THE EFFECT OF PRINCIPAL BEHAVIORS ON STUDENT, TEACHER AND SCHOOL OUTCOMES: A SYSTEMATIC REVIEW AND META-ANALYSIS OF THE EMPIRICAL LITERATURE

David D. Liebowitz[a]
Lorna Porter

University of Oregon, College of Education

March 2019

We thank Matthew Kraft for helpful feedback on this paper. We are also grateful to participants in the EMPL Seminar Series at the University of Oregon. We are particularly indebted to Emily Tanner-Smith for guidance in the planning of this study and helpful feedback on earlier 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.

[a] Corresponding author: daviddl@uoregon.edu, 102S Lokey, 5267 University of Oregon, Department of Educational Methodology, Policy and Leadership, Eugene, OR 97403

THE EFFECT OF PRINCIPAL BEHAVIORS ON STUDENT, TEACHER AND SCHOOL OUTCOMES: A SYSTEMATIC REVIEW AND META-ANALYSIS OF THE EMPIRICAL LITERATURE

ABSTRACT

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 42 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 (0.09-0.17 standard deviations), teacher well-being (0.34 *SD*), teacher instructional practices (0.34 *SD*), and school organizational health (0.69 *SD*); (2) we find that prior literature may overstate the unique importance of instructional management as a tool 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 non-existent. 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 school leaders to improve school climate, instructional practice and student outcomes. School leaders are responsible for a dizzying set of responsibilities (cf. 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. Principals come to the profession from a wide variety of backgrounds (Grissom & Loeb, 2011) and with widely varying skill sets (Grissom, Bartanen, & Mitani, 2018; Grissom & Loeb, 2011, 2017). Despite methodologic difficulties in measuring school leaders' impact on student learning (Chiang, Lipscomb, & Gill, 2016; Grissom, Kalogrides, & Loeb, 2015), strong causal evidence suggests that school leaders 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).

Several well-identified studies suggest not only that principals have meaningful effects on student learning outcomes, but that there is wide variation in principal quality. The strongest identification strategies on principal quality rely on instances in which principals transition from one school to another. Branch, Hanushek and Rivkin (2012) estimate as a lower bound that a one standard deviation unit more effective principal predicts a 0.05 standard deviation units (*SD*) increase in students' rate of achievement growth *per year* in Texas; their upper-bound estimates suggest effects on the order of 0.20 *SD*. The Canadian province of British Columbia (BC) 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 one standard deviation 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 one standard deviation more effective

principal will generate an approximately one-third standard deviation higher graduation rate and a one standard deviation higher English score in BC 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). 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 school leaders, however, the mechanisms by which their behaviors affect outcomes remain unclear. While the field of education leadership has been widely studied, much of the early empirical research literature sought to design and test the impacts of broad leadership styles, e.g., *transformative* (e.g., Leithwood & Jantzi, 2000, 2005), *distributed* (e.g., Heck & Hallinger, 2009) or *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 school leaders affect school and student outcomes (e.g., Hallinger & Heck, 1996). These studies generally found transformative or collaborative leadership had positive direct effects on 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 non-traditional pathways to the principalship, outside of university preparation programs. A multi-year 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 to times when they were not principal of these same schools (Gates et al., 2019, 2014). While statistically significant, the magnitude of these effects are relatively small and vary substantially across 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 vary slightly, 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, estimating the contribution of principal training programs to principal skills remains a challenge.

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 improves with their tenure at a school. Béteille, Kalogrides and Loeb (Béteille, Kalogrides, & Loeb, 2012) find that in Miami-Dade County principal turnover negatively impacts 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 as they have longer tenures in their schools. Recently, Husain, Matsa and Miller (2018) document that male teachers are more likely to leave schools led by female principals.

While these lines of research can be instructive, they do not provide direct guidance on the particular actions a current school leader 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 school leaders allocate their time across school leadership tasks relate to student learning gains (Grissom, Loeb, & Master, 2013; Horng, Klasik, & Loeb, 2010), or how teachers perceive their principals' effectiveness in particular leadership behaviors predict 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 synthesize 42 studies in a meta-analysis estimating the effects of five categories of school leader behaviors on student, teacher and school outcomes. There have been previous efforts to

synthesize the effects of school leaders (defined operationally in our study as school principals and assistant principals, to which we refer henceforth collectively as "principals") on schools (Hitt & Tucker, 2016; Marzano, Waters, & McNulty, 2005; Osborne-Lamplin, Folsom Sidler, & Herrington, 2015; Robinson, Lloyd, & Rowe, 2008; Witziers, Bosker, & Krüger, 2003). Our systematic review 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 (*cf.*, (Blase & Blase, 2004; Zepeda, 2013), we find limited evidence of the relative importance of instructional management strategies compared to other leadership behaviors. Third, the preceding findings are based almost entirely on observational studies because the causal evidence base on school leadership behaviors is non-existent. 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.

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

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. 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 pre-service preparation of principals. Nevertheless, there were over 90,000 public school principals currently employed in the 2015/16 school year, 40 percent of whom had at least four years of experience in their current schools, and for whom there was an exit rate from the principalship of only 10 percent in 2016/17 (U.S. Department of Education National Center for Education Statistics, 2017). Thus, 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.

In Figure 1, we present a theory of action around how principal behaviors influence student, teacher and school 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. 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 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 bi-directional 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). While most of these guides acknowledge the many competing managerial demands of the principalship, they depart 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 and colleagues (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. We return to this topic in our discussion in an effort to synthesize what conclusions can and cannot be drawn about which behaviors to prioritize.

We recognize that we conflate in our model principals' allocation of time to particular behaviors and their skill in effectively executing those leadership behaviors. 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 interaction and non-linearity 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, Bosker and Krüger'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 (Hallinger & Heck, 1996; Leithwood & Jantzi, 2000) on student achievement, Witziers and colleagues document a small but significant direct effect (.02 to .04 *SD*) of school leaders on student achievement across 37 studies published between 1986 and 1996. Witziers' study focuses on *whether* school leaders influence student achievement, but not on *how* they do so.

The most extensively referenced meta-analysis of school leader behaviors is Marzano, Waters and McNulty'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 school leaders. 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 and co-authors review is both different in nature than ours and limited in its methodology. First, while Marzano characterizes 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. Next, Marzano and colleagues select one effect size from each study and average the effect sizes across studies. This has the potential of overstating the strength of the modeled relationship as the selected effect size is not representative of all relationships estimated in the study. Finally, like our study, Marzano's analysis relies exclusively on observational studies.

Robinson, Lloyd and Rowe (2008) compare the effects of two different types of leadership style and attempt to capture the effects that the behaviors which emerge from these styles have on student outcomes. Comparing 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.

More recently, Hitt and Tucker (2016) conduct a qualitative systematic review of frameworks within the empirical research literature that have been used to estimate the relationship between school leader 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, external relations).

Our framework is most similar to Osborne-Lampkin, Sidler Folsom and Herrington's (2015) systematic review of the relationship between principal behaviors and student achievement. Osborne-Lampkin and co-authors 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.

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

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

II. Methods and Measures

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) (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, Econ Lit, ProQuest, and PsycINFO databases, including both 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 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 September 25, 2018. After identifying articles of potential interest through the title search, the same author reviewed abstracts; if the articles met our initial inclusion criteria, they were retained. 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.

Inclusion Criteria

We restrict 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 require that studies be conducted in the United States and other developed countries, published in English, have examined K-12 school samples, and have been completed after 2000.[1] We make this last choice for both substantive reasons and due to the coverage of prior education leadership meta-analyses. First, we theorize that the 2001 re-authorization of the Elementary and Secondary Education Act ushered in a modern era of accountability both in the United States and globally (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, Waters and McNulty's (2005) sample ends in 2001, Witziers, Bosker and Kruger's (2003) in 1996, and Robinson, Lloyd and Rowe's (2008) include only five studies from after 2000. We also exclude studies for which we could not identify at least one measured relationship between a principal behavior and outcome of interest as defined in the next section. Finally, we include studies whose measures of the principal behavior-outcome relationship was quantitative in nature; thus permitting their inclusion in a meta-analysis. Notably, we do 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).

While 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), while 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, Sidler Folsom and Herrington's (2015) research synthesis, we find only one study in our sample that meets either of the top two categories of the Institute for Education Sciences (IES) levels of evidence (Strong or Moderate). The shortage of

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

studies of education leadership is an important topic to which we return in our discussion. However, we argue that by following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) and the Meta-Analyses of Observational Studies in Epidemiology (MOOSE) 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 dataset. 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 69 studies initially retained for review. The intra-class correlation for the unreconciled codes across the 49 studies for which at least one of the authors identified a potential behavior-outcome pair was 0.799.[2] We then reconciled instances of discrepancy, including exclusion or inclusion, through a consensus decision-making process. We used the following codes to characterize study features which constitute aspects of the inclusion criteria and the 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)

[2] This intra-class correlation statistic excludes the private/public school code. In our final coding of the studies, all 42 studies include public schools in their sample and only four include private schools. Since there is no "true" variability in private-public schooling, a one-item difference between coders reduces the overall ICC to 0.741. Since so few private school studies are in our sample, we do not include this variable in our analysis.

administration, and (5) external relations. Grissom and Loeb identify 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, B., 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 impact of principals' attention to staff relationships and well-being.

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 non-instructional staff, while at the same time also capturing actions taken to develop a safe school environment, respond to staff concerns, and engage with other school leaders. As an example, one study examines principals' time spent on different aspects of organizational redesign, looking at active management of roles and responsibilities in their schools (Day, C. Sammons, P. Hopkins, D. Harris, A.

Leithwood, K. Gu, Q. Brown, E. Ahtaridou, E. Kington, 2009), a behavior we consider organizational management.

Administration. Grissom and Loeb (2011) define these operationally 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 entailing compliance activities, standardized assessment implementation and school schedule management, as well as student service management, student supervision, and managing school attendance. For example, May, Huff and Goldring (2012) relate the time principals spend on building operations with measures of student growth, which we code 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 that entailed communication with the district office, community members, partners, or other outside stakeholders, as well as fundraising efforts. As one example, we coded for external relations when one article measured the total on-the-job time principals spent interacting with stakeholders outside the school (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, which included measures of teacher's emotional state in relation to teaching career, such as satisfaction, engagement, and collegial relationships; and (3) teaching practices which included observed or reported assessments of teaching practices determined by the 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 (or in the case of unpublished dissertations the year in which the dissertation was submitted for graduation purposes) 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 RAND.

***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 (Kindergarten-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 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 69 studies in our first review, there were relatively few sample or researcher overlaps. Thus, we created identifiers for 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 must include a plausibly exogenous variation in treatment, thereby restricting studies in this group to randomized control trials, regression discontinuities, difference-in-difference estimates, or instrumental variable estimation techniques. Given the lack of exogenous variation, we did not code any studies employing matching or structural equation modeling (SEM) as causal. 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 non-transparency (Higgins et al., 2013) to the meta-analytic process.

When the article included a predictor-outcome dyad, we extracted the coefficient of interest, recording which outcome category it fell in with the accompanying predictor from the five categorized 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.[3] 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 non-significant paths. In such instances, we contacted the authors for these details.

[3] 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., $p<0.05$ = t-statistic of 1.96) and used these to estimate the standard error.

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 SEM estimates.

Meta-Analytic Methods

We calculate individual effect sizes following standard procedures for calculating mean differences, correlation coefficients, partial correlations and standardized regression coefficients (A. M. Aloe & Thompson, 2013; Borenstein, 2009).[4] 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. Specifically, we specify the following model using Borenstein's nomenclature:

$$Y_{ij} = \mu + \zeta_j + \varepsilon_{ij}, \text{ where} \quad (1)$$

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.

[4] Specifically, we convert all outcomes into standardized effect sizes (Cohen's *d*) and associated standard errors (SE_d) using the following formulas: (a) mean difference: $d = (\overline{X}_1 - \overline{X}_2)/SD_{within}$, where $SD_{within} = \sqrt{((n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2)/(n_1 + n_2 - 2)}$ and $SE_d = \sqrt{((n_1 + n_2)/n_1 n_2) + (d^2/2(n_1 + n_2))}$; (b) correlation coefficient (r): $d = 2r/\sqrt{1 - r^2}$ and $SE_d = \sqrt{(4\,(1 - r^2)^2/(n - 1))/(1 - r^2)^3}$; (c) 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 (b), $SE_d = \sqrt{\left(4\left(1 - r_p^2\right)^2/df\right)/(1 - r^2)^3}$; and (d) standardized slope: *d* is the coefficient and $SE_d = \sqrt{\left((1 - R_y^2)/df\right)(1/(1 - R_x^2))}$.

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

$$Y_{ij} = \mu + \beta' \overrightarrow{PrinBehav_{ij}} + \zeta_j + \varepsilon_{ij}, \text{ where} \quad (2)$$

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 to the relationship between instructional management and the outcome.

Our inclusion criteria produce 527 effect sizes across 42 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 (2017), 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 non-independence of effect sizes within studies.[5] 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. In contrast, estimates from studies that contribute multiple effect size estimates contribute less to the overall estimate of the treatment effect.

[5] The robust variance estimation procedure (Tanner-Smith & Tipton, 2014) constructs weights as follows: $w_{ij} = 1/\{(V_j + \tau^2)(1 + (k_j - 1)\rho)\}$, 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 measures the expected correlation of effect sizes within study. For all estimates presented in the paper, we use the value ρ=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.

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 trimming the top and bottom five percent of the 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 re-estimate our results. Finally, while our main results pool bivariate and partial correlations, we test for differences in the signs and magnitudes across model types.

III. 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 101 studies. Both authors then engaged in a further screening process, reviewing the abstract, sample and methods sections of these 101 studies to determine whether the study measures of both principal behaviors and outcomes matched our analytic framework, and the study's sample was drawn from the United States or other developed 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 developing contexts such as, for example, India. This narrowed our sample to 69 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. We excluded 27 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.[6] Ultimately, after reconciling both authors' judgments, we narrowed our sample to 42 studies that included a quantitative measure of principal behavior and a student-, teacher- or school-level outcome.[7]

Study Characteristics

In Table 1, we report the characteristics for the 42 studies and 527 effect sizes included in the analysis. We present the full list of studies with selected codes in Appendix Table A1. The majority of studies were published in peer-reviewed journals, 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 31 included elementary samples, and 28 secondary samples. Over half of studies draw on principals as the unit of analysis, while 17 studies draw on teachers and three on students. Just one study of the 42 uses a research design that permits causal inferences, while the remaining 41 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, upon 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.

[6] A common example of the first instance was when a study summarized principal time 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.

[7] We identified 18 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 emails to 13 authors requesting additional information to permit the study to be included. We queried 10 authors (in five cases, the same authors as in the previous step) for information to allow us to use the precise standard error rather than an estimate.

As shown in Table 1, the majority of effect sizes are linked to principals' instructional management behaviors, and this type of principal behavior is included in over 85 percent of articles. Internal relations is the next most prevalent principal behavior, present in 38 percent of the studies. The remaining effect sizes are fairly evenly distributed across the organizational management, administration, and external relations predictors, all between five and seven percent of total effect sizes.

We identify seven different outcome categories across the included studies. As with principal behaviors, there is differential representation across the outcomes. Almost half of the effect sizes (44 percent) capture the impact of leadership behaviors on student academic achievement while less than one percent of effect sizes have student behavior as an outcome. Across teacher-level outcomes, both teacher well-being and teacher practices are the most prevalent, at 12 and 10 percent of total effect sizes, respectively. Very few of the studies examine teacher retention (2 percent of effect sizes). At the school level, school organizational health is the most prevalent outcome (30 percent of effect sizes), while fewer than one percent of effect sizes measure the effect of principal behaviors on their remaining 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 of the effect sizes the relationships between all principal behaviors and (a) student achievement, (b) teacher well-being, (c) teaching practices and (d) school organizational health outcomes. The distribution of the relationship 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 0.36 *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 -0.02 *SD* and 0.43 *SD*. However, the relationship between principal behaviors and teaching practices has an evident bimodal distribution and a substantially larger spread of effect sizes (IQR=0.20-0.85 *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 0.13 and 1.05 *SD*. We present density plots for each outcome, by principal behavior in Appendix Figure A1.

In Table 2, we present the main results of our meta-analysis in which we find 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 148 effect sizes nested in 20 studies that measure the relationship between principals' instructional management time allocation and skill and students' achievement a one standard deviation unit difference in principal instructional management is associated with a 0.14 *SD* difference [95% confidence interval (CI): 0.02, 0.26, 95% prediction interval (PI): -0.18, 0.47] in students' achievement.[8] A similar difference in principals' internal relations time and skill is associated with a 0.17 *SD* difference [95% CI: -0.02, 0.36, 95% PI: -0.27, 0.62] in student achievement scores across 30 effect sizes in 10 studies, though these estimates are only marginally significant. 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 four studies.

We find suggestive evidence that principals' focus on generalized support for teachers is more closely related to their sense of well-being, while principals' focus on instruction-specific support relates more closely to improvements in teachers' practices. Column 2 of Table 2 indicates a 0.34 *SD* positive relationship [95% CI: -0.01, 0.68, 95% PI: -0.64, 1.32] between well-being effect sizes and principals' efforts to improve internal relations. We fail to reject the null that there is no relationship between principals' instructional management efforts and measures of teacher well-being. Across 9 studies and 46 effect sizes in Column 3 of Table 2, we note a 0.34 *SD* relationship [95% CI: 0.09, 0.58, 95% PI: -0.32,

[8] We note important heterogeneity across the effect sizes we sample. One substantively meaningful way to capture heterogeneity in effect sizes across studies is the 95 percent 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_μ is the variance of the pooled mean. 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 -0.18 and 0.47 *SD*. We return to this topic in our Discussion.

0.99] 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 are likely heavily 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 (0.69 *SD*, 95% CI: 0.34, 1.03, PI: -0.54, 1.92) and support internal relations on school organizational health effect size outcomes, though only the first is 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 of 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 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; though the sub-setting of our sample of effect sizes increases the noise in 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 much more modest 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 (0.87 and 0.82 *SD*, respectively) at the elementary level; however, these differences are much smaller at the secondary levels, and indistinguishable from zero in the latter case. These 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, A. 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-United States 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 (0.18 *SD*, 95% CI: 0.03, 0.33, 95% PI: -0.29, 0.66) and school organizational health (0.81 *SD*, 95% CI: 0.29, 1.34, 95% PI: -1.01, 2.63) effect sizes.

Do different principal behaviors affect the same outcomes differently? Do the same principal behaviors affect different outcomes differently?

Of critical interest to the resource-limited practicing principal is how to allocate the scarcest of resource: time. Thus, while our meta-analysis reveals general positive correlations between all principal behaviors and our outcomes, principals would benefit from understanding whether some behaviors have a stronger effect than others on desired improved student, teacher and school outcomes. Ultimately, the best 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 multi-collinearity 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 study data.

We find supporting evidence that a principal's focus on instructional management is related to similar signed and sized student achievement and school organizational health effects as does a principal's focus on internal relations. Panel A of Figure 4 reveals the positive relationship across a small sample of eight studies with student achievement effect sizes related to both instructional management

and internal relations behaviors. We find a strong correlation between the two sets of principal behaviors' effect sizes (r=0.65). 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. We next test these ideas more formally in a meta-regression framework.

We find suggestive evidence that while principals' instructional management behaviors have important relationships with student achievement, teaching practices and school organizational health, these appear to be no different from the relationship between other principal behaviors and the same outcomes. In Table 5, we present results of a meta-regression 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, there is a significant and moderately sized relationship between principals' time spent and skill level devoted to instructional management (0.158 *SD*, 95% CI: 0.02, 0.30), 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 noise 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 in 95 out of 100 instances the summary effect in our study population of the benefits of time spent and skill in Internal Relations will be no less than 0.2 standard deviations less effective and no more than 0.3 standard deviations more effective at improving student achievement than time spent and skill in Instructional Management. The other three principal behaviors are similarly bounded in their 95 percent 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 than other principal behaviors. However, for the three process outcome measures these estimates are quite imprecisely estimated, and we are unable to rule out very substantial pooled effect size differences.

To summarize the results from our meta-regression, we find evidence that 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 larger 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 to the other. This nuanced understanding of the value of instructional leadership accords with Grissom and Loeb's findings (2011; 2013). 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 find suggestive evidence that the joint 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, we find a negative relationship between achievement and organizational health outcomes (r=-0.25). Again, while we urge caution in comparing across these 10 observational studies, we believe these results are suggestive that principal actions produce multi-dimensional 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 percent 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. While 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 conduct 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 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. The imputation method in Column 2 reveals slightly attenuated results, but they are substantively the same as our main results in Table 2, with effect sizes on student achievement for instructional management and internal relations of .10 *SD* and .11 *SD*, respectively.

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 (A.

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 A2. We again interpret this sample subset cautiously given our small 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 to the bivariate relationships. Given that only one of our studies uses a randomized design to compare "treatment" and "control" groups, we take this as evidence that those studies that include covariate adjustments are more successful at accounting for omitted variable bias. While we are uncomfortable arguing that these are any more accurate results given their observational nature, we hypothesize that future meta-analyses in which a pool of causal studies exist alongside the studies surveyed in this paper might benefit from analyses that compare these future potential causal bivariate relationships to the observational partial correlations in this review.

IV. Discussion

We reach two major substantive conclusions in our meta-analysis. First, in a review of 42 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 strength of these relationships implies that a one standard deviation 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 non-randomized studies the difference in mean effect was negligible (0.05 standard deviation units difference from groups with standard deviations of 0.28 and 0.36 *SD*). We find no substantive variation in these patterns based on the level of the school or in comparing studies conducted in the United States to elsewhere. These results 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 non-instructional tasks towards instructional ones as some have suggested (e.g., Bambrick-Santoyo & Peiser, 2012) may be misguided. 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, Loeb and Master (2013) document, in Miami-Dade, principals spent only 12.7 percent 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 non-instructional tasks are critical.

Taken at face value, our findings imply that principals must effectively engage in these five leadership behaviors with little opportunity for relative efficiencies by focusing only on some. This is likely cold comfort to U.S. principals who report in the 2015/16 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. 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). If each school in the United States hired an additional principal, even if the newly hired principals allocated time identically, this could permit increases in the amount of time spent on instructional management on the order of 18 hours, or one-and-a-half standard deviations more than is currently spent. Based on the strong assumption of non-diminishing returns, the implied effect on student achievement would be on the order of 0.21 *SD*. For comparison sake, this is one-third the size of the black-white achievement gap (Bloom, Hill, Black, & Lipsey, 2008; S. Reardon, 2011). The average principal earned $95,700 in 2015/16 (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 while 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 0.21 *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. 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 strength 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 0.18 and positive 0.47 *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 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 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 or negative than the averages in our sample. Most importantly, as we note repeatedly, our meta-analysis relies on only one study[9] permitting causal inferences, compared to 41 that are observational in nature. In addition to the now familiar calls of all meta-analysts for future authors 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 either stressed 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

[9] This study (Silva et al., 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.

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, Loeb and Master's (Grissom et al., 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.

V. Conclusion

We pool results from 42 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.

References

**Indicates reference included in meta-analysis*

*Alig-Mielcarek, J. M. (2003). A model of school success: Instructional leadership, academic press, and student acheivement.

*Allen, Nancy Grigsby, Bettye 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. Retrieved from https://eric.ed.gov/?id=EJ1083099

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

Aloe, A., Tanner-Smith, E., Becker, B., & Wilson, D. (2016). *Campbell methods policy note on synthesizing bivariate and partial effect sizes*. Oslo. https://doi.org/10.4073/cmpn.2016.2

Avolio, B., Bass, B. (1997). Multifactor Leadership Questionnaire. Mind Garden.

Bambrick-Santoyo, P., & Peiser, B. M. (2012). *Leverage Leadership : A Practical Guide to Building Exceptional Schools*. San Francisco: Jossey-Bass.

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

*Bednar, J. (2018). Effective school leaders and student acheivement: An examination of 265 schools in northern Illinois. Proquest Dissertations Publishing. Retrieved from https://search.proquest.com/docview/2055729418?pq-origsite=gscholar

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*(4), 904–919. https://doi.org/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*(4), 289–328. https://doi.org/10.1080/19345740802400072

Borenstein, M. (2009). *Introduction to Meta-Analysis* (1st ed.). West Sussex, UK: Wiley & Sons.

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. https://doi.org/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*. 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*(1), 126–145. https://doi.org/10.3102/0162373716668028

*Burkhauser, S., Gates, S. M., Hamilton, L. S., & Ikemoto, G. S. (2012). *First-year principals in urban*

school districts. Santa Monica: RAND Corporation. 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. https://doi.org/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*(3), 283–309. https://doi.org/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: The 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*(1), 92–109. https://doi.org/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*(2), 232–253. https://doi.org/10.3102/0162373712437206

*Day, C. Sammons, P. Hopkins, D. Harris, A. Leithwood, K. Gu, Q. Brown, E. Ahtaridou, E. Kington, A. (2009). *The impact of school leadership on pupil outcomes*.

*Devries, R. (2017). *The instructional leadership practices of elementary principals of average needs/resource capacity school districts in New York State*. ProQuest Dissertations Publishing. Retrieved from https://search.proquest.com/docview/2010534116?pq-origsite=gscholar

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, *47*(2), 634–663. https://doi.org/10.1111/caje.12086

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

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*(2), 455–463. https://doi.org/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*(2), 110–141.

*Egley, R. Jones, B. (2005). Principals' inviting leadership behaviors in a time of test-based accountability. *Scholar-Practitioner Quarterly*, *3*(1), 13–24.

*Fancera, S. Bliss, J. (2011). Instructional leadership influence on collective teacher efficacy to improve school achievement. *Leadership and Policy in Schools*, *10*(3), 349–370.

Fullan, M. (2001). *The New Meaning of Educational Change*. 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 | RAND*. Santa Monica, CA. Retrieved from https://www.rand.org/pubs/research_reports/RR2812.html?adbsc=social_20190227_2707441&adbid=1100578268338577408&adbpl=tw&adbpr=22545453

*Gates, S. M., Hamilton, L. S., Burkhauser, S., Heaton, P., Pierson, A., Baird, M. D., … Rand Education (Institute). (2014). *Preparing principals to raise student achievement: implementation and effects of the New Leaders program in ten districts*. *RAND Corporation*. Santa Monica, CA: RAND Corporation. 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. *The Elementary School Journal*, *109*(4), 406–427. https://doi.org/10.1086/593940

*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.

Grissom, J. A., Bartanen, B., & Mitani, H. (2018). *Principal sorting and the distribution of principal quality* (Vanderbilt University Working Paper). Nashville, TN. Retrieved from https://s3.amazonaws.com/vu-my/wp-content/uploads/sites/2824/2018/06/20012209/distributions_june2018.pdf

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. https://doi.org/10.3102/0162373714523831

*Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness. *American Educational Research Journal*, *48*(5), 1091–1123. https://doi.org/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*(3), 369–395. https://doi.org/10.1162/EDFP_a_00210

*Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders. *Educational Researcher*, *42*(8), 433–444. https://doi.org/10.3102/0013189X13510020

Grissom, J. A., Mitani, H., & Woo, D. S. (2019). Principal Preparation Programs and Principal Outcomes. *Educational Administration Quarterly*, *55*(1), 73–115. https://doi.org/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*(1), 5–44. https://doi.org/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*(6), 654–678. https://doi.org/10.1177/1741143210379060

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*(1), 3–41.

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*(3), 659–689. https://doi.org/10.3102/0002831209340042

Heck, R. H., & Hallinger, P. (2010). Collaborative leadership effects on school improvement: Integrating unidirectional- and reciprocal-effects models. *The Elementary School Journal*, *111*(2), 226–252. https://doi.org/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.

https://doi.org/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. https://doi.org/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*(2), 531–569. https://doi.org/10.3102/0034654315614911

*Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*(4), 491–523. https://doi.org/10.1086/653625

*Houchard, M. (2005). Principal leadership, teacher morale, and student achievement in seven schools in Mitchell County, North Carolina. East Tennessee State University Digital Commons.

*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*(5), 728–771. https://doi.org/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*. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3259638

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

Kraft, M. A., Blazar, D., & Hogan, D. (2017). *The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence*. Providence. Retrieved from https://scholar.harvard.edu/files/mkraft/files/kraft_blazar_hogan_2016_teacher_coaching_meta-analysis_wp_w_appendix.pdf

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*(5), 711–753. https://doi.org/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.

*Leana, C. R., & Pil, F. K. (2006). Social capital and organizational performance: Evidence from urban public schools. *Organization Science*, *17*(3), 353–366. https://doi.org/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*(2), 112–129. https://doi.org/10.1108/09578230010320064

Leithwood, K., & Jantzi, D. (2005). A Review of Transformational School Leadership Research 1996–2005. *Leadership and Policy in Schools*, *4*(3), 177–199. https://doi.org/10.1080/15700760500244769

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*(12), 1181–1209. https://doi.org/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*(3), 315–336.

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*(4), 417–439. https://doi.org/10.1080/09243453.2012.678866

*May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*(2), 332–352. https://doi.org/10.1177/0013161X10383411

*Mees, G. W. (2008). The relationships among principal leadership, school culture, and student achievement in Missouri middle schools.

*Minckler, C. H. (2014). School leadership that builds teacher social capital. *Educational Management Administration & Leadership*, *42*(5), 657–679. https://doi.org/10.1177/1741143213510502

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

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), e1000097. https://doi.org/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. https://doi.org/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*(2), 174–190.

*O'Donnell, R. J., & White, G. P. (2005). Within the accountability era: Principals' instructional leadership behaviors and student achievement. *NASSP Bulletin*, *89*(645), 56–71. https://doi.org/10.1177/019263650508964505

OECD. (2013). *Synergies for better learning : An international perspective on evaluation and assessment.* Paris: OECD Publishing, Inc. https://doi.org/https://doi.org/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*(1), 18–70. https://doi.org/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. Retrieved from http://ies.ed.gov/ncee/edlabs

*Quinn, D. M. (2002). The impact of principal leadership behaviors on instructional practice and student engagement. *Journal of Educational Administration*, *40*(5), 447–467. https://doi.org/10.1108/09578230210440294

*Quint, Janet C. Akey, Theresa M. Rappaport, Shelley Willner, C. J. (2007). *Instructional leadership,*

teaching quality and student achievement: Suggestive evidence from three urban school districts. *MDRC*. New York,NY: MDRC. 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 developmen. *Leadership and Policy in Schools*, *10*(1), 63–83. https://doi.org/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: 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*(5), 635–674. https://doi.org/10.1177/0013161X08321509

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

*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*(4), 445–459. https://doi.org/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*(2), 200–230. https://doi.org/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*(5), 772–793. https://doi.org/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*(3), 293–331. https://doi.org/10.1080/00220270903527623

Spillane, J. P., & Lee, L. C. (2014). Novice school principals' sense of ultimate responsibility. *Educational Administration Quarterly*, *50*(3), 431–465. https://doi.org/10.1177/0013161X13505290

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., … Group, for the M. O. O. S. in E. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *JAMA*, *283*(15), 2008. https://doi.org/10.1001/jama.283.15.2008

Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, *46*(1), 31–56. https://doi.org/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. https://doi.org/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*(3), 496–536. https://doi.org/10.1177/0013161X11400185

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression.

Psychological Methods, *20*(3), 375–393. https://doi.org/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*(5), 891–899. https://doi.org/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.

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. https://doi.org/10.1002/jrsm.1064

*Ware, H. W., & Kitsantas, A. (2011). Predicting teacher commitment using principal and teacher efficacy variables: An HLM approach. *The Journal of Educational Research*, *104*(3), 183–193.

*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. Retrieved from http://digitalcommons.auctr.edu/challenge/vol15/iss1/3

Witziers, B., Bosker, R. J., & Krüger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*(3), 398–425. https://doi.org/10.1177/0013161X03253411

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

Figures

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

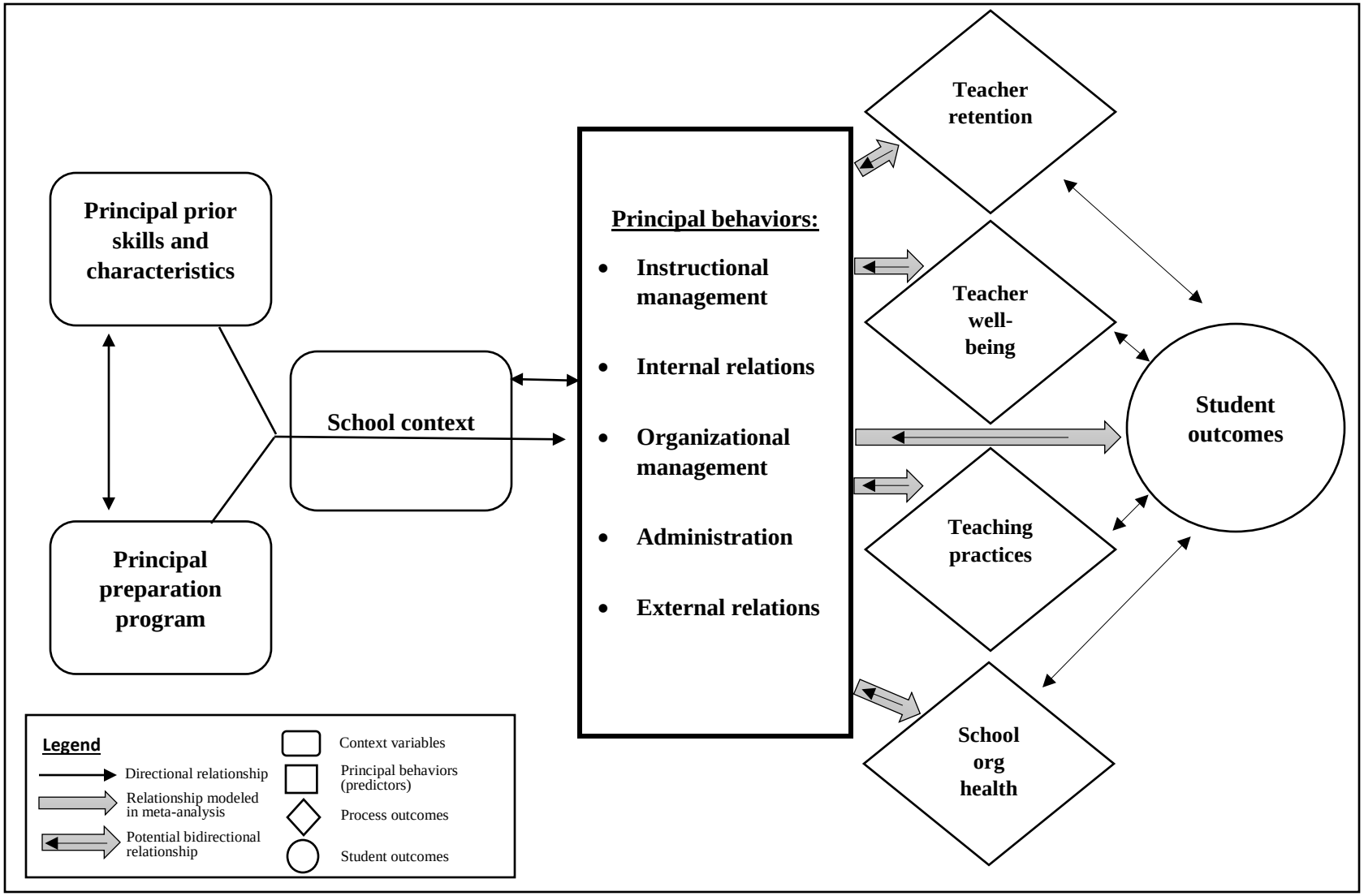
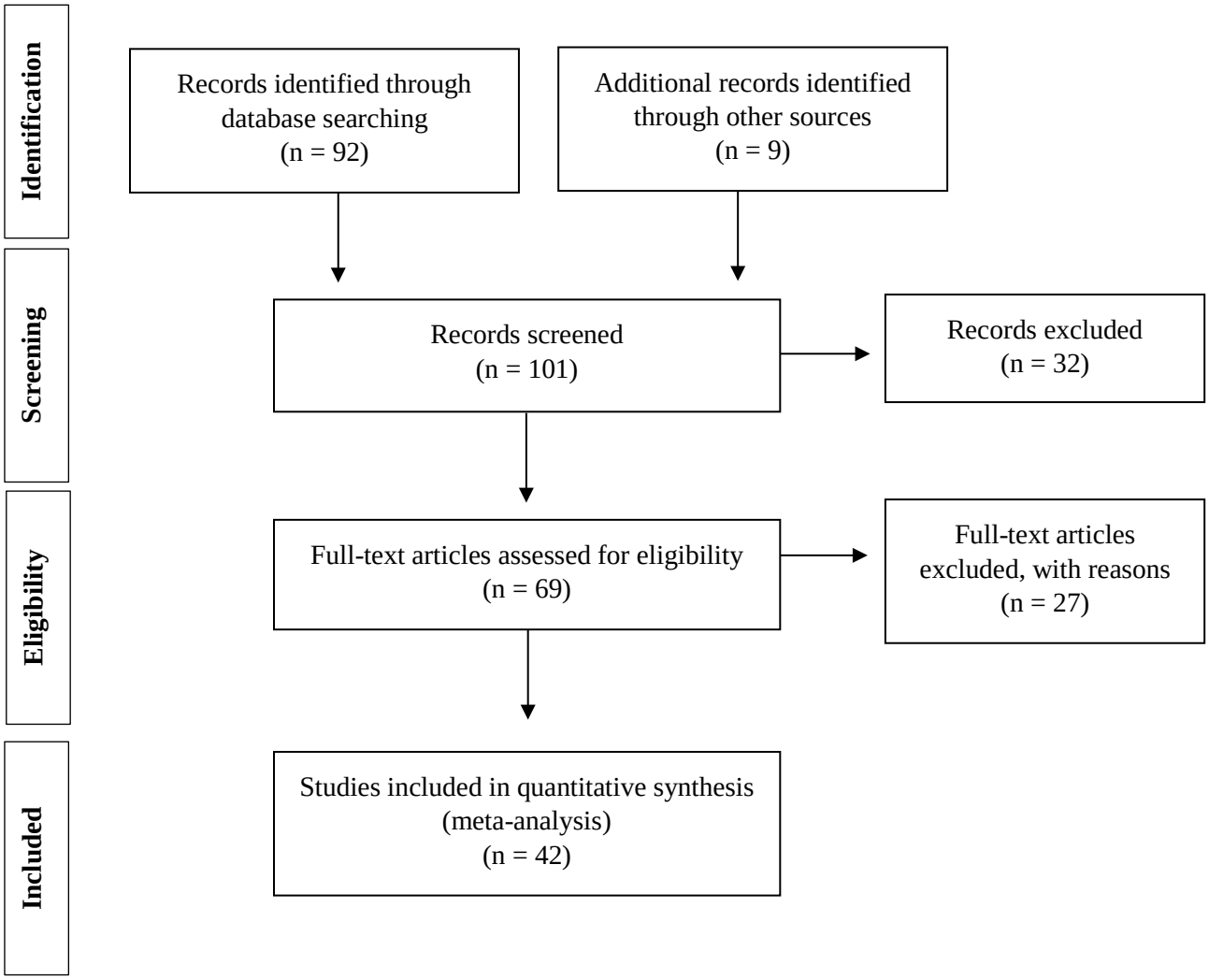
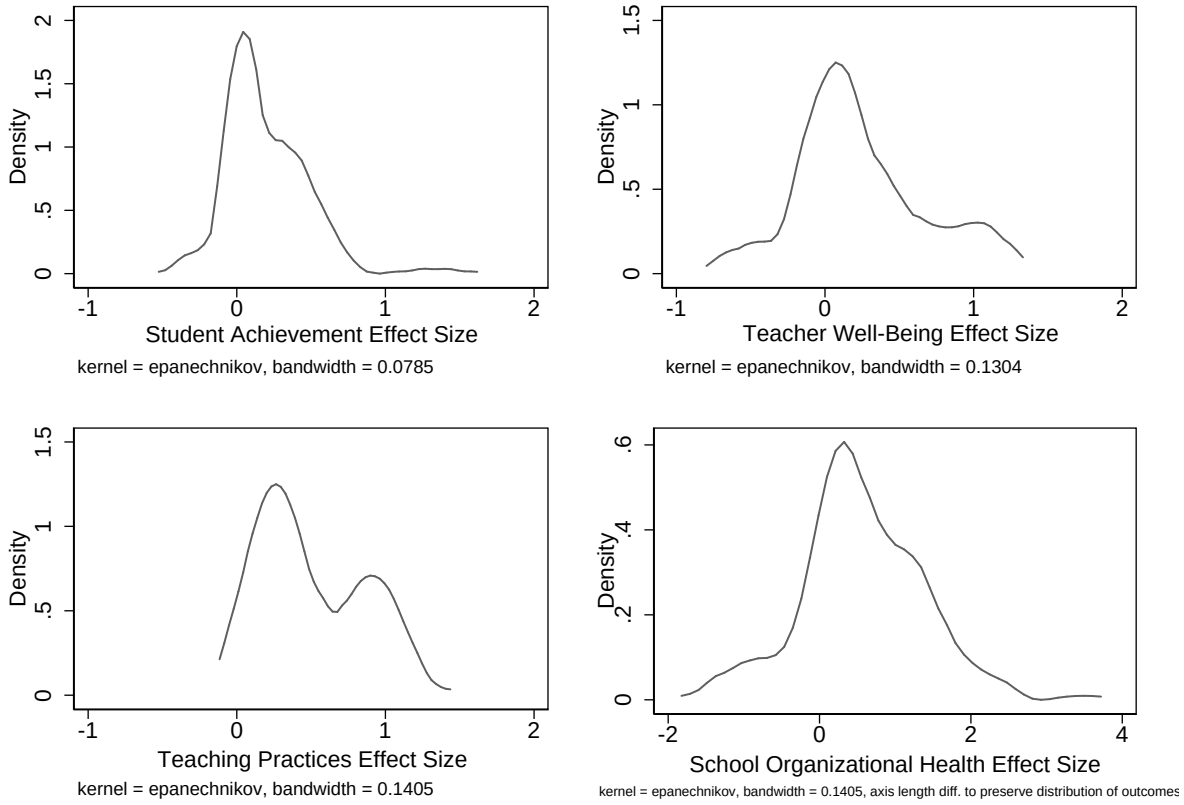


Figure 2. **PRISMA flow chart of literature identification, screening, eligibility and inclusion process**



Note: derived from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Moher, Liberati, Tetzlaff, Altman, & Group, 2009)

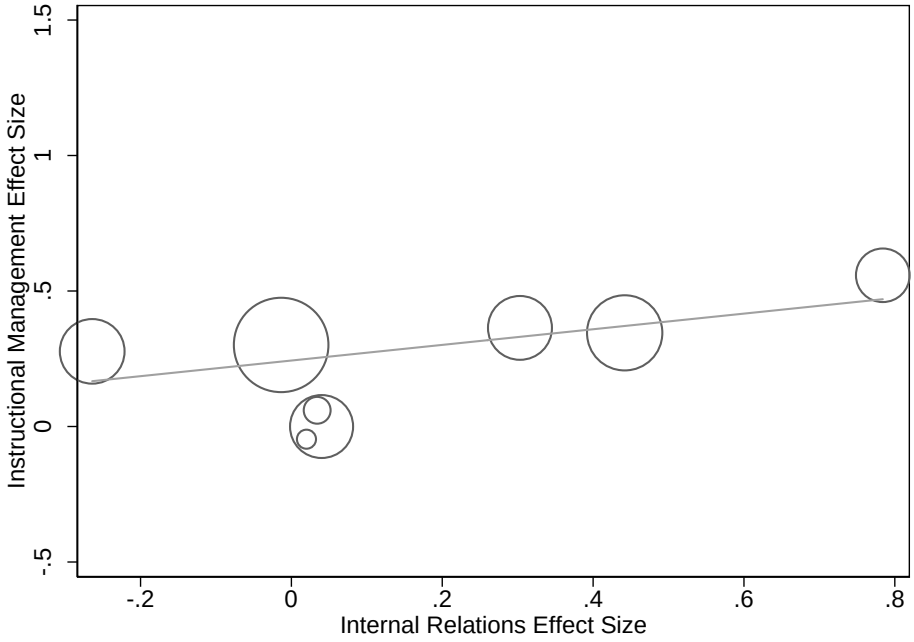
Figure 3. Kernel density of effect sizes of principal behaviors on student achievement, teacher well-being, teacher practices and school organizational health



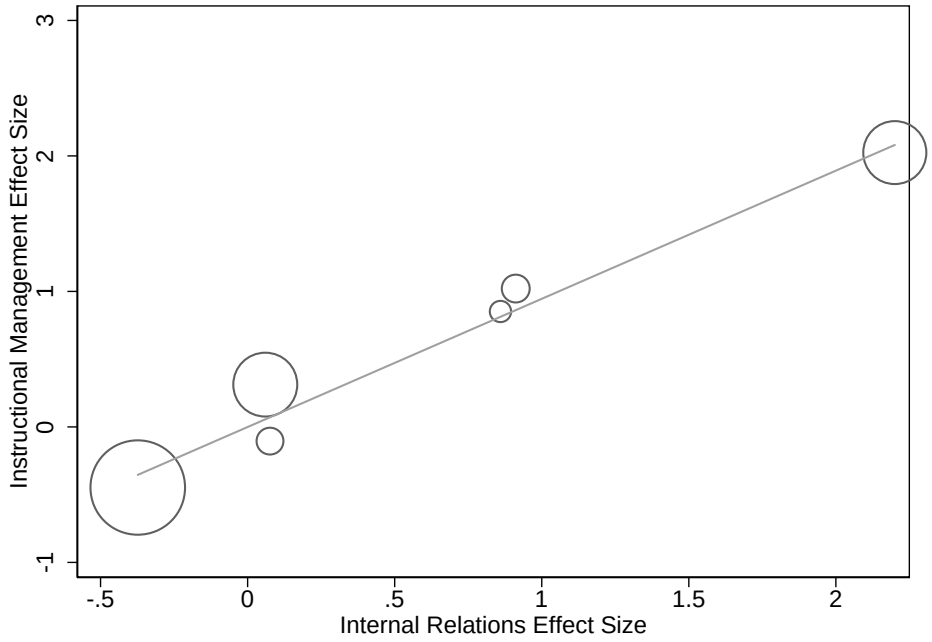
Note: k=230 for student achievement, 65 for teacher well-being, 55 for teaching practices and 160 for school organizational health.

Figure 4. Relationship between instructional management and internal relation behaviors effects on student achievement and organizational health

Panel A. Leadership Effects on Student Achievement

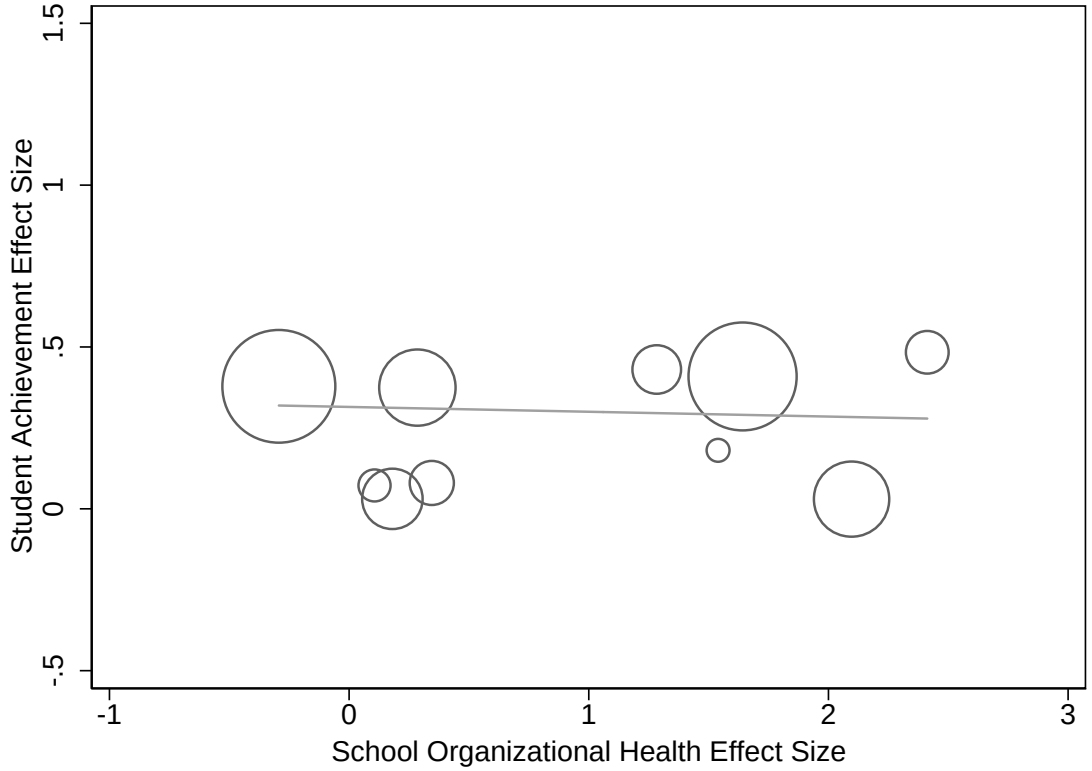


Panel B. Leadership Effects on Organizational Health



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. N=8 studies for student achievement and 6 studies for organizational health.

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. N=10 studies.

Tables

Table 1. Characteristics of studies included in meta-analysis

Code	Count [cases] studies	Proportion [cases] studies
Publication year		
2001-2005	[35]7	[0.07]0.17
2006-2010	[139]13	[0.26]0.31
2011-2015	[325]20	[0.62]0.48
2016-2018	[28]2	[0.05]0.05
Unit of analysis		
Student	[49]3	[0.09]0.07
Teacher	[194]17	[0.37]0.41
Principal	[284]26	[0.54]0.62
Publication type		
Peer-reviewed journal	[353]31	[0.67]0.74
Technical report	[120]5	[0.23]0.12
Dissertation/working paper	[54]6	[0.10]0.14
Country of study		
United States	[453]33	[0.86]0.79
International	[74]9	[0.14]0.21
School level		
Elementary	[349]31	[0.66]0.74
Secondary	[335]28	[0.64]0.67
Research design		
Causal (exogenous variation in treatment)	[1]1	[0.00]0.02
Observational	[526]41	[1.00]0.98
Principal behaviors		
Instructional management	[350]37	[0.66]0.88
Internal relations	[93]16	[0.18]0.38
Organizational management	[35]5	[0.07]0.12
Administration	[24]5	[0.05]0.12
External relations	[25]4	[0.05]0.20
Outcomes		
Student academic achievement	[232]23	[0.44]0.55
Student behavior	[2]1	[0.00]0.02
Teacher well-being	[65]10	[0.12]0.24
Teacher practices	[55]10	[0.10]0.24
Teacher retention	[11]3	[0.02]0.07

Principal retention	[2]1	[0.00]0.02
School organizational health	[160]22	[0.30]0.52
[k cases]N studies	**[527]42**	

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

Table 2. Pooled effect size estimates of the effect of principal behaviors on student, teacher and school outcomes

	Student Achievement	Teacher Well-Being	Teaching Practices	School Organizational Health
Instructional Management	0.143**	0.212	0.336*	0.689***
	(0.055)	(0.122)	(0.104)	(0.164)
n[k]	148[20]	31[8]	46[9]	116[19]
Internal Relations	0.173~	0.337~	na	0.616
	(0.079)	(0.140)		(0.323)
n[k]	30[10]	26[7]	8[3]	26[8]
Organizational Management	0.139	na	na	0.781
	(0.072)			(0.433)
n[k]	21[4]	4[2]	0[0]	10[4]
Administration	0.091	na	na	na
	(0.075)			
n[k]	18[4]	1[1]	0[0]	2[1]
External Relations	0.086	na	na	na
	(0.034)			
n[k]	15[4]	3[1]	1[1]	6[2]

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Cells with "na" are not estimated due to too few or no data.

Table 3. Pooled effect size estimates of the effect of principal behaviors on student, teacher and school outcomes, by school level

	Student Achievement	Teacher Well-Being	Teaching Practices	School Organizational Health
A. Elementary				
Instructional Management	0.091	0.080	0.249*	0.865***
	(0.052)	(0.069)	(0.070)	(0.220)
n[k]	104[16]	20[5]	35[7]	59[13]
Internal Relations	0.073	0.260	na	0.819
	(0.054)	(0.142)		(0.470)
n[k]	20[6]	23[5]		13[5]
Organizational Management	0.075	na	na	na
	(0.020)			
n[k]	19[3]			
Administration	0.019	na	na	na
	(0.013)			
n[k]	11[3]			
External Relations	0.079	na	na	na
	(0.031)			
n[k]	13[3]			
B. Secondary				
Instructional Management	0.106	0.160	0.327~	0.272*
	(0.084)	(0.122)	(0.139)	(0.115)
n[k]	78[10]	21[7]	21[7]	64[10]
Internal Relations	0.192	0.237	na	0.067
	(0.133)	(0.183)		(0.263)
n[k]	23[7]	20[5]		21[5]
Organizational Management	0.160	na	na	na
	(0.113)			
n[k]	19[3]			
Administration	0.093	na	na	na

	(0.074)			
n[k]	17[4]			
External Relations	0.080	na	na	na
	(0.036)			
n[k]	14[4]			

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Cells with "na" are not estimated due to too few or no data.

Table 4. Pooled effect size estimates of the effect of principal behaviors on student, teacher and school outcomes, United States results only

	Student Achievement	Teacher Well-Being	Teaching Practices	School Organizational Health
Instructional Management	0.183*	0.140	0.378~	0.811**
	(0.069)	(0.133)	(0.168)	(0.241)
n[k]	146[18]	19[6]	20[6]	99[13]
Internal Relations	0.173~	0.326	na	0.574
	(0.079)	(0.168)		(0.377)
n[k]	30[10]	21[6]		23[7]
Organizational Management	0.139	na	na	0.977
	(0.072)			(0.588)
n[k]	21[4]			8[3]
Administration	0.091	na	na	na
	(0.075)			
n[k]	18[4]			
External Relations	0.086	na	na	na
	(0.034)			
n[k]	15[4]			

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Cells with "na" are not estimated due to too few or no data.

Table 5. Meta-regression estimates of the relationship between principal behaviors and student, teacher and school outcomes, with instructional management as omitted principal behavior

	Student Achievement	Teacher Well-Being	Teaching Practices	School Organizational Health
Internal Relations	0.064	0.110	0.137	-0.086
	(0.098)	(0.219)	(0.237)	(0.551)
Organizational Management	-0.041	-0.175	na	-0.156
	(0.088)	(0.167)		(0.338)
Administration	-0.069	na	na	na
	(0.081)			
External Relations	-0.048	-0.385	na	-1.005
	(0.067)	(0.228)		(0.461)
Intercept (Instructional Management)	0.158*	0.252	0.316**	0.716***
	(0.063)	(0.149)	(0.098)	(0.163)
k[N]	232[23]	64[10]	54[9]	158[22]

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Cells with "na" are not estimated due to too few or no data.

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 percent of effect sizes for each outcome

	Student Achievement	Teacher Well-Being	Teaching Practices	School Organizational Health
Instructional Management	0.151**	0.209	0.378*	0.655***
	(0.044)	(0.115)	(0.107)	(0.135)
n[k]	134[18]	27[8]	41[8]	106[19]
Internal Relations	0.105	0.353*	na	0.472
	(0.053)	(0.136)		(0.261)
n[k]	25[8]	24[7]		22[7]
Organizational Management	0.077	na	na	0.308*
	(0.031)			(0.029)
n[k]	20[3]			9[3]
Administration	0.073	na	na	na
	(0.066)			
n[k]	15[4]			
External Relations	0.086	na	na	na
	(0.034)			
n[k]	15[4]			

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Trimming the top and bottom 5% removes 23 effect sizes from student achievement, 6 effect sizes from teacher well-being, 5 effect sizes from teaching practices, and 15 effect sizes from school organizational health. Cells with "na" are not estimated due to too few or no data.

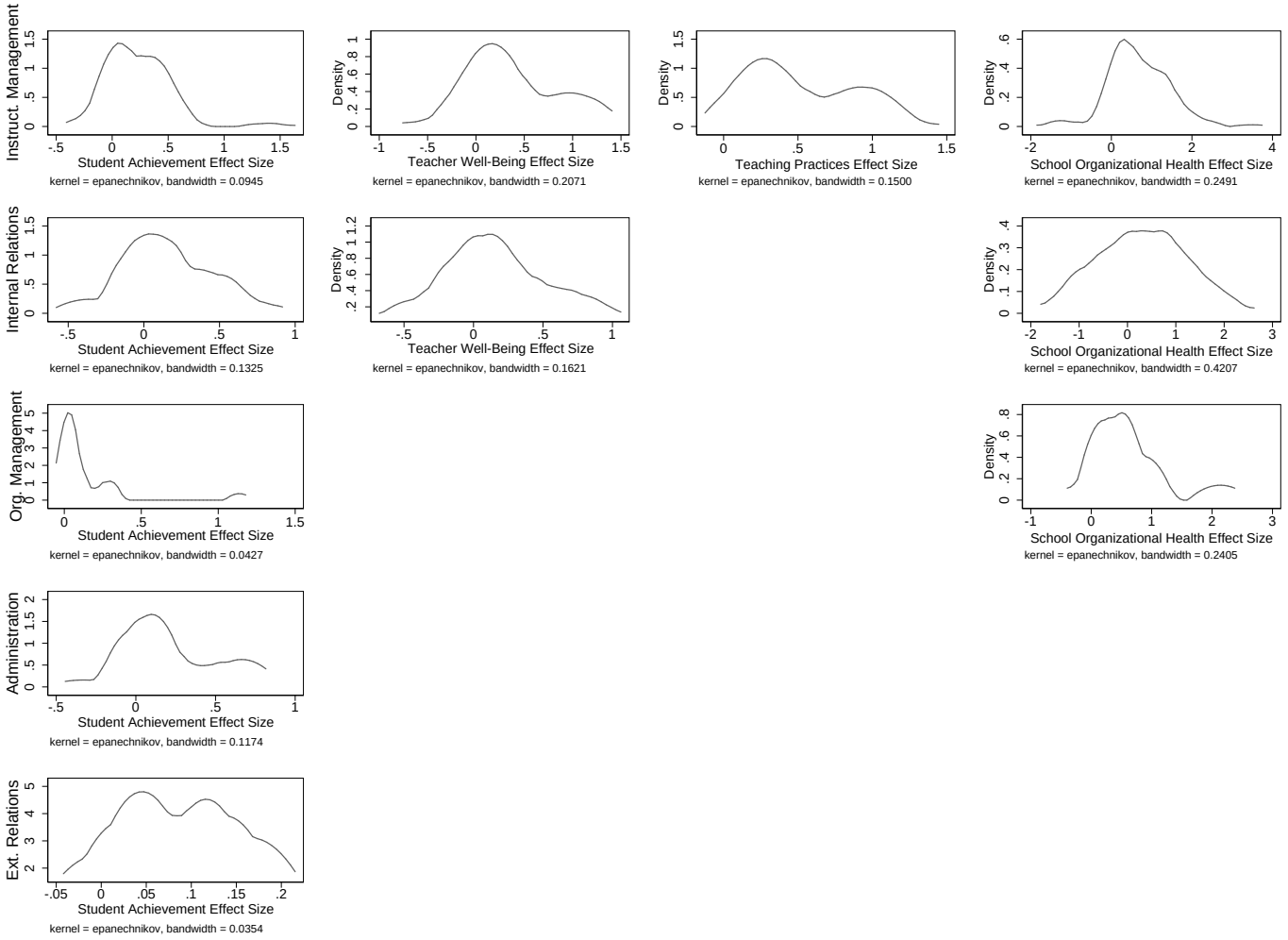
Table 7. Sensitivity analysis of effect size estimates of the effect of principal behaviors on student, teacher and school outcomes using study-level trim-and-fill method

	Unadjusted Study-Level Estimates	Study-Level Estimates with Imputed "Missing" Studies
Instructional Management	0.154**	0.101~
	(0.053)	(0.056)
n	[20]	[23]
Internal Relations	0.167**	0.114
	(0.063)	(0.075)
n	[10]	[13]
Organizational Management	0.042	0.034
	(0.065)	(0.064)
n	[4]	[5]
Administration	0.052	0.052
	(0.065)	(0.065)
n	[4]	[4][a]
External Relations	0.075	0.070
	(0.052)	(0.052)
n	[4]	[5]

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. [a] No "missing" studies found through trim-and-fill method. Study-level effect size estimates with standard errors clustered at study level reported in parentheses. n is the number of studies.

Appendix

Figure 1A. Kernel density of effect sizes of principal behaviors on student achievement, teacher well-being, teacher practices and school organizational health, by principal behavior



Note: different axis lengths to preserve full range of data and distributional shape of outcomes

Table A1. Studies Included in Meta-Analysis

Author	Year	Publication Type	Unit of Analysis	Effective Principal Sample Size	Locale	School Level	Principal Behaviors	Outcomes
Alig-Mielcarek	(2003)	Dissertation	Principal	146	United States	Elementary	Instructional management	Student academic achievement, School organizational health
Allen, Grigsby, & Peters	(2015)	Peer review	Teacher	6	United States	Elementary	Instructional management	School organizational health, Student academic achievement
Barnett & McCormick	(2004)	Peer review	Teacher	41	International	Secondary	Instructional management	School organizational health, Teacher practices
Bednar	(2018)	Dissertation	Principal	265	United States	Elementary	Instructional management	Student academic achievement, Teacher retention, Principal retention, School organizational health
Braun, Gable, & Kite	(2008)	Working paper	Principal	75	United States	Elementary & Secondary	Instructional management	Teacher practices
Burkhauser et al.	(2012)	Technical report	Principal	56	United States	Elementary & Secondary	Instructional management, Internal relations, Administration	Student academic achievement

Cancio, Albrecht, & Johns	(2013)	Peer review	Teacher	Not available[1]	United States	Elementary & Secondary	Instructional management, Internal relations	Teacher retention, Teacher well-being
Day et al.	(2009)	Technical Report	Principal	363	International	Elementary & Secondary	Organizational management	School organizational health
DeVries	(2017)	Dissertation	Principal	99	United States	Elementary	Instructional management	Student academic achievement
Dumay, Boonen, & Van Damme	(2013)	Peer review	Principal	120	International	Elementary	Instructional management	School organizational health, Student academic achievement
Ebmeier	(2003)	Peer review	Teacher	222	United States	Elementary & Secondary	Instructional management	Teacher practices, School organizational health
Egley & Jones	(2005)	Peer review	Teacher	Not available[1]	United States	Elementary	Internal relations	Student academic achievement, Teacher well-being, School organizational health
Fancera & Bliss	(2011)	Peer review	Principal	53	United States	Secondary	Instructional management, Internal relations, Administration	Student academic achievement, School organizational health
Gates et al.	(2014)	Technical report	Principal	249	United States	Elementary & Secondary	Instructional management	Student academic achievement

Geijsel et al.	(2009)	Peer review	Teacher	18	International	Elementary	Instructional management	Teacher practices, School organizational health
Graczewski, Knudson, & Holtzman	(2009)	Peer review	Teacher & Principal	49	United States	Elementary	Instructional management	School organizational health
Grissom & Loeb	(2011)	Peer review	Teacher, Principal, & Student	147	United States	Elementary & Secondary	Instructional management, Internal relations, Administration, Organizational management, External relations	Student academic achievement, Teacher well-being, School organizational health
Grissom, Loeb, & Master	(2013)	Peer review	Principal	127	United States	Elementary & Secondary	Instructional management	Student academic achievement
Horng, Klasik, & Loeb	(2010)	Peer review	Principal	39	United States	Secondary	Instructional management, Internal relations, Organizational management, External relations	Student academic achievement, Teacher well-being, School organizational health
Houchard	(2005)	Dissertation	Teacher	7	United States	Elementary & Secondary	Internal Relations	Teacher well-being, School organizational health
Hulpia, Devos, & Van Keer	(2011)	Peer review	Teacher	46	International	Secondary	Instructional management	School organizational health

Krüger, Witziers, & Sleegers	(2007)	Peer review	Principal	98	International	Secondary	Instructional management	School organizational health
Leana & Pil	(2006)	Peer review	Principal	88	United States	Elementary & Secondary	External relations	Student academic achievement, Teacher practices
Louis, Dretzke, & Wahlstrom	(2010)	Peer review	Principal	103	United States	Elementary & Secondary	Instructional management	Student academic achievement, Teacher practices
May & Supovitz	(2011)	Peer review	Teacher	51	United States	Elementary & Secondary	Instructional management	Teacher practices
May, Huff, & Goldring	(2012)	Peer review	Student	39	United States	Elementary & Secondary	Instructional management, Internal relations, Organizational managements, Administration, External relations	Student academic achievement
Mees	(2008)	Dissertation	Principal	79	United States	Secondary	Instructional management, Internal relations	Student academic achievement
Minckler	(2014)	Peer review	Teacher	13	United States	Secondary	Instructional management, Internal relations	Teacher well-being, Teacher practices, School organizational health
Mulford & Silins	(2011)	Peer review	Principal	131	International	Elementary	Instructional management	Student academic achievement, Student behavior
Ndoye, Imig, & Parker	(2010)	Peer review	Teacher	Not available[1]	United States	Elementary & Secondary	Instructional management, Internal	Teacher retention

							relations, Administration	
O'Donnell & White	(2005)	Peer review	Teacher	75	United States	Secondary	Instructional management	Student academic achievement
Orr & Orphanos	(2011)	Peer review	Principal	176	United States	Elementary & Secondary	Instructional management	School organizational health
Quinn	(2002)	Peer review	Principal	24	United States	Elementary & Secondary	Instructional management, Internal relations	Teacher practices
Quint et al.	(2007)	Technical report	Principal	42	United States	Elementary	Instructional management	Student academic achievement, School organizational health
Reardon	(2011)	Peer review	Principal	31	United States	Elementary	Instructional management	Student academic achievement
Shatzer et al.	(2014)	Peer review	Principal	45	United States	Elementary	Instructional management, Internal relations	Student academic achievement
Shen et al.	(2012)	Peer review	Principal	7,674	United States	Elementary & Secondary	Instructional management	Teacher well-being
Silva, White, & Yoshida	(2011)	Peer review	Student	3	United States	Secondary	Internal relations	Student academic achievement
Thoonen et al.	(2011)	Peer review	Teacher	32	International	Elementary	Instructional management, Internal relations	School organizational health, Teacher well-being, Teacher practices

Tuytens & Devos	(2011)	Peer review	Teacher	32	International	Secondary	Instructional management	Teacher well-being
Ware & Kitsantas	(2011)	Peer review	Principal	6,711	United States	Elementary & Secondary	Instructional management	Teacher well-being, School organizational health
Williams	(2009)	Peer review	Principal	81	United States	Elementary	Instructional management, Internal relations, Organizational management	Student academic achievement, School organizational health

[1] There are three articles that do not provide enough information to determine the effective principal sample size. The Cancio, Albrecht, & Johns (2013) sample includes 408 teachers, Egley & Jones (2005) sample 708 teachers from 30 districts. Ndoye, Imig, & Parker (2010) sample 328 teachers.

Table A2. Sensitivity analysis of effect size estimates of the effect of principal behaviors on student, teacher and school outcomes comparing bivariate and partial correlation study designs

	Student Achievement		Teacher Well-Being		Teaching Practices		School Organizational Health	
	Bivariate	Partial	Bivariate	Partial	Bivariate	Partial	Bivariate	Partial
Instructional Management	0.374**	0.052	na	0.028	0.633*	0.116*	1.242**	0.253*
	(0.085)	(0.047)		(0.026)	(0.184)	(0.033)	(0.238)	(0.096)
n[k]	88[9]	60[11]		16[6]	38[4]	8[5]	60[8]	56[11]
Internal Relations	0.237	na	0.534~	na	na	na	0.842~	na
	(0.125)		(0.177)				(0.378)	
n[k]	17[6]		16[4]				19[6]	
External Relations	na	0.086	na	na	na	na	na	na
		(0.034)						
n[k]		15[4]						

Notes: ~ p<.10, * p<.05, ** p<.01, *** p<.001. Pooled effect size estimates with robust-variance estimated standard errors reported in parentheses. For sample size, k is the number of effect sizes and n is the number of studies. Cells with "na" are not estimated due to too few or no data. Insufficient data for all organizational management and administration outcomes.