requests to leading scholars in the field of school leadership seeking their input on any studies we may have overlooked. These supplemental searches resulted in an additional 51 studies, for a total of 253 studies for initial screening.

Inclusion Criteria

We restricted the inclusion of articles in the systematic review based on three central criteria: the location and timing of the sample, the principal behaviors and outcomes studied, and the research design. We required that studies be conducted in the United States and other high-income, OECD-member countries, published in English, have examined K–12 school samples, and have been completed after 2000.[3] We made this last choice for both substantive reasons and due to the coverage of prior education leadership meta-analyses. First, we theorize that the 2001 reauthorization of the Elementary and Secondary Education Act ushered in a modern era of accountability in the United States (similar accountability shifts occurred in many countries around this time; OECD, 2013), and that the nature of the relationship between principal behaviors and student, teacher, and school outcomes may have shifted in its aftermath. Second, prior education leadership meta-analyses cover studies published primarily before 2001: Marzano et al.'s (2005) sample ends in 2001, Witziers et al.'s (2003) in 1996, and Robinson et al.'s (2008) include only five studies from after 2000. The samples for the more recent Leithwood and Sun (2012, 2015) meta-analyses end in 2008 and only 12 of these dissertations are from after 2005. We also excluded studies for which we could not identify at least one measured relationship between a principal behavior and an outcome of interest as defined in the next section. Finally, we included studies in which the measures of the principal behavior–outcome relationship were quantitative in nature, thus permitting their inclusion in a meta-analysis. Notably, we did not restrict our study sample to studies with experimental or quasi-experimental designs, generally understood to be the only types of study design that permit causal inferences (Shadish, Cook, & Campbell, 2002).

Although our inclusion criteria do not permit us to make causal claims about the effects of principal behaviors, careful methodological design and reporting permit us to provide valuable insight to the education leadership research and practice fields via this meta-analysis. As various meta-analytic expert groups have noted (Higgins et al., 2013; Stroup et al., 2000; Valentine & Thompson, 2013), although meta-analyses of causal studies are superior to observational ones, many fields do not lend themselves easily to exogenous variation in treatment. Consistent with Osborne-Lampkin et al.'s (2015) research synthesis, we found only one study in our sample that meets either of the top two categories of the Institute for Education Sciences levels of evidence (Strong or Moderate). The shortage of studies of education leadership is an important topic to which we return in our discussion. However, we argue that by following the PRISMA (Moher et al., 2009) and the Meta-Analyses of Observational Studies in Epidemiology guidelines (Stroup et al., 2000) for background, search strategy, methods, results, and discussion reporting, our meta-analysis can add meaningfully to the knowledge base on school leadership, while carefully noting the limits of causal inferences we might draw from our findings.

Coding Procedures

After the initial search and screening process, we coded studies to define our final analytic sample and to build our data set. We took a number of steps to maximize coding accuracy, aware of the challenges inherent in categorizing the range of measurement approaches across studies. We engaged in a calibration exercise using previously excluded articles, reconciling differences and refining our codebook before moving to the included articles. Both authors conducted a duplicate blind coding of each of the 87 studies retained for full eligibility review. The intraclass correlation for the unreconciled codes across the 61 studies for which at least one of the authors identified a potential behavior–outcome pair was 0.814.[4] We then reconciled instances of discrepancy, including exclusion or inclusion, through a consensus decision-making process. In the next section, we describe the codes used to characterize aspects of the inclusion criteria and measures in our analysis.

Measures

Principal Behaviors

We required that all included studies model at least one principal behavior as a predictor within the five categories articulated by Grissom and Loeb (2011) to topologize principal behaviors. These five categories are (1) instructional management, (2) internal relations, (3) organizational management, (4) administration, and (5) external relations. Grissom and Loeb empirically derive these five latent leadership constructs from a 42-item principal survey using factor analysis, with an eigenvalue threshold of 1.0. Additional details on the development of these constructs appear in their article. Here, we describe the criteria used to code the articles for each category.

Instructional management. Instructional management broadly encompasses principal behaviors focused on, or linked to, schools' instructional practices and curricular program implementation (Grissom & Loeb, 2011). This includes supporting teachers' instructional practices through teacher evaluation, observation, and feedback, as well as planning teachers' professional development. Instructional management also includes any behaviors related to planning or developing education programs, including what many leadership scales refer to as developing and enacting a schoolwide vision (Avolio & Bass, 1997). Other behaviors include data use related to the school's education program and aspects of program evaluation.

Internal relations. The internal relations code captures the relational aspects of principals' behavior, focused on within-school interpersonal relationships (Grissom & Loeb, 2011). This includes behaviors related to developing and sustaining student and family relationships and attending school activities, as well as handling staff conflicts and engaging informally and socially with staff. Most often in our coding internal relations behaviors are those measuring the relationship between principals' attention to staff relationships and teachers' well-being. For example, Egley and Jones (2005) measured the effects of principals' personally inviting behaviors, including items such as "takes time to talk with faculty and staff about their out-of-school activities" and "makes an intentional effort to treat others with trust and respect" (p. 17), behaviors we coded as internal relations.

Organizational management. Organizational management includes behaviors tied to managing the operational functions of the school related to medium- and long-term strategic goals (Grissom & Loeb, 2011). This code captures principal behaviors such as budgetary tasks, facility planning, and managing noninstructional staff, while at the same time also capturing actions taken to develop a safe school environment, hiring teachers and staff, and networking with other principals. As an example, May, Huff, and Goldring (2012) surveyed principals on how long they spent each day preparing budgets and seeking grants, behaviors we coded as organizational management.

Administration. Grissom and Loeb (2011) define these operational-focused actions as distinct from organizational management as these are "characterized by more routine administrative duties and tasks" (p. 1102). In our coding, we looked for behaviors such as compliance activities, standardized assessment implementation, and school schedule management as well as student service management, student supervision, and managing school attendance. For example, May et al. (2012) relate the time principals spend on building operations with measures of student growth, which we coded as administration.

External relations. The fifth category is external relations, a dimension that captures principals' engagement with stakeholders beyond the school building (Grissom & Loeb, 2011). More specifically, we looked for behaviors such as communication with the district office, community members, partners, or other outside stakeholders as well as fundraising efforts. As one example, we coded total on-the-job time principals spent interacting with stakeholders outside the school as external relations (Leana & Pil, 2006).

Outcomes

We restricted our sample to studies that included at least one outcome measure that captures the following definitions of student-, teacher-, or school-level outcomes.

Student outcomes. We coded for the presence of four forms of student outcomes: (1) the level or growth in student achievement measured on a standardized content-area achievement test, (2) students' grades or grade point average, (3) measures of students' timely progress through their schooling (i.e., graduation or grade repetition rates), and (4) student behavioral outcomes, including measures of student engagement at school, attendance rates, and behavioral and disciplinary measures. Prior to conducting the literature search, we intended to code each of the school progress and behavioral outcomes separately. However, given the small number of studies captured in our initial inclusion search that included any of these types of outcomes, we made the decision to bin outcome types (3) and (4) for the purpose of reporting descriptively the number of each of these categories of outcomes. In the end, we found no studies with outcomes related to students' grades, graduation, or repetition rates in our study sample.

Teacher outcomes. We coded for three different kinds of teacher outcomes: (1) teacher retention, including intent to return or leave the school; (2) teacher well-being, including measures of teacher's emotional state in relation to teaching career, such as satisfaction, engagement, and collegial relationships; and (3) teaching practices which includes observed or reported assessments of teaching practices determined by the primary study framework to be beneficial or harmful in nature.

School outcomes. We coded for two different school outcomes: (1) school organizational health, an expansive code encompassing measures of overall school organizational efficacy and functionality, school climate, including reported safety, student or teacher satisfaction with school practices, and community relationships and (2) principal retention, including intent to return or turnover.

Study Features

Publication year and type of publication. We coded the year of publication and whether the study appeared in a peer-reviewed journal, an institute report, or a dissertation or unpublished working paper. Institute reports were typically those prepared by large-scale contract firms such as the RAND Corporation or Mathematica Policy Research.

Country of study. We coded for whether the sample was U.S. or internationally based.

School level. We created indicators for whether the study included elementary schools (K–5th grade) and secondary schools (6th–12th grades). In cases where studies sampled both, we coded for both.

Unit of analysis. We coded for whether each effect size's unit of analysis was a district, principal/school, teacher, or student.

Sample size. For each effect size, we coded the relevant principal, teacher, or student sample size.

Data source and research group. To test for the potential for hierarchical nesting of study outcomes within studies emerging from the same data sources or research groups, we coded for researcher and sample identification. Based on our initial screen of the 87 studies in our first round review, there were relatively few sample or researcher overlaps. We created identifiers for only three research groups and one data source; we coded all others as "unique."

Research design. We organized studies into two categories: causal or observational. We stipulated that causal studies include a plausibly exogenous variation in treatment, thereby restricting studies in this group to randomized control trials, regression discontinuities, difference-in-differences estimates, or instrumental variable estimation techniques. Given the lack of exogenous variation, we did not code studies employing matching or structural equation modeling as causal. Based on meta-analytic expert guidance (Higgins et al., 2013), we did not code for any qualitative dimension of "study quality" other than the objective assessment of whether the study relied on exogenous variation in treatment to avoid introducing additional layers of nontransparency to the meta-analytic process.

When the article included a predictor–outcome dyad, we extracted the coefficient of interest, recording the outcome category into which it fell with the accompanying predictor from the five principal behaviors. Along with the coefficient value, we extracted the sample size, standard error when available or other values to support robust variance estimation, whether the coefficient described a partial or bivariate correlation, as well as the covariates included in the authors' final model.[5] We coded each separate predictor and outcome pair as a separate effect size, but when authors estimated multiple models of the same predictor–outcome relationship, we coded only the authors' preferred model specification.

In some studies, however, the preferred model was missing necessary information to estimate the effect size or failed to include important parameters of interest. For example, for structural equation models, some studies did not report nonsignificant paths. In such instances, we contacted the authors for these details. When we were unable to obtain such values, we extracted the coefficient from the most complete model, which in some cases was the bivariate correlation matrix. We did not code for indirect effects in structural equation modeling estimates.

Meta-Analytic Methods

We calculate individual effect sizes following standard procedures for calculating mean differences, correlation coefficients, partial correlations, and standardized regression coefficients (Aloe & Thompson, 2013; Borenstein, 2009).[6] We construct effect sizes to reflect the relationship between a one standard deviation change in principals' time allocation or assessed ability in one of the five categories of principal behaviors (our predictors) with a given outcome, expressed in standard deviation units.

We estimate a standard random-effects meta-analytic model where we model effect size outcomes as a sample of results drawn from a distribution of true effects (Borenstein, 2009) produced by a range of different relationships between principal behaviors and outcomes. Concretely, we specify the following model using Borenstein's nomenclature:

$$Y_{ij} = \mu + \zeta_j + \varepsilon_{ij}, \tag{1}$$

where Y_{ij} measures an effect size i for one of the outcomes defined above in study j. Mu, μ, captures the pooled effect of a particular principal behavior. We fit models for each separate principal behavior–outcome pairing separately. Zeta, ζ_j, is the study-level random effect and ε_{ij} is the mean-zero stochastic error term.

We explore how different principal behaviors predict our various effect size outcomes by extending this model to fit a meta-analytic regression:

$$Y_{ij} = \mu + \beta' \overrightarrow{PrinBehav_{ij}} + \zeta_j + \varepsilon_{ij}, \tag{2}$$

where μ, our intercept, captures the effect of instructional management on our outcome of interest, $\overrightarrow{PrinBehav_{ij}}$ is a vector of effect-size-level indicators for a particular principal action and each β captures the marginal effect of a particular leadership behavior, compared with the relationship between instructional management and the outcome.

Our inclusion criteria produce 655 effect sizes across 51 studies. Many studies contribute multiple effect sizes, both across outcomes and for a particular outcome. In the first case, this occurs because studies estimate different outcomes. The second case occurs when studies use multiple measures to assess an outcome, multiple "treatment" groups, or measures of the same kind are estimated at different moments in time. Relying heavily on the clear analytic approach of Kraft, Blazar, and Hogan (2018), we address the clustered nature of our outcomes by estimating all models using robust variance estimation methods (Hedges, Tipton, & Johnson, 2010; Tanner-Smith & Tipton, 2014), which account both for the different precision of estimates across studies as well as for the nonindependence of effect sizes within studies.[7] This method is similar to clustering standard errors; it estimates standard errors asymptotically in small sample sizes with as few as four degrees of freedom (Tipton, 2015). The robust variance estimates give more weight to effect sizes that are estimated with greater precision due to differences in sample size, standard deviations of the predictors and outcomes, predictive power of covariates, and other study characteristics. Effect size estimates from studies that include multiple effect sizes contribute less to the overall estimate of the treatment effect.

We also conduct several sensitivity analyses and tests for publication bias. We compare results across models pooling effect sizes that estimate relationships in elementary and secondary as well as samples based in the United States separately. We examine whether our results are driven by outlying effect sizes by comparing results which include all effect sizes to estimates in which we exclude the top and bottom 5% of outcomes. To test for publication bias, we use a rank-based augmentation technique to identify potential "missing" studies with null or negative findings, impute the estimated values of these studies and reestimate our results. Finally, although our main results pool bivariate and partial correlations, we test for differences in the direction and magnitude of effects across these two model design types.

Results

Study Selection

In Figure 2, we describe the results of the screening, eligibility assessment, and inclusion process. In the initial identification stage, one author conducted a database and reference list search which yielded 202 studies. Over the course of drafting the article, both authors identified an additional 51 studies through supplemental Google Scholar searches, reference list reviews, and expert recommendations for a total starting pool of 253 studies. One author reviewed each study's abstract to screen out ineligible studies, resulting in the exclusion of 134 records. In Appendix Figure S1 (in the online version of the journal), we further outline the study selection process, including exemplars of articles that were excluded at each stage and rationales for exclusion. Next, both authors engaged in a second screening process, reviewing the abstract, sample, and methods sections of the remaining 119 studies to determine whether the measures of both principal behaviors and outcomes matched our analytic framework, and the study's sample was drawn from the United States or other high-income country context in a K–12 setting. In this screening step, we excluded 32 studies. Reasons for study exclusion at this stage included that predictor measures did not capture a particular behavior or action by principals, outcome measures did not align with any of the above categories, or the sample was from an educational system that did not meet our inclusion criteria such as, for example, India. This narrowed our sample to 87 studies.

Our next step was the coding process, in which both authors reviewed each article in full, coding as described above and assessing final eligibility for inclusion. We met periodically to discuss our decisions regarding study eligibility. After reconciling both authors' judgments, we excluded 36 studies at this stage for a number of reasons, primarily related to a study's failure to directly model our defined outcome constructs as a function of particular principal behaviors or incomplete statistical information about the modeled relationship which prevented their inclusion in the meta-analysis.[8] Ultimately, we were left with a sample of 51 studies that included a quantitative measure of principal behavior and a student-, teacher-, or school-level outcome.[9]

Study Characteristics

In Table 1, we report the characteristics for the 51 studies and 655 effect sizes included in the analysis. We present the full list of studies with selected codes in Appendix Table S1 (in the online version of the journal) and make available online our full effect-size-level analytic database (https://scholar.harvard.edu/files/dliebowitz/files/meta_analytic_dataframe_master.csv). The majority of studies appeared in peer-reviewed journals, were conducted in the United States, and published after 2005. There is a fairly even distribution of studies focused on elementary and secondary school samples, as 36 included elementary schools and 34 included secondary schools. Over half of the studies draw on principals as the unit of analysis. Eighteen studies draw on teachers, four on students, and one on the district. Just one study of the 51 uses a research design that permits causal

TABLE 1

Characteristics of studies included in meta-analysis

Code	Count cases (studies)	Proportion cases (studies)
Publication year		
2001–2005	20 (6)	0.03 (0.12)
2006–2010	201 (15)	0.31 (0.29)
2011–2015	372 (24)	0.57 (0.47)
2016–2019	62 (6)	0.09 (0.12)
Unit of analysis		
Student	55 (4)	0.08 (0.08)
Teacher	187 (18)	0.29 (0.35)
Principal	408 (33)	0.62 (0.65)
District	5 (1)	0.01 (0.02)
Publication type		
Peer-review journal	372 (33)	0.57 (0.65)
Technical report	135 (7)	0.21 (0.14)
Dissertation/working paper	148 (11)	0.23 (0.22)
Country of study		
International	74 (9)	0.11 (0.18)
United States	581 (42)	0.89 (0.82)
School level		
Elementary	368 (36)	0.56 (0.71)
Secondary	444 (34)	0.68 (0.67)
Research design		
Observational	654 (50)	1 (0.98)
Causal (exogenous variation in treatment)	1 (1)	0 (0.02)
Principal behavior		
Instructional management	439 (46)	0.67 (0.9)
Internal relations	111 (20)	0.17 (0.39)
Organizational management	50 (9)	0.08 (0.18)
Administration	26 (6)	0.04 (0.12)
External relations	29 (5)	0.04 (0.1)
Outcomes		
Student academic achievement	346 (31)	0.53 (0.61)
Student behavior/attendance	5 (3)	0.01 (0.06)
Teacher retention	12 (4)	0.02 (0.08)
Teacher well-being	57 (10)	0.09 (0.2)
Teaching practices	59 (11)	0.09 (0.22)
Principal retention	6 (2)	0.01 (0.04)
School organizational health	170 (24)	0.26 (0.47)
N cases (*K* studies)	655 (51)	

Note. Unit of analysis, level, behaviors, and outcomes are not exclusive by study, so sum to greater than 100%.

inference; the remaining 50 are observational in nature. We note our surprise at this finding. Our initial scan of the literature revealed a small handful of studies that used causal designs, and our original intent was to compare the findings of these studies to the larger sample of observational studies. However, on closer inspection, we found that studies that, for example, used a random assignment of treatment to evaluate the overall effect of participating in a leadership training program were not designed in such a way as to causally test whether differences in principals' behaviors as a result of participation in the program led to differences in student outcomes (Gates et al., 2014).

All five principal behaviors are present in our final analysis, however two outcomes, instructional management and internal relations, appear more frequently in both the number of articles and effect sizes. As shown in Table 1, principals' instructional management behaviors are included in 90% of articles and represent 67% of effect sizes. Internal relations is the next most prevalent principal behavior, present in 39% of studies and constituting 17% of effect sizes. The remaining effect sizes are fairly evenly distributed across the organizational management, administration, and external relations constructs with each representing between 4% and 8% of total effect sizes.

We identify seven different outcome categories across the included studies. As with principal behaviors, the effect sizes cluster around certain outcomes. Over half of the effect sizes (53%) capture the relationship between leadership behaviors and student academic achievement, whereas just 1% of effect sizes have student behavior as an outcome. Across teacher-level outcomes, both teacher well-being and teacher practices are the most prevalent, each at 9% of total effect sizes. Very few of the studies examine teacher retention (2% of effect sizes). At the school level, school organizational health is the most prevalent outcome (26% of effect sizes), whereas 1% of effect sizes measure the effect of principal behaviors on whether they remain at their school.

Effects of Principal Behaviors on Student, Teacher, and School Outcomes

We begin by assessing the overall distribution of the relationship between principal behaviors and the outcomes of interest in our study. In Figure 3, we present kernel density plots displaying the distribution of relationships between principal behaviors and (1) student achievement, (2) teacher well-being, (3) teaching practices, and (4) school organizational health outcomes. The distribution of the relation between principal behaviors and student achievement effect sizes is approximately normal with a long right tail. The magnitude of effect sizes are relatively narrowly distributed, with an interquartile range between −0.01 *SD* and 0.32 *SD*. The relationship between principal behaviors and teacher well-being effect sizes is similarly distributed with a positive skew and an effect size interquartile range between 0.00 *SD* and 0.51 *SD*. However, the relationship between principal behaviors and teaching practices has an evident bimodal distribution and a substantially larger spread of effect sizes (interquartile range = 0.20–0.84 *SD*). Finally, the relationship between principal behavior and school organizational health effect sizes is roughly normally distributed, but with a much larger interquartile range between 0.20 and 1.18 *SD*. We present density plots for each outcome, by principal behavior in Appendix Figure S2 (in the online version of the journal).

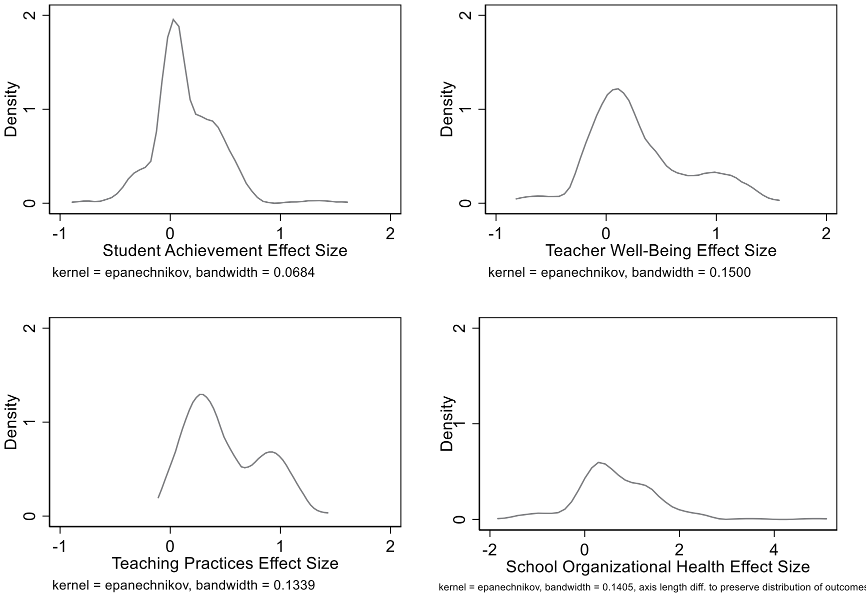

FIGURE 3. *Kernel density of effect sizes of principal behaviors on student achievement, teacher well-being, teacher practices, and school organizational health.*
Note. n = 346 effect sizes for student achievement, 57 for teacher well-being, 59 for teaching practices, and 170 for school organizational health.

In Table 2, we present the main results of our meta-analysis in which we find significant and substantively meaningful relationships between principals' skill and time allocation to instructional management and internal relations on students' achievement in their schools. In column 1 of Table 2, we find across 216 effect sizes nested in 28 studies that measure the relationship between principals' instructional management time allocation and skill and students' achievement a 1 *SD* unit difference in principal instructional management is associated with a 0.11 *SD* difference, 95% confidence interval [CI: 0.03, 0.20]; 95% prediction interval [PI: −0.12, 0.35], in students' achievement.[10] A similar difference in principals' internal relations time and skill is associated with a 0.16 *SD* difference, 95% CI [0.02, 0.31]; 95% PI [−0.21, 0.54], in student achievement scores across 58 effect sizes in 14 studies. We note evidence of a similar magnitude relationship between principals' organizational management, administration, and external relations efforts, but caution that these estimates rely on only between five and eight studies.

We find marginally significant evidence that principals' focus on instructional management and internal relations are associated with higher levels of teacher well-being. Column 2 of Table 2 indicates a 0.34 *SD* positive relationship, 95% CI [−0.04, 0.71], 95% PI [−0.28, 0.96], between well-being effect sizes and principals' efforts to manage instruction and a 0.38 *SD* positive relationship, 95% CI

TABLE 2

Pooled effect size estimates of the effect of principal behaviors on student, teacher, and school outcomes

Principal behavior	Student achievement	Teacher well-being	Teaching practices	School organizational health
Instructional management	0.113* (0.039)	0.339~ (0.161)	0.348** (0.093)	0.812*** (0.172)
n (*k*)	216 (28)	32 (9)	50 (10)	127 (21)
Internal relations	0.163* (0.065)	0.383~ (0.150)	na	0.726* (0.276)
n (*k*)	58 (14)	17 (6)	8 (3)	25 (8)
Organizational management	0.084 (0.051)	na	na	0.781 (0.433)
n (*k*)	33 (8)	4 (2)	0 (0)	10 (4)
Administration	0.084 (0.064)	na	na	na
n (*k*)	20 (5)	1 (1)	0 (0)	2 (2)
External relations	0.081 (0.034)	na	na	na
n (*k*)	19 (5)	3 (2)	1 (1)	6 (2)

Note. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, *n* is the number of effect sizes and *k* is the number of studies. Cells with “na” are not estimated due to insufficient data ($k < 4$).
~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

[−0.01, 0.77], 95% PI [−0.64, 1.41], between well-being effect sizes and principals’ time and skill in building internal relations. In line with its implied effect, we find relatively strong relationships between principals’ focus on instruction-specific support and more effective teaching practices. Across 10 studies and 50 effect sizes in column 3 of Table 2, we note a 0.35 *SD* relationship, 95% CI [0.14, 0.56], 95% PI [−0.27, 0.97], between principals’ instructional management efforts and the quality of teachers’ instructional practices.

We interpret the relationships between principals’ behaviors and school organizational health effect sizes cautiously. We observe large pooled effect sizes, though recognize the strong likelihood that these may be influenced by omitted variable bias in these observational studies. Nevertheless, we observe positive relationships in column 4 of Table 2 between principals’ efforts to improve instruction, 0.81 *SD*, 95% CI [0.45, 1.17], 95% PI [−0.70, 2.32]; support internal relations, 0.73 *SD*, 95% CI [0.07, 1.38], 95% PI [−0.99, 2.44]; and manage their organization on school organizational health effect size outcomes, though only the first two are precisely estimated.

Do Effects Vary by School Level or Country of Study?

We explore the potential that there may be different relationships between principal behaviors and student, teacher, and school outcomes depending on the grade levels of the school in Table 3. We interpret results that rely on parsing our sample of studies cautiously given the already small number of studies that

TABLE 3

Pooled effect size estimates of the effect of principal behaviors on student, teacher, and school outcomes, by school level

Principal behaviors	Student achievement	Teacher well-being	Teaching practices	School organizational health
A. Elementary				
Instructional management	0.067~ (0.035)	0.246 (0.172)	0.283** (0.068)	1.014*** (0.215)
n (*k*)	112 (20)	21 (6)	39 (8)	70 (15)
Internal relations	0.073 (0.054)	0.307 (0.159)	na	1.136~ (0.441)
n (*k*)	20 (6)	14 (4)		7 (4)
Organizational management	0.030 (0.025)	na	na	na
n (*k*)	21 (4)			
Administrative duties	na	na	na	na
n (*k*)				
External relations	na	na	na	na
n (*k*)				
B. Secondary				
Instructional management	0.090 (0.056)	0.314 (0.178)	0.327~ (0.139)	0.437~ (0.203)
n (*k*)	140 (15)	22 (8)	21 (7)	67 (11)
Internal relations	0.165~ (0.090)	0.283 (0.213)	na	0.279 (0.202)
n (*k*)	51 (11)	11 (4)		20 (5)
Organizational management	0.083 (0.055)	na	na	na
n (*k*)	31 (7)			
Administrative duties	0.088 (0.063)	na	na	na
n (*k*)	19 (5)			
External relations	0.075* (0.036)	na	na	na
n (*k*)	18 (5)			

Note. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, n is the number of effect sizes and k is the number of studies. Cells with "na" are not estimated due to insufficient data ($k < 4$).
~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

contribute to each predictor–outcome pairing. For our results highlighting the relationship between principals' instructional management behaviors and student and teacher outcomes, we find little substantive difference across elementary and secondary settings. However, the subsetting of our sample of effect sizes reduces the precision of each estimate. The same is true for other principal behaviors and student and teacher outcomes, though these estimates rely on much smaller samples. Interestingly, however, we find a more modest

TABLE 4

Pooled effect size estimates of the effect of principal behaviors on student, teacher, and school outcomes, United States results only

Principal behaviors	Student achievement	Teacher well-being	Teaching practices	School organizational health
Instructional management	0.133* (0.050)	0.313 (0.209)	0.385* (0.139)	0.962** (0.233)
n (k)	214 (26)	20 (7)	24 (7)	110 (15)
Internal relations	0.163* (0.065)	0.379 (0.186)	na	0.703~ (0.325)
n (k)	58 (14)	12 (5)		22 (7)
Organizational management	0.084 (0.051)	na	na	na
n (k)	33 (8)			
Administration	0.084 (0.064)	na	na	na
n (k)	20 (5)			
External relations	0.081 (0.034)	na	na	na
n (k)	19 (5)			

Note. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, n is the number of effect sizes and k is the number of studies. Cells with "na" are not estimated due to insufficient data ($k < 4$).
~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

relationship between principals' behaviors and organizational health effect sizes at the secondary level than at the primary. Differences in instructional management and internal relations are associated with large differences in measures of organizational health (1.01 and 1.14 *SD*, respectively) at the elementary level; however, these differences are smaller at the secondary levels, and indistinguishable from zero in the latter case. The attenuated relationships that principals' influence have on organizational dynamics at the secondary level is consistent with the literature on the resistance of secondary schools' cultures to change (Fullan, 2001; Hargreaves & Goodson, 2006).

We next explore the extent to which our results are specific to particular national education systems. Given the small number of studies in our sample and the range of countries represented in the non–U.S. studies, we are unable to measure the relationship between principal behaviors and outcomes in non-U.S. contexts. Instead, in Table 4 we present results for a meta-analysis that restricts our sample to only U.S.-based studies. Unsurprisingly, given our largely U.S.-based sample, the results we present in Table 4 essentially mirror those from Table 2, though with slightly stronger evidence of the relationship between principals' instructional management time allocation/skill and student achievement, 0.13 *SD*, 95% CI [0.03, 0.23], 95% PI [−0.18, 0.44]; teaching practices, 0.39, 95% CI [0.04, 0.73], 95% PI [−0.47, 1.24]; and school organizational health, 0.96 *SD*, 95% CI [0.46, 1.46], 95% PI [−1.22, 3.15], effect sizes.[11]

Do Different Principal Behaviors Affect the Same Outcomes Differently? Do the Same Principal Behaviors Affect Different Outcomes Differently?

Of critical interest to a principal practicing in a resource-limited environment is how to allocate the scarcest of resource: time. Thus, although our meta-analysis reveals generally positive correlations between all principal behaviors and our outcomes, principals would benefit from understanding whether some behaviors have a stronger effect than others on improving desired student, teacher, and school outcomes. Ultimately, the most rigorous way to test this sort of question is through a series of well-designed experiments in which principals are randomly selected to allocate more time (or improve their skill) in one dimension or another of school leadership. We discuss future research avenues below. In addition to the above caveats about the observational nature of our study sample with its potential for omitted variable bias and reverse causality, it is possible that comparisons across principal behaviors reflect the multicollinearity between principal ratings across different dimensions. Nevertheless, our data allow us to present some suggestive evidence on the relative value of different principal behaviors. In particular, we examine whether current trends emphasizing instructional management strategies for principals over other behaviors are supported by our primary study data.

We begin by reporting evidence that a principal's focus on instructional management is related to similar magnitude outcomes as a principal's focus on internal relations. In Panel A of Figure 4, we present the relationship between student achievement outcomes related to principals' instructional management and principals' internal relations behaviors for 12 studies in our sample that include both predictor–outcome pairs. We find a strong correlation between the two sets of principal behavior relationships to student achievement outcomes ($r = 0.72$). Similarly, Panel B reveals a near perfect correlation ($r = 0.98$) between studies with measures of both instructional and relational leadership effects on school organizational health outcomes. We next test these ideas more formally in a metaregression framework.

We find suggestive evidence that although principals' instructional management behaviors have important relationships with student achievement, teacher well-being, teaching practices, and school organizational health outcomes, these appear to not be substantively stronger than the relationship between other principal behaviors and the same outcomes. In Table 5, we present results of a metaregression in which we estimate the relationship between different principal behaviors and our student, teacher, and school outcomes. Column 1 of Table 5 shows that consistent with our meta-analytic results in Table 2, across 346 effect sizes in 31 studies there is a significant and moderately sized relationship between principals' time spent and skill level devoted to instructional management and student achievement, 0.13 *SD*, 95% CI [0.03, 0.23], reported in this table as the omitted category intercept. However, we fail to reject the null hypothesis that there are no differences in the direction or strength of the relationships between instructional management and the other four principal behaviors and student achievement. There is some imprecision in the estimates in column 1, which precludes us from stating conclusively that there are no differences in student achievement outcomes by principal behaviors. However, our results suggest that

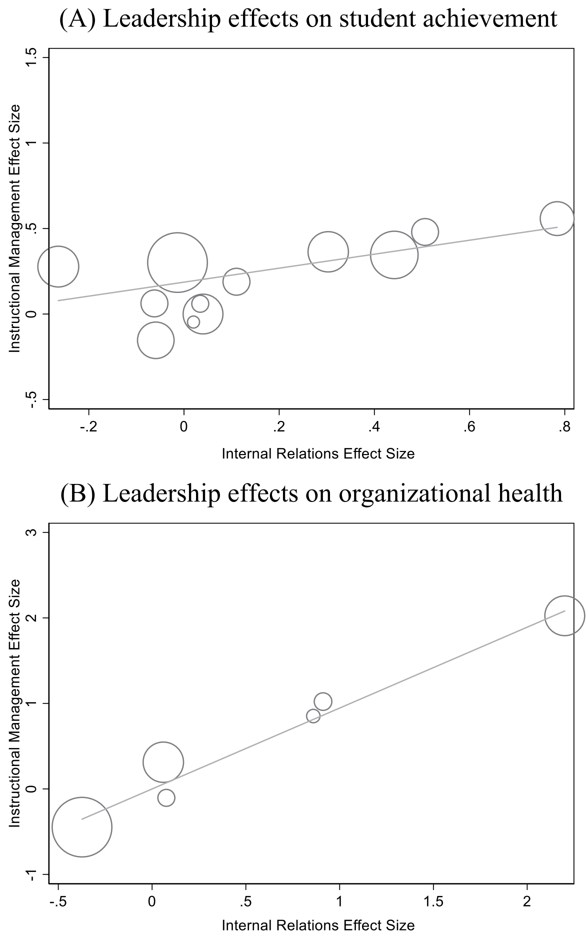

FIGURE 4. *Relationship between instructional management and internal relation behaviors effects on student achievement and organizational health.*
Note. Data points are calculated by averaging effect sizes for a particular principal behavior and outcome within the same study and weighting by the precision of the estimates. $k = 12$ studies for student achievement and 6 studies for organizational health.

in 95 out of 100 instances the summary effect in our study population of the benefits of time spent and skill in, for example, Internal Relations will be no more than 0.12 *SD*s less effective and no more than 0.24 *SD*s more effective at improving student achievement than time spent on and skill in Instructional Management. The other three principal behaviors are similarly bounded in their 95% confidence interval differences from Instructional Management. We also fail to reject the null hypotheses in columns 2, 3, and 4 that instructional management predicts teacher well-being, teaching practice quality, or school organizational health differently

TABLE 5

Meta-regression estimates of the relationship between principal behaviors and student, teacher, and school outcomes, with instructional management as omitted principal behavior

Principal behaviors	Student achievement	Teacher well-being	Teaching practices	School organizational health
Internal relations	0.059	0.029	0.120	0.004
	(0.068)	(0.253)	(0.238)	(0.361)
Organizational management	−0.062	−0.326	na	−0.280
	(0.057)	(0.215)		(0.353)
Administration	−0.045	na	na	na
	(0.068)			
External relations	−0.019	−0.563	na	−1.163
	(0.053)	(0.268)		(0.463)
Intercept (instructional management)	0.128*	0.411~	0.332**	0.852***
	(0.046)	(0.200)	(0.088)	(0.172)
n (*k*)	346 (31)	56 (10)	58 (10)	168 (24)

Note. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, *n* is the number of effect sizes and *k* is the number of studies. Cells with "na" are not estimated due to insufficient data.
~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

than other principal behaviors. However, for the three process outcome measures these estimates are quite imprecisely estimated, and we are unable to rule out large pooled effect size differences.

To summarize the results from our meta-regression, we find evidence that prior literature may overstate the unique importance of instructional management as a tool to improve student achievement outcomes. At the least, we conclude that in our sample of studies the effects of time spent on, and skill in, instructional management on student achievement outcomes are similar in magnitude to time and skill in other types of principal behaviors. These relationships should be interpreted with caution due to the observational nature of the underlying studies, but if these findings were to hold in a set of carefully constructed causal studies, it would imply no comparative efficiencies to be gained from focusing (either in time spent or skill development) on one type of school leadership activity compared with the other. This nuanced understanding of the value of instructional leadership accords with Grissom and Loeb (2011) and Grissom et al.'s (2013) findings. In addition to important questions regarding the relative value of different principal behaviors, it is also instructive to explore whether similar principal behaviors may relate to different outcomes differently.

We also find suggestive evidence that the effects of principals' behaviors are not similarly related to student and school outcomes. Specifically, in Figure 5 we observe that when we examine studies in which effect sizes for both student achievement and school organizational health can be connected to the same principal behavior construct, we find a negative relationship between achievement and organizational health outcomes ($r = -0.14$). Again, though we urge caution

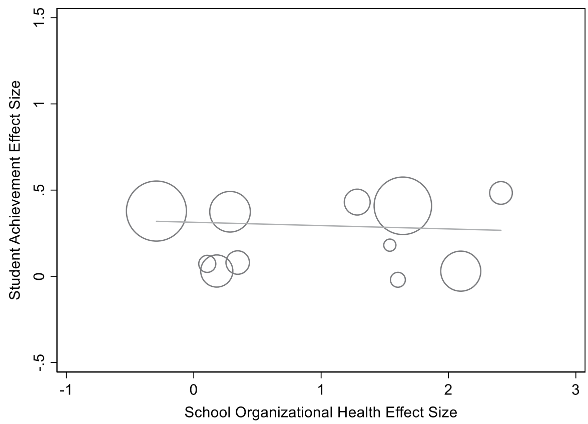

FIGURE 5. *Relationship between all principal behavior effects on student achievement and school organizational health outcomes.*
Note. Data points calculated averaging effect sizes for a particular outcome from the same study and weighting by precision of estimates. $k = 11$ studies.

in comparing across these 11 observational studies, we believe these results are suggestive that principal actions produce multidimensional student and school outcomes.

Sensitivity Analyses

We examine the sensitivity of our estimates to three sources of potential bias: outliers, missing study data, and study design. As Figure 3 reveals, each of our outcomes has relatively long tails. We test the sensitivity of our results to removing the lowest and highest 5% of effect sizes for each outcome. We report in Table 6 the results of our main analysis with trimmed outlying results. A comparison with the results in Table 2 reveals all of our substantive findings hold, though some of the student achievement and school organizational health results are attenuated. In fact, given the long tail of school organizational health effect size outcomes, the results from Table 6 are perhaps closer to realistic estimates of these effects.

In addition to extreme effect size results, we recognize the risk of potential bias in our results as a product of missing values in the distribution of effect sizes when studies that do not find statistically significant effects are not submitted or accepted for publication, as well as when authors do not include all outcome results in an article. Although it is impossible to fully test for the absence of unknown null results, we note two strategies to address these concerns. First, we include in our search criteria unpublished results and we follow-up with authors to request unreported relationships in studies. Second, we conduct a modified version of the Duval and Tweedie (2000) trim and fill method. This rank-based augmentation technique estimates the number of missing effect sizes using a funnel density plot and imputes these theoretically missing data points. It involves calculating the hypothetical data points needed to balance the spread of effect

TABLE 6

Sensitivity analysis of pooled effect size estimates of the effect of principal behaviors on student, teacher, and school outcomes, excluding top and bottom 5% of effect sizes for each outcome

Principal behaviors	Student achievement	Teacher well-being	Teaching practices	School organizational health
Instructional management	0.107** (0.031)	0.229~ (0.115)	0.386** (0.093)	0.714*** (0.131)
n (*k*)	198 (26)	28 (8)	45 (9)	116 (21)
Internal relations	0.115~ (0.047)	0.383~ (0.150)	na	0.591* (0.217)
n (*k*)	49 (12)	17 (6)		22 (7)
Organizational management	0.057 (0.039)	na	na	0.781 (0.433)
n (*k*)	30 (7)			10 (4)
Administration	0.061 (0.049)	na	na	na
n (*k*)	15 (5)			
External relations	0.081 (0.034)	na	na	na
n (*k*)	19 (5)			

Note. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, *n* is the number of effect sizes and *k* is the number of studies. Cells with "na" are not estimated due to insufficient data ($k < 4$).
~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

sizes across the mean effect size estimate derived from the random effects model in Equation 1. Given the nested nature of our effect sizes within studies, we collapse our data at the study level by averaging effect sizes and variances within a particular principal behavior–outcome pair. We conduct the trim-and-fill analysis and find that the Duval and Tweedie method identifies hypothetically missing studies only for our student achievement results. Therefore, we report in Table 7, the results of the trim-and-fill method only for our student achievement results. Despite collapsing the multiple effect sizes to single studies, we find essentially identical relationships between principal behavior and student achievement in column 1 of Table 7 as in our main results in Table 2. The imputation method in column 2 of Table 7 reveals slightly attenuated results, but they are substantively the same as our main results, with effect sizes on student achievement for both instructional management and internal relations of 0.10 *SD*.

Finally, we recognize that the design of the studies from which we sample in which some compare a bivariate relationship between a principal behavior and an outcome of interest and others in which a partial correlation is estimated—using various covariates to account for differences in contexts—creates potential issues of comparability across methodological designs. The Campbell Collective (Aloe, Tanner-Smith, Becker, & Wilson, 2016) offers various approaches to synthesize partial and bivariate relationships. We formally test for the differences in methodological design in Appendix Table S2 (in the online version of the journal). We

TABLE 7

Sensitivity analysis of effect size estimates of the effect of principal behaviors on student achievement outcomes using study-level trim-and-fill method

Principal behaviors	Unadjusted study-level estimates	Study-level estimates with imputed "missing" studies
Instructional management	0.135**	0.097*
	(0.041)	(0.044)
k	28	31
Internal relations	0.165**	0.097
	(0.049)	(0.059)
k	14	18
Organizational management	0.083	0.052
	(0.046)	(0.064)
k	8	9
Administration	0.042	0.042
	(0.044)	(0.044)
k	5	5[a]
External relations	0.068	0.063
	(0.036)	(0.039)
k	5	6

Note. Study-level effect size estimates with standard errors clustered at study level reported in parentheses. *k* is the number of studies.

[a]No "missing" studies found through trim-and-fill method.

~$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

again interpret this sample subset cautiously given our small starting sample of studies. Nevertheless, for instances in which we are able to compare bivariate and partial correlations in the meta-analytic framework, we find smaller signed but same directioned results for the partial correlations compared with the bivariate relationships. Given that only one of our studies uses a randomized design to compare "treatment" and "control" groups, we take this as suggestive evidence that those studies that include covariate adjustments are more successful at accounting for omitted variable bias. We are uncomfortable arguing that these are any more accurate results given their observational nature, but we hypothesize that future meta-analyses in which a pool of causal studies exist alongside the studies surveyed in this article might benefit from analyses that compare these future potentially causal bivariate relationships to the observational partial correlations in this review.

Discussion

We reach two substantive conclusions in our meta-analysis. First, in a review of 51 empirical studies relating principal behaviors to student, teacher, and school outcomes, we find consistently positive relationships between increased principal time or skill and student achievement, teacher well-being, instructional practices, and school organizational health. The magnitude of these relationships implies that a 1 *SD* difference in principal time or skill in instructional management, internal relations, organizational management, administration, or external relations is associated with between one tenth and one third of a standard deviation difference in student achievement, teacher well-being, and instructional practices. Based on Kraft's (2018) empirically derived schema for educational effect sizes, these represent moderate- to large-effect sizes. The magnitude of this relationship is much stronger for school organizational health, with average effect sizes on the order of two-thirds of a standard deviation. We recognize that there is a strong risk our findings are inflated due to omitted variable bias, which we discuss in further detail below. Nevertheless, we note that in Lipsey and Wilson's (1993) comparison of 74 meta-analyses that compared randomized and nonrandomized studies the difference in mean effect was negligible (0.05 *SD* units difference from groups with *SD* of 0.28 and 0.36). We find no substantive variation in our results based on the level of the school or in comparing studies conducted in the United States to elsewhere. Our findings are robust to various sensitivity checks for the presence of outliers, publication bias, and the inclusion of partial or bivariate relationships.

Second, we find that previous literature may overstate the unique student achievement effects of principals' time spent on and skill in instructional leadership behaviors. In fact, the effects of four other leadership behaviors are statistically indistinguishable from the effects of instructional management. We conclude from this that an exclusive focus on diverting time or skill development away from other noninstructional tasks toward instructional ones as some have suggested (e.g., Bambrick-Santoyo & Peiser, 2012) may be misguided. In this way, our conclusions are similar to Sebastian, Allensworth, Wiedermann, Hochbein, and Cunningham (2018) who find that principals conceive of their leadership skills unidimensionally across instructional and organizational management, and that these jointly predict stronger student outcomes. Note that our findings do not imply that instructional leadership is not important nor that it does not merit more attention. In fact, as Grissom et al. (2013) document, in Miami-Dade, principals spent only 12.7% of their time on average on instructional management–related tasks. Thus, a more equal balance of time across the task categories may be of value. Alternatively, instructional management may in fact have a unique role in improving outcomes, but it must be paired with other strategies to leverage its unique status. Our study design does not allow us to test this hypothesis; however, such a finding would still imply that other noninstructional tasks are critical.

Taken at face value, our findings suggest that principals must effectively engage in these five leadership behaviors with little opportunity for relative efficiencies gained by focusing only on some. This is likely cold comfort to U.S. principals who report in the 2015–2016 school year average work weeks of 58.6 hours (U.S. Department of Education, National Center for Education Statistics, 2017). If the findings of this meta-analysis hold in well-designed causal studies, the returns to quality school leadership are substantial, but constrained by the amount of time the current principal force can expend on any one of these behaviors. This constraint might be addressed by dividing responsibility for these behaviors over more educators.

If each school in the United States hired an additional principal or assistant principal, even if the newly hired principals allocated time identically, this could permit increases in the amount of time spent on each behavior. Principals reported spending 17.5 hours in an average week on Curriculum and Teaching–related tasks, with a standard deviation of 12.3 hours (U.S. Department of Education, National Center for Education Statistics, 2017). Adding a school administrator could therefore permit, for example, increases in time spent on instructional management by 18 hours, or 1.5 standard deviations more than is currently spent. Based on the strong assumption of nondiminishing returns, the implied effect on student achievement would be on the order of 0.17 *SD*. For comparison's sake, this is about one-quarter the size of the Black–White achievement gap (Bloom, Hill, Black, & Lipsey, 2008; S. Reardon, 2011). The average principal earned $95,700 in 2015–2016 (U.S. Department of Education, National Center for Education Statistics, 2017). A back-of-the-envelope calculation, using these nationally representative U.S. figures implies that although the cost of hiring 91,000 principals would be around $9 billion, the benefits to the 50 million United States school children could be on the order of 0.17 *SD* for $180 per student. Alternatively, an investment in improving the skill levels of school principals in any one of these areas could generate substantial returns at an even lower cost. Similar benefits might be realized in the quality of teachers' instruction.

We intend these rough figures to be helpfully illustrative of an upper bound estimate of the benefits of increased school leadership capacity and its associated costs. We recognize that the magnitude of these effects depends on equivalent skill levels among newly hired principals, nondiminishing returns to their behaviors, unbiased estimates in the observational studies from our meta-analysis, and more. However, even if estimates for the potential benefits of increasing the intensity or quality of school leadership behaviors are substantially overstated, benefits as small as half of these would compare favorably to other commonly advocated educational interventions such as substantial class size reductions (~0.20 *SD*), intensive tutoring (~0.25 *SD*), or intensive teacher evaluation (~0.10 *SD*).

Of course, all of these conclusions should be taken in the context of a limited knowledge base. As we note above, there is substantial heterogeneity in the direction and magnitude of the relationships between principal behaviors and student, school, and organizational outcomes in the studies underlying our meta-analysis. Even in our more precisely estimated pooled effects, for example, the relationship between instructional management and student achievement, we are only able to confidently predict that a future study drawing from the same underlying population of schools and principals would estimate a relationship of between negative 0.12 and positive 0.35 *SD*. The between-study variability, and therefore our prediction intervals, are even wider for our teacher well-being, instructional practices, and school organizational health outcomes. One important conclusion from the heterogeneity of our prediction intervals is that principals' actions matter in different ways in different contexts. Further study to better understand the extent to which different accountability, school culture, or demographic contexts influence principal behavior effects would help inform future policy. Despite our tests for the presence of studies or relationships that are not reported due to publication bias, it is impossible to rule out the possibility that research teams have not publicly released various studies in which the relationship between principal behaviors and student, teacher, and school outcomes is either smaller than, or in the opposite direction of, the effects found in our sample. Most important, as we note repeatedly, our meta-analysis relies on only one study[12] permitting causal inferences, compared with 50 that are observational in nature. In addition to the now familiar calls of all meta-analysts for future authors of primary studies to report all critical study information, including sample size for every estimate, standard errors (or deviations) and measure construction details, we propose several lines of future causal research.

Despite the challenges in randomly assigning principals to improve their skill or time dedicated to particular behaviors, we suggest several methods by which future research might estimate the causal effects of principals' behaviors. For example, studies might randomly assign principals to a professional development activity that emphasizes either building instructional or operational management skills. Trained observers could observe principals prior to and after the training and record either skill or time spent on the five leadership behaviors. Instrumental variable estimates might then capture the portion of the change in leadership skill or behavior attributable to the exogenous assignment to different professional development activities. These behavioral changes would presumably be the only difference, on average, between the two groups of principals and any differences in student, teacher, and school outcomes across the two groups could be credibly argued to be caused by changes in principal behavior. Other studies might capitalize on mandatory school-size-to-administrator ratios in a regression discontinuity framework to examine the effect of an additional assistant principal on various outcomes. Combined with time-use data, estimates of the causal effect of more total school administrator time spent on different tasks might be compared between schools that just fall short of receiving an additional assistant principal and those that are assigned one. Of particular interest, motivated by Grissom et al.'s (2013) nuanced findings that the type of behavior within each of these categories matters, is further causal research focused on the effects of different types of actions within the five behavioral constructs on outcomes. Any such studies would benefit from an explicit preregistration of the differential effects of principal behaviors across various school contexts and replication studies examining the generalizability of the findings.

While there are clear benefits to the types of causal research designs detailed above, limits to what questions may be addressed or to which populations such results may be generalized highlight the value of complementary research that estimates the effects of principal behaviors in observational studies. For instance, the above theoretical study exploiting assistant principal assignment thresholds would only be generalizable to schools with student enrollment figures close to the thresholds. Thus, pairing such a study with matching or correlational estimates, adjusted for various dimensions of school context, in a broadly representative sample would permit tentative conclusions further away from the assignment threshold. More broadly, learning more about how principals spend their time, in what areas are their skills most developed, and how these relate to relevant educational outcomes will prove fruitful.

In crafting a model for school leadership, a behavior-based approach is one piece of a larger puzzle. The study of principal behaviors, while critical, does not account for other potential dimensions of school leadership, importantly, principals' knowledge base, both of academic content (Stein & Nelson, 2003) and how teachers learn (Stein & Spillane, 2005). It is also possible that the sum of any particular behaviors and skills does not capture the value of the integration of behaviors into principal style, attributes, and experiences. While we feel our conclusions provide valuable insights for the principalship, school leadership models should also focus on how school leaders conceptualize their role, knowledge, and orientation around school improvement and instruction (Hallinger, Leithwood, & Murphy, 1993; Stein & Spillane, 2005), as well as acknowledge the important role that teachers as learners play in their interpretation and receipt of principal behaviors (Stein & Spillane, 2005).

Conclusion

We pool results from 51 quantitative studies on the relationship between five principal behaviors and student, teacher, and school outcomes. We find moderate- to large-positive effects across all leadership behaviors on student achievement, teacher well-being, instructional practices, and school organizational health. We find suggestive evidence that instructional management strategies are no more strongly related to these outcomes than other critical principal behaviors. We argue that our findings imply the value of investing in the capacity of school leaders, either through more leadership staff or through building the capacity of current leaders. Finally, we signal an important caution related to these findings, namely that they are based on research that cannot support causal inferences. We articulate a research agenda for future scholars interested in better understanding the causal relationship between school leaders' actions and skills and student, teacher, and school outcomes.

Notes

We thank Matthew Kraft, Susanna Loeb, Nicholas Parr, Brendan Bartanen, and three anonymous referees for helpful feedback on this article. We thank Jason Grissom for providing coefficients of interest. We received valuable comments from participants in the EMPL Seminar Series at the University of Oregon and the AEFP 2019 Conference. We are particularly indebted to Emily Tanner-Smith for guidance in the planning of this study and helpful feedback on initial drafts. We are also grateful to the many authors of studies included in the meta-analysis who responded to our queries. All errors are our own.

[1]The alpha coefficients from the principal behavior constructs in Grissom and Loeb (2011) are as follows: instructional management, 0.90; internal relations, 0.82; organizational management, 0.83; administration, 0.82; and external relations, 0.73.

[2]More broadly, the large body of research generated by Leithwood, Hallinger, Heck, and their coauthors has generated valuable insights to the theoretical and empirical study of school leadership. For the most part, however, while we discuss the insights from their work in our motivating sections, we are unable to include their studies in our meta-analysis due to our targeted interest in understanding ways in which particular principal behaviors relate to student, teacher, and school outcomes. We are unable to extract evidence on the principal behaviors of interest from their analyses of overarching leadership styles or practices.

[3]We included both published and unpublished studies. For published studies, we included those published in 2001 and beyond. For unpublished studies, primarily doctoral dissertations, we took as their completion date the most recent date listed on the document, typically the student's date of graduation.

[4]This intraclass correlation statistic excludes the private/public school code. In our final coding of the studies, all 51 studies include public schools in their sample and only four include private schools. As there is no "true" variability in public schooling, a one-item difference between coders reduced the overall ICC to 0.757. Because so few private school studies are in our sample, we do not include this variable in our analysis.

[5]Some studies did not report standard errors or *t* statistics but did include stars denoting levels of significance. We queried the authors for this information, and if we received an insufficient response, we calculated the lower bound *t* statistic (e.g., $p < .05 = t$ statistic of 1.96) and used these to estimate the standard error.

[6]Specifically, we convert all outcomes into standardized effect sizes (Cohen's *d*) and associated standard errors (SE_d) using the following formulas: (1) mean difference: $d = \left(\bar{X}_1 - \bar{X}_2\right) / SD_{\text{within}}$, where $SD_{\text{within}} = \sqrt{\left((n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2\right) / (n_1 + n_2 - 2)}$ and $SE_d = \sqrt{\left((n_1 + n_2)/n_1 n_2\right) + \left(d^2 / 2(n_1 + n_2)\right)}$; (2) correlation coefficient (*r*): $d = 2r / \sqrt{1 - r^2}$ and $SE_d = \sqrt{\left(4\left(1 - r^2\right)^2 / (n - 1)\right) / \left(1 - r^2\right)^3}$; (3) regression coefficient: $r_p = t_f / \sqrt{t_f^2 + df}$, where t_f is the *t* statistic on the regression coefficient, *df* are the degrees of freedom, and r_p is the partial correlation coefficient which we convert to Cohen's *d* following step (2) above, $SE_d = \sqrt{\left(4\left(1 - r_p^2\right)^2 / df\right) / \left(1 - r^2\right)^3}$; and (4) standardized slope: *d* is the coefficient and $SE_d = \sqrt{\left(\left(1 - R_y^2\right) / df\right)\left(1 / \left(1 - R_x^2\right)\right)}$.

[7]The robust variance estimation procedure (Tanner-Smith & Tipton, 2014) constructs weights as follows: $w_{ij} = 1 / \left\{\left(V_j + \tau^2\right)\left(1 + (k_j - 1)\rho\right)\right\}$, where V_j is the mean of the within-study sampling variances for study *j*, τ^2 is the estimate of between-study variance, k_j is the number of effect sizes (cases) for study *j*, and ρ (rho) measures the expected correlation of effect sizes within study. For all estimates presented in the article, we use the value $\rho = 0.9$. We test for values of rho ranging between 0.1 and 0.99 and results are trivially different with coefficients and standard errors varying in the hundredths or thousandths places.

[8]A common example of the first instance was when a study reported principal's time spent on a particular behavior descriptively, but then only modeled student achievement outcomes as a function of an overall leadership ability construct. A common example of the second issue was when a study marked a principal behavior–outcome relationship as significant with an asterisk, but failed to include a mean difference, coefficient, *t* statistic, or any other numerical summary of the relationship.

[9]We identified 19 studies for author follow-up at this stage, either as a result of incomplete information that prevented us from including the study in the meta-analysis entirely, or that as a result of missing information required us to estimate standard errors on the standardized main effect. We sent e-mails to six authors requesting additional information to permit the study to be included. We sent e-mails to 13 authors requesting additional information to permit the inclusion of precise standard errors rather than an estimate and/or information to include nonsignificant coefficients not reported in the original paper. Thirteen authors responded; however, only one provided information on unreported standard errors (Tuytens & Devos, 2011) and two (Valentine & Prater, 2011; Sebastian, Huang, & Allensworth, 2017) sent us articles that used the same data set, but had reported the findings differently, allowing us to include the results (Prater, 2004; Sebastian, Allensworth, & Huang, 2016). Details of the nonincluded studies are available from the authors on request.

[10] We note heterogeneity across the effect sizes we sample. One meaningful way to capture heterogeneity in effect sizes across studies is the 95% prediction interval, calculated as follows: $95\%PI = \mu \pm t_{95\%}\sqrt{\tau^2 + V_\mu}$, where μ is the pooled mean effect, $t_{95\%}$ is the critical t statistic given the degrees of freedom, τ^2 is the between-effect-size variability, and $V\mu$ is the variance of the pooled mean (Borenstein, Higgins, Hedges, & Rothstein, 2017). For example, our prediction intervals on the relationship between instructional management and student achievement outcomes imply that in 95 of 100 future studies drawing from a similar underlying population of principals, we would expect that the magnitude of the effect of instructional management to range between −0.12 and 0.35 *SD*. We return to this topic in our Discussion.

[11]Our level-of-schooling heterogeneity analysis is ill-suited for the inclusion of a moderator interacted with a predictor (e.g., Instruction_Mgmt*Secondary) in a meta-regression framework because some effect sizes include both elementary and secondary schools in their sample. We do test this approach by interacting our predictors with an indicator for having a U.S.-based sample and reestimate all results from Table 4 on the full sample of studies. We find no significant effects for any of the predictors or outcomes, implying no differences in principal behavioral effects across U.S. and international contexts.

[12]This study (Silva, White, & Yoshida, 2011) samples 41 students from a single school, reaches a significant result only after excluding an outlier, and tests a principal intervention that calls for providing direct services to students. Thus, few robust conclusions can be drawn from it.

References

References marked with an asterisk were included in the meta-analysis.

*Alig-Mielcarek, J. M. (2003). *A model of school success: Instructional leadership, academic press, and student achievement* (Doctoral dissertation). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1054144000

*Allen, N., Grigsby, B., & Peters, M. L. (2015). Does leadership matter? Examining the relationship among transformational leadership, school climate, and student achievement. *International Journal of Educational Leadership Preparation*, *10*(2), 1–22.

Aloe, A. M., Tanner-Smith, E., Becker, B., & Wilson, D. (2016). *Synthesizing bivariate and partial effect sizes: Campbell policy note 2*. Oslo, Norway: Campbell Collaboration. doi:10.4073/cmpn.2016.2

Aloe, A. M., & Thompson, C. G. (2013). The synthesis of partial effect sizes. *Journal of the Society for Social Work and Research*, *4*, 390–405. doi:10.5243/jsswr.2013.24

Austin, W., Chen, B., Goldhaber, D., Hanushek, E., Holden, K., Koedel, C., . . . Turaeva, M. (2019). *Path to the principalship and value-added: A cross-state comparison of elementary and middle-school principals* (CALDER Working Paper Series No. 213-0119-1). Washington, DC: American Institutes for Research. Retrieved from https://caldercenter.org/sites/default/files/CALDER WP 213-0119-1.pdf

Avolio, B., & Bass, B. (1997). *Multifactor Leadership Questionnaire*. Menlo Park, CA: Mind Garden.

Bambrick-Santoyo, P., & Peiser, B. M. (2012). *Leverage leadership: A practical guide to building exceptional schools*. San Francisco, CA: Jossey-Bass.

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

*Bednar, J. (2018). *Effective school leaders and student achievement: An examination of 265 schools in Northern Illinois* (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 10747419)

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*, 904–919. doi:10.1016/j.ssresearch.2012.03.003

Blase, J., & Blase, J. (2004). *Handbook of instructional leadership: How successful principals promote teaching and learning* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, *1*, 289–328. doi:10.1080/19345740802400072

Borenstein, M. (2009). *Introduction to meta-analysis* (1st ed.). West Sussex, England: Wiley.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, *8*(1), 5–18. doi:10.1002/jrsm.1230

Branch, G., Hanushek, E., & Rivkin, S. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (NBER Working Paper No. 17803). Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w17803

*Braun, D., Gable, R., & Kite, S. (2008). Relationship among essential leadership preparation practices and leader, school, and student outcomes in K-8 schools. *K-12 Education*, 8. Retrieved from https://scholarsarchive.jwu.edu/k12_ed/8

Burkhauser, S. (2017). How much do school principals matter when it comes to teacher working conditions? *Educational Evaluation and Policy Analysis*, *39*, 126–145. doi:10.3102/0162373716668028

*Burkhauser, S., Gates, S. M., Hamilton, L. S., & Ikemoto, G. S. (2012). *First-year principals in urban school districts*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/technical_reports/TR1191.html

*Cancio, E. J., Albrecht, S. F., & Johns, B. H. (2013). Defining administrative support and its relationship to the attrition of teachers of students with emotional and behavioral disorders. *Education and Treatment of Children*, *36*(4), 71–94. doi:10.1353/etc.2013.0035

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy*, *11*, 283–309. doi:10.1162/EDFP_a_00184

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance* (CALDER Working Paper No. 38, National Center for Analysis of Longitudinal Data in Education Research). Washington, DC: Urban Institute. Retrieved from https://eric.ed.gov/?id=ED509693

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, *31*, 92–109. doi:10.1016/j.econedurev.2011.09.001

Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012). Training your own: The impact of New York City's Aspiring Principals Program on student achievement. *Educational Evaluation and Policy Analysis*, *34*, 232–253. doi:10.3102/0162373712437206

*Crosthwaite, G. (2015). *The correlation between elementary school leadership behaviors and student achievement* (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3689171)

*Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., . . . Kington, A. (2009). *The impact of school leadership on pupil outcomes*. Nottingham, England: University of Nottingham.

*Devries, R. (2017). *The instructional leadership practices of elementary principals of average needs/resource capacity school districts in New York State*. Available from ProQuest Dissertations and Theses database. (UMI No. 10640434)

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, *47*, 634–663. doi:10.1111/caje.12086

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

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. doi:10.1111/j.0006-341X.2000.00455.x

*Ebmeier, H. (2003). How supervision influences teacher efficacy and commitment: An investigation of a path model. *Journal of Curriculum and Supervision*, *18*, 110–141. Retrieved from https://eric.ed.gov/?id=EJ660834

*Egley, R., & Jones, B. (2005). Principals' inviting leadership behaviors in a time of test-based accountability. *Scholar-Practitioner Quarterly*, *3*(1), 13–24. Retrieved from https://files.eric.ed.gov/fulltext/EJ796133.pdf

*Fancera, S., & Bliss, J. (2011). Instructional leadership influence on collective teacher efficacy to improve school achievement. *Leadership and Policy in Schools*, *10*, 349–370. doi:10.1080/15700763.2011.585537

Fullan, M. (2001). *The new meaning of educational change*. New York, NY: Routledge.

*Gates, S. M., Baird, M. D., Doss, C. J., Hamilton, L. S., Opper, I. M., Master, B. K., . . . Zaber, M. A. (2019). *Preparing school leaders for success: Evaluation of new leaders' aspiring principals program, 2012–2017*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/research_reports/RR2812.html?adbsc=social_20190227_2707441&adbid=1100578268338577408&adbpl=tw&adbpr=22545453

Gates, S. M., Baird, M., Master, B., & Chavez-Herrerias, E. (2019). *Principal pipelines: A feasible, affordable, and effective way for districts to improve schools*. Santa Monica, CA: RAND. doi:10.7249/RR2666

*Gates, S. M., Hamilton, L. S., Burkhauser, S., Heaton, P., Pierson, A., & Baird, M. D., . . . Rand Education. (2014). *Preparing principals to raise student achievement: Implementation and effects of the New Leaders program in ten districts*. Santa Monica, CA: RAND. Retrieved from https://eric.ed.gov/?id=ED561152

*Geijsel, F. P., Sleegers, P. J. C., Stoel, R. D., & Krüger, M. L. (2009). The effect of teacher psychological and school organizational and leadership factors on teachers' professional learning in Dutch schools. *Elementary School Journal*, *109*, 406–427. doi:10.1086/593940

Goldhaber, D., Holden, K., & Chen, B. (2019). *Do more effective teachers become more effective principals?* (CALDER Working Paper Series No. 215-0119-1). Washington, DC: RAND. Retrieved from https://caldercenter.org/sites/default/files/CALDER WP 215-0119-1.pdf

*Graczewski, C., Knudson, J., & Holtzman, D. J. (2009). Instructional leadership in practice: What does it look like, and what influence does it have? *Journal for Students Placed at Risk*, *14*(1), 72–96. doi:10.1080/10824660802715460

*Greb, W. (2011). *Principal leadership and student achievement: What is the effect of transformational leadership in conjunction with instructional leadership on student achievement?* (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3468985)

Grissom, J. A., Bartanen, B., & Mitani, H. (2018). *Principal sorting and the distribution of principal quality* (Vanderbilt University Working Paper). Nashville, TN: Vanderbilt University.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3–28. doi:10.3102/0162373714523831

*Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness. *American Educational Research Journal*, *48*, 1091–1123. doi:10.3102/0002831211402663

Grissom, J. A., & Loeb, S. (2017). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy*, *12*, 369–395. doi:10.1162/EDFP_a_00210

*Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders. *Educational Researcher*, *42*, 433–444. doi:10.3102/0013189X13510020

Grissom, J. A., Mitani, H., & Woo, D. S. (2019). Principal preparation programs and principal outcomes. *Educational Administration Quarterly*, *55*, 73–115. doi:10.1177/0013161X18785865

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, *32*, 5–44. doi:10.1177/0013161X96032001002

Hallinger, P., & Heck, R. H. (2010). Leadership for learning: Does collaborative leadership make a difference in school improvement? *Educational Management Administration & Leadership*, *38*, 654–678. doi:10.1177/1741143210379060

Hallinger, P., Leithwood, K., & Murphy, J. (1993). Cognitive perspectives on educational administration: An introduction. *Educational Administration Quarterly*, *29*, 296–301.

Hargreaves, A., & Goodson, I. (2006). Educational change over time? The sustainability and nonsustainability of three decades of secondary school change and continuity. *Educational Administration Quarterly*, *42*, 3–41. doi:10.1177/0013161X05277975

Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, *46*, 659–689. doi:10.3102/0002831209340042

Heck, R. H., & Hallinger, P. (2010). Collaborative leadership effects on school improvement: Integrating unidirectional- and reciprocal-effects models. *Elementary School Journal*, *111*, 226–252. doi:10.1086/656299

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. doi:10.1002/jrsm.5

Higgins, J. P., Ramsay, C., Reeves, B. C., Deeks, J. J., Shea, B., Valentine, J. C., . . . Wells, G. (2013). Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods*, *4*(1), 12–25. doi:10.1002/jrsm.1056

Hitt, D. H., & Tucker, P. D. (2016). Systematic review of key leader practices found to influence student achievement. *Review of Educational Research*, *86*, 531–569. doi:10.3102/0034654315614911

*Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*, 491–523. doi:10.1086/653625

*Hulpia, H., Devos, G., & Van Keer, H. (2011). The relation between school leadership from a distributed perspective and teachers' organizational commitment. *Educational Administration Quarterly*, *47*, 728–771. doi:10.1177/0013161X11402065

Husain, A., Matsa, D. A., & Miller, A. R. (2018). *Do male workers prefer male leaders? An analysis of principals' effects on teacher retention* (NBER Working Paper No. w25263). doi:10.2139/ssrn.3259638

*Jones, C., & Gilman, L. (2019, March). *How are district differences in the implementation of teacher evaluations associated with teacher perceptions of school leadership and teacher job satisfaction?* Paper presented at American Education and Finance Policy Conference, Milwaukee, WI.

*Klinginsmith, E. N. (2007). *The relative impact of principal managerial, instructional, and transformational leadership on student achievement in Missouri middle level schools* (Doctoral dissertation, University of Missouri–Columbia). doi:10.32469/10355/4772

Kraft, M. A. (2018). *Interpreting effect sizes of education interventions* (Brown University Working Papers). Retrieved from https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*, 547–588. doi:10.3102/0034654318759268

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, *52*, 711–753. doi:10.1177/0013161X16653445

*Krüger, M. L., Witziers, B., & Sleegers, P. (2007). The impact of school leadership on school level factors: Validation of a causal model. *School Effectiveness and School Improvement*, *18*(1), 1–20. doi:10.1080/09243450600797638

*Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, *17*, 353–366. doi:10.1287/orsc.1060.0191

Leithwood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, *38*, 112–129. doi:10.1108/09578230010320064

Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, *4*, 177–199. doi:10.1080/15700760500244769

Leithwood, K., & Sun, J. (2012). The nature and effects of transformational school leadership: A meta-analytic review of unpublished research. *Educational Administration Quarterly*, *48*, 387–423. doi:10.1177/0013161X11436268

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209. doi:10.1037/0003-066X.48.12.1181

*Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national U.S. survey. *School Effectiveness and School Improvement*, *21*, 315–336. doi:10.1080/09243453.2010.486586

Marzano, R., Waters, T., & McNulty, B. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

*May, H., Huff, J., & Goldring, E. (2012). A longitudinal study of principals' activities and student performance. *School Effectiveness and School Improvement*, *23*, 417–439. doi:10.1080/09243453.2012.678866

*May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*, 332–352. doi:10.1177/0013161X10383411

*McInnis, T. M. (2009). *Relationship between high school leadership team practices and student achievement* (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3383219)

*Mees, G. W. (2008). *The relationships among principal leadership, school culture, and student achievement in Missouri middle schools* (Doctoral dissertation, University of Missouri–Columbia). doi:10.32469/10355/5540

*Minckler, C. H. (2014). School leadership that builds teacher social capital. *Educational Management Administration & Leadership*, *42*, 657–679. doi:10.1177/1741143213510502

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., & The Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). doi:10.1371/journal.pmed.1000097

*Mulford, B., & Silins, H. (2011). Revised models and conceptualisation of successful school principalship for improved student outcomes. *International Journal of Educational Management*, *25*(1), 61–82. doi:10.1108/09513541111100125

*Ndoye, A., Imig, S. R., & Parker, M. A. (2010). Empowerment, leadership, and teachers' intentions to stay in or leave the profession or their schools in North Carolina charter schools. *Journal of School Choice*, *4*, 174–190. doi:10.1080/15582159.2010.483920

*Nelson, J. L., Hegtvedt, K. A., Haardörfer, R., & Hayward, J. L. (2019). Trust and respect at work: Justice antecedents and the role of coworker dynamics. *Work and Occupations*, *46*, 307–338. doi:10.1177/0730888419835261

*Odegaard, L. (2008). *The relationship between teacher-identified principal leadership behavior and effectiveness and student achievement in South Dakota secondary schools* (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3318826)

*O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' instructional leadership behaviors and student achievement. *NASSP Bulletin*, *89*, 56–71. doi:10.1177/019263650508964505

OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris, France: Author. doi:10.1787/9789264190658-en

*Orr, M. T., & Orphanos, S. (2011). How graduate-level preparation influences the effectiveness of school leaders: A comparison of the outcomes of exemplary and conventional leadership preparation programs for principals. *Educational Administration Quarterly*, *47*, 18–70. doi:10.1177/0011000010378610

Osborne-Lampkin, L., Folsom Sidler, J., & Herrington, C. D. (2015). *A systematic review of the relationships between principal characteristics and student achievement*

(REL 2016-091). Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/edlabs

*Prater, M. (2004). *The relative impact of principal managerial, instructional and transformational leadership on student achievement* (Unpublished doctoral dissertation). University of Missouri, Columbia.

*Quinn, D. M. (2002). The impact of principal leadership behaviors on instructional practice and student engagement. *Journal of Educational Administration*, *40*, 447–467. doi:10.1108/09578230210440294

*Quint, J. C., Akey, T. M., Rappaport, S., & Willner, C. J. (2007). *Instructional leadership, teaching quality and student achievement: Suggestive evidence from three urban school districts*. New York, NY: Manpower Demonstration Research. Retrieved from https://eric.ed.gov/?id=ED499788

*Reardon, R. M. (2011). Elementary school principals' learning-centered leadership and educational outcomes: Implications for principals' professional development. *Leadership and Policy in Schools*, *10*(1), 63–83. doi:10.1080/15700760903511798

Reardon, S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–116). New York, NY: Russell Sage Foundation.

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*, 635–674. doi:10.1177/0013161X08321509

*Sebastian, J., Allensworth, E., & Huang, H. (2016). The role of teacher leadership in how principals influence classroom instruction and student learning. *American Journal of Education*, *123*, 69–108. doi:10.1086/688169

Sebastian, J., Allensworth, E., Wiedermann, W., Hochbein, C., & Cunningham, M. (2018). Principal leadership and school performance: An examination of instructional leadership and organizational management. *Leadership and Policy in Schools*. Advance online publication. doi:10.1080/15700763.2018.1513151

Sebastian, J., Huang, H., & 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*, 463–488. doi:10.1080/09243453.2017.1319392

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

*Shatzer, R. H., Caldarella, P., Hallam, P. R., & Brown, B. L. (2014). Comparing the effects of instructional and transformational leadership on student achievement. *Educational Management Administration & Leadership*, *42*, 445–459. doi:10.1177/1741143213502192

*Shen, J., Leslie, J. M., Spybrook, J. K., & Ma, X. (2012). Are principal background and school processes related to teacher job satisfaction? A multilevel study using schools and staffing survey 2003–04. *American Educational Research Journal*, *49*, 200–230. doi:10.3102/0002831211419949

*Silva, J. P., White, G. P., & Yoshida, R. K. (2011). The direct effects of principal–student discussions on eighth grade students' gains in reading achievement: An experimental study. *Educational Administration Quarterly*, *47*, 772–793. doi:10.1177/0013161X11404219

Spillane, J. P., & Hunt, B. R. (2010). Days of their lives: A mixed-methods, descriptive analysis of the men and women at work in the principal's office. *Journal of Curriculum Studies*, *42*, 293–331. doi:10.1080/00220270903527623

Spillane, J. P., & Lee, L. C. (2014). Novice school principals' sense of ultimate responsibility. *Educational Administration Quarterly*, *50*, 431–465. doi:10.1177/0013161X13505290

Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis*, *25*, 423–448. doi:10.3102/01623737025004423

Stein, M. K., & Spillane, J. (2005). What can researchers on educational leadership learn from research on teaching? Building a bridge. In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research in educational leadership* (pp. 28–45). New York, NY: Teachers College Press.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., & Rennie, D., . . . MOOSE Group. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA Journal of the American Medical Association*, *283*, 2008–2012. doi:10.1001/jama.283.15.2008

Sun, J., & Leithwood, K. (2012). Transformational school leadership effects on student achievement. *Leadership and Policy in Schools*, *11*, 418–451. doi:10.1080/15700763.2012.681001

Sun, J., & Leithwood, K. (2015). Direction-setting school leadership practices: A meta-analytic review of evidence about their influence. *School Effectiveness and School Improvement*, *26*, 499–523. doi:10.1080/09243453.2015.1005106

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, *46*, 31–56. doi:10.1177/1094670509353043

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13–30. doi:10.1002/jrsm.1091

*Thoonen, E. E. J., Sleegers, P. J. C., Oort, F. J., Peetsma, T. T. D., & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly*, *47*, 496–536. doi:10.1177/0013161X11400185

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*, 375–393. doi:10.1037/met0000011

*Tuytens, M., & Devos, G. (2011). Stimulating professional learning through teacher evaluation: An impossible task for the school leader? *Teaching and Teacher Education*, *27*, 891–899. doi:10.1016/j.tate.2011.02.004

U.S. Department of Education, National Center for Education Statistics. (2017). *National Teacher and Principal Survey (NTPS), Public School Principal Data File, 2015-16 and Principal Follow-Up Survey*. Washington, DC: Author.

Valentine, J., & Prater, M. (2011). Instructional, transformational, and managerial leadership and student achievement: High school principals make a difference. *NASSP Bulletin*, *95*(1), 5–30. doi:10.1177/0192636511404062

Valentine, J. C., & Thompson, S. G. (2013). Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods*, *4*(1), 26–35. doi:10.1002/jrsm.1064

*Ware, H. W., & Kitsantas, A. (2011). Predicting teacher commitment using principal and teacher efficacy variables: An HLM approach. *Journal of Educational Research*, *104*, 183–193. doi:10.1080/00220671003638543

*Williams, E. (2009). Evaluation of a school systems plan to utilize teachers' perceptions of principal leadership to improve student achievement. *Challenge*, *15*(1), 13–30.

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*, 398–425. doi:10.1177/0013161X03253411

Zepeda, S. J. (2013). *The principal as instructional leader: A practical handbook* (3rd ed.). New York, NY: Routledge.

Authors

DAVID D. LIEBOWITZ is an assistant professor in the Department of Educational Methodology, Policy and Leadership at the University of Oregon, 102S Lokey Education Building, Eugene, OR 97403, USA; email: *daviddl@uoregon.edu*. He studies the challenges facing schools with concentrations of traditionally underserved students and effective leadership strategies to address these challenges.

LORNA PORTER is a PhD student in the Department of Educational Methodology, Policy and Leadership at the University of Oregon, 102 Lokey Education Building, Eugene, OR 97403, USA; email: *Lporter@uoregon.edu*. Her research focuses on the impacts of educational policy decisions on the opportunities and outcomes for students classified as English learners and immigrant students.