ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-581

Why Do School Districts Matter? An Interdisciplinary Framework and Empirical Review

David Blazar
University of Maryland

Beth Schueler
University of Virginia

What guidance does research provide school districts about how to improve system performance and increase equity? Despite over 30 years of inquiry on the topic of effective districts, existing frameworks are relatively narrow in terms of disciplinary focus (primarily educational leadership perspectives) and research design (primarily qualitative case studies). To bridge this gap, we first review the theoretical literatures on how districts are thought to affect student outcomes, arguing that an expanded set of disciplinary perspectives—organizational behavior, political science, and economics—have distinct theories about why districts matter. Next, we conduct a systematic review of quantitative studies that estimate the relationship between district-level inputs and performance outcomes. This review reveals benefits of district-level policies that cross disciplinary perspectives, including higher teacher salaries and strategic hiring, lower student-teacher ratios, and data use. One implication is that future research on district-level policymaking needs to consider multiple disciplinary perspectives. Our review also reveals the need for significant additional causal evidence and provides a multidisciplinary map of theorized pathways through which districts could influence student outcomes that are ripe for rigorous testing.

VERSION: May 2022

Suggested citation: Blazar, David, and Beth Schueler. (2022). Why Do School Districts Matter? An Interdisciplinary Framework and Empirical Review. (EdWorkingPaper: 22-581). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/58m4-fs65

Why Do School Districts Matter?
An Interdisciplinary Framework and Empirical Review

David Blazar, *University of Maryland*,
Address: 2205 Benjamin Building, 3942 Campus Drive, College Park, MD 20740
Contact Information: dblazar@umd.edu; 301-405-3324

Beth Schueler, *University of Virginia*
Address: Ridley Hall 268, 405 Emmet Street S., Charlottesville, VA 22904
Contact Information: beth_schueler@virginia.edu; 434-243-8620

Abstract

What guidance does research provide school districts about how to improve system performance and increase equity? Despite over 30 years of inquiry on the topic of effective districts, existing frameworks are relatively narrow in terms of disciplinary focus (primarily educational leadership perspectives) and research design (primarily qualitative case studies). To bridge this gap, we first review the theoretical literatures on how districts are thought to affect student outcomes, arguing that an expanded set of disciplinary perspectives—organizational behavior, political science, and economics—have distinct theories about why districts matter. Next, we conduct a systematic review of quantitative studies that estimate the relationship between district-level inputs and performance outcomes. This review reveals benefits of district-level policies that cross disciplinary perspectives, including higher teacher salaries and strategic hiring, lower student-teacher ratios, and data use. One implication is that future research on district-level policymaking needs to consider multiple disciplinary perspectives. Our review also reveals the need for significant additional causal evidence and provides a multidisciplinary map of theorized pathways through which districts could influence student outcomes that are ripe for rigorous testing.

The authors thank students at the University of Maryland, College Park Modibo Abbo, Kayla Bill, Shilpa Bista, Nardos Ghebreab, Rebecca Mann, and Xinyi Zhong for excellent research assistance; and Brooks Bowden, Dan Goldhaber, Meredith Honig, Cara Jackson, Matthew Kraft, Jal Mehta, Jennifer King Rice, Viviana Rodriguez, Laura Stapleton, and Tina Trujillo for valuable feedback on ideas and drafts. Address correspondence to David Blazar.

Introduction

Historically, school districts have been the primary actors charged with running U.S. school systems (Jacobsen & Saultz, 2012). Despite trends toward greater centralization of education policy over the past 50 years (Henig, 2013; Peterson, 2016) and the fact that states have constitutional authority for overseeing public education, the U.S.'s commitment to local control over education means that districts are viewed as a key mechanism for bringing to scale policy change in the service of improving student outcomes and increasing equity (Spillane & Thompson, 1997; Supovitz, 2006). Compared to school- or classroom-level programs, district-level policymaking may provide a more efficient approach to scaling reform. For example, salary schedules and recruitment efforts meant to attract high-quality and diverse teacher candidates often are set at the district level (Hanushek et al., 2004; Lankford et al., 2002). School boards and district leadership teams typically oversee the adoption of instructional resources, including curriculum materials and professional development (Little, 1989; Polikoff, 2018). Districts, often in collaboration with teachers' unions, generally set salary schedules (Grissom & Strunk, 2012), and have been the primary implementers of a new generation of teacher evaluation and compensation systems (McGuinn, 2012). The passage of the federal Every Student Succeeds Act in 2015 devolved even more authority to local education agencies.

But, what guidance do education researchers have for district leaders about how to run effective school systems? Despite longstanding interest in district-level policymaking (e.g., Anderson & Young, 2018; Leithwood, Sun, & McCullogh, 2019; Murphy & Hallinger, 1988; Opfer, Henry, & Mashburn, 2008; Rorrer, Skrla, & Scheurich, 2008; Trujillo, 2016), existing frameworks are relatively narrow in scope. What many researchers refer to as the "district effectiveness literature" is dominated by scholars in the educational leadership and organizational behavior traditions, who generally have focused on district policies related to instructional improvement and development. Without downplaying the importance of this line of inquiry, we argue that additional attention should be paid

to other disciplinary perspectives and evidence bases, including political scientists—often focused on district governance arrangements—and economists—often focused on the supply of and demand for teacher talent. Further, because of the dominance of the educational leadership perspective that traditionally has been grounded in qualitative case study analysis, prior reviews and frameworks may not have captured causally-oriented analyses that are more common in other traditions (despite the strengths of deep-dive qualitative case study work). In a recent review on this topic from the educational leadership perspective, Anderson and Young (2018) identified just eight quantitative and 21 mixed-methods studies seeking to identify components of effective districts. In comparison, when we expand the scope of our systematic review to cross disciplinary perspectives, we find 110 quantitative studies that fit our inclusion criteria. Relatedly, in Anderson and Young's review and ones like it, quantitative studies generally are treated in one broad category of research design, rather than being coded for rigor and the ability to identify practices that result in improved student outcomes.

This is not to say that the existing literature base on district effectiveness is without guidance for agency leaders. Qualitative case studies, descriptive statistics of survey items, and correlational analyses linking district policies to student outcomes reveal important hypothesized dimensions of effective districts, including building instructional leadership and reorienting the organization towards reform (Leithwood, 2010; Rorrer et al., 2008; Thompson et al., 2008; Trujillo, 2016). A central descriptor of effective districts that arises from this work is "coherent instructional systems", meaning those that bring together bureaucratic activities initiated by district administrators (e.g., standards adoption, testing requirements) with activities that support school leaders and teachers to operationalize and implement these higher-up policies in their classroom instruction (e.g., professional development) (Elmore & Burney, 1998; Honig & Hatch, 2004; O'Day, 2002). These frameworks directly guide our work, but where we pay closer attention to quantitative and causal inquiry.

To expand upon existing district effectiveness frameworks, we begin by reviewing the theoretical literatures on how districts affect student outcomes, looking across three disciplinary approaches to the topic: (i) organizational behavior with a close link to educational leadership, (ii) political science, and (iii) economics. We then systematically review the empirical literature testing these theorized pathways, regardless of whether the authors of these studies were trained in the disciplinary tradition typically associated with the district-level input under investigation. We limit studies in our review to those that use quantitative research designs, and we code studies and describe patterns of findings along a hierarchy of evidence for the purpose of drawing causal inference—with randomized trials at the top and "observational" (i.e., non-causal) studies towards the bottom. We argue that this approach allows us to make concrete recommendations for policy and future research. We further limited studies to the pre-Covid-19 period, while also recognizing a growing literature base that compares the district-level decisions made in response to the pandemic (e.g., mode of instruction; Halloran et al., 2021). The decisions that district leaders made in response to Covid-19 are distinct from those that constitute standard practice, and thus are set up for a separate analysis.

We find that studies most likely to support causal claims reveal benefits of district-level policies that cross disciplinary perspectives: higher teacher salaries and strategic hiring/replacement of teachers (a focus of the economics literature); lower student-teacher ratios/smaller class sizes and data use (a focus of organizational behavior and educational leadership); and school choice systems (a focus of political science), though findings on choice are more mixed. Because prior scholarship has tended to occur in disciplinary silos, our approach reveals a more complete picture of district effectiveness.

A second finding from our analysis, though, is that the empirical work could do a better job of using quantitative data to rigorously test existing theoretical frameworks on why and how districts matter. The majority of the quantitative articles we reviewed were observational in nature (72%), meaning that they are unlikely to support causal conclusions on their own. We recognize and discuss

the difficulty of randomly assigning policies at the district level. At the same time, "natural-experimental" designs (i.e., regression discontinuity and difference-in-differences), which also can credibly isolate the causal effect of a policy input (Murnane & Willett, 2010) too were relatively rare.

In addition, studies generally focused narrowly on student test scores to measure efficacy. Yet, theory describes a much more dynamic logic model linking district-level policies to a broad set of performance measures. For example, how do short-term test score gains translate into higher rates of high school graduation and college enrollment? What are the key mediating pathways by which districts affect these outcomes? All three disciplinary perspectives describe how district-level policies are likely to benefit students insomuch as they first influence teacher and teaching quality. However, our review of the quantitative literature identified just 14% of included studies that looked at teacher-level outcomes. In light of these limitations, our study provides a map of theorized pathways through which districts *could* influence desired outcomes and that are ripe for more rigorous testing.

Motivation

Why do scholars study effective school districts? The pursuit of "district practices and structures that matter for school performance and student achievement" (Anderson & Young, 2018, p. 2) are aligned to over six decades of research documenting large differences in outcomes between students from advantaged and less advantaged backgrounds, motivating the implementation of myriad policies—from an array of policy actors and school system levels—aimed at reducing inequality. Stemming back to the Coleman report from the 1960s, researchers often have focused on disparities in student outcomes and access to resources across school buildings (Peterson, 2016). While Coleman et al. (1966) found less variation in student outcomes across schools than they originally hypothesized (10% to 20% of the total variation), this analysis set the stage for examination of school policies and within-school resources (e.g., teachers) that drive student outcomes (e.g., Clotfelter et al., 2006).

The study of effective districts also stems from this earlier work in that scholars have explored districts as an additional source of variation in student outcomes (e.g., Bowers, 2010; Bowers, Ni, & Esswein, 2018). However, this literature base is relatively new as compared to the quantitative literature on schools, likely due to data constraints. While theoretical frameworks on district effectiveness were developed many decades ago (Murphy & Hallinger, 1988), statewide datasets that allow researchers to compare student outcomes across districts only became available on a broad scale in the decade after passage of the federal No Child Left Behind Act of 2001. In one of the earliest studies of its kind, Hedges and Hedberg (2013) used data from 11 states to decompose the variation in test-score gains across districts versus schools. A smaller degree of variation at the district level (roughly 5%, depending on the state and grade) compared to the school level (roughly 15% to 20%) was expected, given that districts are farther removed from students than schools. That said, in similar analyses of North Carolina data, Chingos et al. (2015) argued that documented variation in student test scores across districts (roughly 2%) was substantively meaningful. A 1 standard deviation (SD) difference in district effectiveness was associated with a 0.11 SD difference in student test scores, roughly equivalent to nine weeks of schooling. Following closely with the tradition of Coleman, Fahle and Reardon (2018) also documented substantial between-district variation in student test-score growth across all U.S. states, further arguing that these differences may be due to economic segregation along racial lines.

Additional studies have examined differences in varied measures of teacher quality across districts, bridging the longstanding variance decomposition literature with newer attention paid to the critical role that teachers play in improving student outcomes (e.g., Hanushek & Rivkin, 2010). For example, Goldhaber et al. (2018) found large teacher quality gaps across districts in North Carolina and Washington, as measured by teaching experience, licensure scores, and contributions to student test scores. Most of the inequitable sorting of disadvantaged students to lower-quality teachers occurred across districts and schools, rather than within schools. These patterns also can be found in

other states (e.g., Glazerman & Max, 2011). Similarly, Blazar, Litke, and Barmore (2016) found very large differences in measures of teachers' instructional practice across several urban school districts (upwards of 1 SD). Using the same dataset, Hill and colleagues (2015) found that districts explained much more of the variation in math-specific teaching practices than schools; the reverse was true for content-neutral teaching practices (e.g., classroom management). The authors suggest that these patterns may be related to instructional coherence, where districts often drive the adaptation of content-specific materials (e.g., standards, curricula) but less so for more generalized teaching practices. Together, these analyses reveal that variation in different metrics of teacher and teaching quality across districts translate into differences in student outcomes, suggesting that teachers may be one of the key mechanisms explaining the overall effectiveness of some districts over others.

Though the extant literature describing the magnitude of district effects is small relative to the literature on school and teacher effects, substantively meaningful differences in performance across districts motivate our search to better understand the sources of this variation and the policies best suited to closing gaps in outcomes. Therefore, we next provide findings from our synthesis of the theoretical literature on how school districts affect student outcomes, which guides the process by which we then conduct a systematic review of the quantitative studies examining this topic.

Synthesis of Frameworks on District Capacity to Influence Student Performance

To begin answering our driving question regarding characteristics of effective districts and effective district policymaking, we first describe and bridge theoretical frameworks across the three primary disciplinary approaches to the topic: (i) organizational behavior with a focus on educational leadership, (ii) political science, and (iii) economics. As we argue below, each has approached the topic of district effectiveness from distinct theoretical frames, meaning that focusing on just one perspective is likely to provide only limited guidance for how best to run effective school systems. By looking across these perspectives, we are able to generate a fairly exhaustive list of the major district-level

policies and inputs scholars have theorized should influence performance. (We list these policy inputs in Tables 2a-2c.)

Organizational Behavior and Educational Leadership

Prior reviews and treatments of "district effectiveness" stem largely from scholars in the traditions of organizational behavior and the closely related field of educational leadership, who have largely focused on district inputs related to instructional development. Scholars from these arenas historically argued that districts lacked the conditions necessary to make substantive impacts on teachers and students (DiMaggio & Powell, 1983; Meyer & Rowan, 1977), emphasizing the "loose coupling" between district policymaking and classroom environments (Orton & Weick, 1990; Weick, 1976) and the need to focus on the instructional core of educational practice (Elmore, 1996).

More recent treatments, though, suggest that districts have more capacity than originally proposed, particularly in the way central offices structure and draw on leadership teams in the service of supporting schools, teachers, and students (e.g., Supovitz, 2006). Although the educational leadership tradition's approach to studying district effectiveness has sometimes been described as a-theoretical (Trujillo, 2013), one key district-level feature with a richer theoretical basis relates to "coherence." Scholars highlight the role that central offices can play in crafting coherence by aligning standards, curricula, assessments, and policy (Honig & Hatch, 2004; Childress, Doyle, & Thomas, 2009). Through these efforts, district offices are thought to serve as key mediators of instructional reform (Cohen & Hill, 2008; Johnson et al., 2015). By shifting from a compliance-based to a school-support orientation, central offices also are thought to develop school capacity for reform, including through efforts to apply evidence to decision-making (Honig, 2012; Honig & Coburn, 2008).

Much of the organizational theory and educational leadership framework has been tested empirically in the form of single case studies profiling districts that had some success in educational improvement efforts. One prominent example is District #2 in New York City, which gained attention

due to its efforts turning around literacy and then mathematics outcomes for its lowest-performing students by adopting performance standards and intensive and coherent teacher professional development aligned to these standards (Elmore & Burney, 1998). Another example is San Diego, where the former superintendent of District #2 tried to replicate results more rapidly (Hightower, 2002). Expanding this work to several districts in Michigan, Spillane (2000) found that some were successful in bridging the gap between reformers' proposals and teachers' implementation of these ideas, while other districts failed to understand "the spirit" of reform and translate it into practice. These and other case studies highlight the importance of professional development, coordinated and coherent approaches to in-service learning, and strong instructional leadership (Firestone et al., 2005; Hightower, 2002; Little, 1989).

These patterns related to instructional leadership and system-wide coherence also are highlighted in several prior reviews on the study of district effectiveness, which all tend to focus on the organizational behavior and educational leaderships traditions (Leithwood, 2010; Opfer et al., 2008; Thompson et al., 2008; Trujillo, 2013). For example, in their synthesis of the literature, Rorrer et al. (2008) describe four key components of effective districts: (i) providing instructional leadership, (ii) reorienting the organization (e.g., reform-minded), (iii) establishing policy coherence, and (iv) maintaining an equity focus. Anderson and Young (2018) organize the literature similarly.

Political Science

When theorizing about how and why districts matter for student outcomes, political scientists have tended to focus on district governance structures, often arguing that these factors can create barriers to district-wide improvement. As such, they have theorized about the merits of reforms focused on agents more proximal to students, namely principals and teachers who serve as "street-level bureaucrats" enacting policy in classrooms (Weatherly & Lipsky, 1977). This perspective informs,

in part, the work of educational leadership scholars who have argued for bridging across the district bureaucracy and the classroom context in the form of purposeful and coherent systems.

More specifically, critiques of school board politics and dysfunction that are thought to disrupt district-level policymaking (Moe, 2011; Anzia, 2014) motivate proposals for alternative district governance arrangements such as school-level autonomy, school choice (Chubb & Moe, 1990), mayoral control (Henig, 2009; Wong et al., 2007), and portfolio management models (Henig & Bulkley, 2010). Many of these approaches are theorized to shift the venue of decision-making in ways that decrease the dominance of interest groups in the policymaking process in favor of policies that would prioritize student interests.

Politically oriented scholars have been somewhat hamstrung from rigorously examining the impact of novel governance arrangements on district-level performance given that alternatives to the traditional elected local school board are relatively rare and tend to be scattered across states (rather than across districts within states). Between-state variation—in both policy inputs and student outcomes—that is necessary to compare the efficacy of unique governance structures has been difficult to capitalize on; until recently, each state administered its own assessment, making cross-state outcome comparisons quite difficult without methodological compromises (Ho, 2008). Instead, political scientists have tended to focus again on case studies of districts that have, for example, abandoned traditional collective bargaining arrangements, increased school choice or school-level autonomy, or embraced portfolio management models (e.g., Bulkley et al., 2010) to generate theory about which district-level policies work and which do not.

Economics

Labor market theory identifies districts as key agents in the human capital and human resource pipeline. Economists of education have therefore tended to focus on district inputs related to recruitment and retention of teachers as a key pathway for influencing student achievement outcomes.

This perspective overlaps to some extent with discussion from organizational behaviorists and political scientists on policy being enacted by teachers inside schools and classrooms. In order to build coherent instructional systems, districts need to identify, recruit, hire, and retain teaching talent.

At the same time, the economics literature focuses much more on who teachers are and where they come from, rather than on teachers' day-to-day work with students. For example, a broad literature on teacher labor markets indicates that teachers' decisions about where to teach are influenced by contextual factors at the district level, including geography or proximity (Boyd et al., 2004) and student demographics (Clotfelter et al., 2011). In turn, following theory on the supply of and demand for teacher talent, the economics literature related to district effectiveness has focused on the importance of salary and financial incentives in determining whether and where individuals teach (e.g., Hanushek et al., 2004; Lankford et al., 2002; Murnane & Olsen, 1990). The Roy model of occupational choice (Roy, 1951) states, more specifically, that individuals' decisions about whether to teach or enter another labor market depends on: (i) average wages in one profession versus the other, where teachers' wages generally are set at the district level (Grissom & Strunk, 2012); (ii) the extent to which higher-skilled employees earn higher wages; and (iii) the extent to which skills relevant to each profession are positively correlated. Economists emphasize that non-random sorting on these factors likely contribute to differences in teacher effectiveness and, in turn, student outcomes across districts.

A more recent line of research involving some economists of education has examined the efficacy of district-wide turnaround reforms that often include but are not limited to human resource policies (e.g., Harris & Larsen, 2016; Chin et al., 2019). We view district turnaround efforts as often sitting at the intersection of all three disciplinary perspectives. Theoretical foundations for these policies tend to consider the politics and governance reforms that result in a district turnaround, the implementation of the reform often through market-based approaches, and the specific reform features that often are aligned with instructional improvement and educational leadership perspectives.

However, we introduce district turnarounds in this section on economics, as published quantitative studies have focused primarily on using econometric methods to estimate the *causal* impact of a package of reforms, rather than the politics and governance structures guiding them, or instructional resources that come with these changes. To do so, researchers have compared student achievement growth in a turnaround district to similar districts elsewhere in the state not experiencing this policy. This is unlike the political science and educational leadership literatures, described above, which have more often used single case studies without formal comparison groups to generate theory on district effects. Further, when researchers have speculated about sources of turnaround success, they often have focused on the introduction of competition and new approaches to teacher compensation and evaluation (e.g., Chin et al., 2019; Harris & Larsen, 2016), both core tenets of economic theory.

Need for Cross-Disciplinary Synthesis

Our review of these three theoretical literatures identifies multiple ways in which district-led policymaking *is hypothesized* to benefit district-wide performance, while also identifying key gaps in our knowledge base. As noted above, prior reviews covered the organizational behavior and educational leadership research traditions that have focused on ways in which district leaders interact with schools and teachers primarily around instructional reform. Yet, our review of broader theoretical literatures suggests that there likely are other pathways by which districts could affect student outcomes.

Not accounting for these other potential "omitted variables" also could limit readers' ability to draw robust, causal conclusions. As Anderson and Young (2018), Leithwood (2010), and Trujillo (2013) all point out, efforts to identify characteristics of effective districts in the educational leadership research they reviewed often suffered from the use of "outlier" designs. That is, districts were selected for qualitative case study either as success stories, or low performers undergoing turnaround. Selection on the dependent variable makes it unclear whether such characteristics vary across districts, let alone whether these characteristics explain variation in outcomes.

Comparatively, other research traditions (e.g., economics) have been more attune to drawing causal conclusions. At the same, this approach has required scholars to consider a relatively narrow set of inputs, and generally has focused on human resource policies and financial incentive schemes that are a focus of labor economics theory. We therefore see a need to examine the quantitative empirical literature, testing the full range of district effectiveness theories across disciplines.

Methods for Systematic Review of the Quantitative Empirical Evidence

To fill gaps in the current literature on district effectiveness, we next describe the methods we used to search and synthesize the quantitative literature testing myriad district-level inputs across disciplinary perspectives that are thought to impact district performance and student outcomes.

Inclusion Criteria

There were three primary ways that we restricted the set of quantitative empirical studies we included in our systematic review, which were related to the: (i) unit of analysis under study, (ii) outcome measures examined, and (iii) research design. We did not impose any restrictions on publication type, nor on geographic location as long as the context was one in which schools have been governed by some local authority (such as local education agencies in the U.S). While we did not place any lower bound on publication date, we limited studies to the pre-Covid-19 period (i.e., policy inputs and student outcomes captured before spring 2020) to focus on district-level policies that generalize beyond a pandemic. Though not an inclusion criterion per se, we also note for readers that we excluded studies focused only on district-level expenditures. Given our theoretical grounding, we are interested in the specific policy levers that districts pick up, rather than the financial resources that allow them to do so (for discussion of this topic, see Jackson, 2018a).

Districts as the Unit of Analysis. First, we included only quantitative studies that both focused on policies implementable at the district level and examined variation in policies across districts. This means, for example, that we excluded studies of state takeover of districts. Although

this governance change may vary across districts, it is primarily the state—rather than the district—that makes the determination about whether to place a particular district under takeover, and we were interested in understanding how districts can best impact key outcomes. (That said, our sample includes some studies of district turnarounds that occurred in the context of state takeover but that were not attempting to isolate the takeover component.) Similarly, we recognize that there are many interventions that *could* be rolled out at the district level (e.g., class size reduction, teacher evaluation) but that researchers have studied using variation between schools, teachers, or students.

We set this criterion for both substantive and practical reasons. Substantively, our focus is on district-level policymaking. Studies not conducted at the district level likely ignore general equilibrium effects of rolling out an intervention districtwide that are important for district-level policymakers to consider. For example, we excluded studies of school closure that used student- or school-level variation only. Knowing the effect of school closure on the students who attended the closed school is important, but does not provide an answer about the overall effect of closure on district-wide productivity. Short-term, negative effects may be offset by avoiding sending future students to a closed, low-performing school. Similarly, school choice systems that operate within school district boundaries may have different effects for students exercising choice than for students remaining in neighborhood schools. As a result, we included only those studies that examined district-level policy changes and effects, and leveraged this district-level variation in their analyses. Focusing on district-level variation was also practically necessary because we quickly realized that including every study of a policy that *could potentially be implemented* at the district level would explode our sample and make the task unmanageable.

Focus on Student or Teacher Outcomes. Second, we focused only on studies that quantitatively examined the relationship between some district-level input and either a measure of student performance or teacher quality. We view students as the ultimate beneficiary of any

educational policy, and so prioritized studies that examined links to varied student outcomes. More specifically, student outcomes include test-based achievement, course grades, school behaviors (e.g., absences, suspensions), high school graduation, and college-going.

We also included studies that examined teacher quality measures given that all three disciplinary traditions describe the critical role of teachers in mediating the effect of district policies on students. Teacher-level outcome measures include value-added to student test scores, observed quality of instruction, and teacher characteristics that have been shown to impact student outcomes including classroom experience (Papay & Kraft, 2015) and turnover (Ronfeldt et al., 2013). We also included studies that used as outcome measures teachers' certification, education, and licensure scores. While there is debate about whether these are good proxies for teacher quality (Wayne & Youngs, 2003), in practice, districts use such measures for hiring, compensation, and promotion decisions.

Research Design. Third and finally, we required that studies included in our analytic sample attempted, in some way, to limit biases due to the endogeneity of district-level policies. In other words, districts often self-select the policies and practices that leadership teams deem best suited for their schools, teachers, and students, which could lead one to overstate (or understate) the true impact of these policies on desired outcomes. We decided on a moderate to low bar for supporting causal claims, as very few studies accounted fully for non-random selection of district-level policies using experimental designs. Further, if we focused only on "natural experiments" (i.e., regression discontinuity and difference-in-differences), we would have been left with studies of those district-level inputs generally of interest to the economists historically most likely to wield these analytic tools. Therefore, we required only that studies control, in some way, for non-random selection of district-level policies. Studies could not be purely correlational.

At the same time, we grouped studies into two categories to describe research design quality, and we use this schema when interpreting results. The first category includes three research designs

that generally are thought of as being able to support causal conclusions: randomized control trials, the gold standard of causal inference; regression discontinuity designs, which exploit variation in treatment status due to strict eligibility thresholds and meet the highest evidence standards of the federal What Works Clearinghouse (WWC) repository for education research; and difference-in-differences designs, which examine how changes in outcomes over time differ between one set of units (in this case districts) that experienced a policy shock, relative to changes in outcomes for comparison units (districts) that did not experience this shock (Murnane & Willett, 2010). The second broad research design category includes all observational studies, which we define as those studies that attempt in some way to limit bias due to non-random selection but generally do not capitalize on exogenous variation and are therefore thought to fall short of supporting causal claims on their own. This set of designs includes multivariate regression, propensity score matching, panel data with one-way fixed effects, and instrumental variables estimation without the presence of a policy shock. Including studies with variation in research design quality facilitates a direct discussion of the tradeoff between methodological rigor and comprehensiveness of district-level policies examined.

Search Procedures

To identify the studies that fit these inclusion criteria, we conducted a systematic review of the literature in three phases. First, we identified articles using the electronic databases Academic Search Premier, Econ Lit, Ed Abstracts, ERIC, Google Scholar, ProQuest, and PsycINFO. We searched databases using terms that represented the broad array of district-level inputs identified in our review of the theoretical literature and scanned titles and abstracts for signs of whether studies met our inclusion criteria.[1] Second, to cross-check our search process, we reviewed the reference lists from

[1] Specific combination of search terms included: "education" AND "school district" OR "local education agency" OR "local education authority" OR "local government authority" OR "local authority" OR "central office" OR "statewide" AND "student achievement" OR "student test" OR "effective" OR "teacher effective" OR "teacher quality" OR "teacher retention" OR "teacher retain" OR "teacher recruit" AND "salary" OR "benefits" OR "hiring" OR "transfer" OR "evaluate" OR "school board" OR "mayoral control" OR "teacher union" OR "collective bargain" OR "portfolio model"

prior reviews (e.g., Anderson & Young, 2018; Leithwood, 2010; Trujillo, 2013) and from all studies that met our inclusion criteria. Finally, we contacted leading scholars in the field including many authors of the articles included in this analysis to solicit help identifying analyses we missed.

Coding and Analysis

We coded the resulting sample of studies on a number of dimensions related to their features and findings. We coded publication year, publication type, and country. We tracked the outcome measures for each study, as well as the independent variables that map onto our theory-based list of district-level inputs (see Tables 2a-2c). We also coded the research design used in each study. Finally, we generated brief summaries of the findings of each study, including the magnitude of the relationship between a given district-level policy and outcomes.[2] We analyzed the resulting data by tabulating descriptive statistics on each of our codes, and then by reviewing and synthesizing the findings by district-level input.

Findings from Systematic Review of the Quantitative Empirical Evidence

Characteristics of Included Studies

Our search identified a total of 110 quantitative studies that met our inclusion criteria. In Table 1, we describe that the majority of studies in our sample (65%) were published in peer-reviewed journals, though the second largest category was unpublished dissertations (22%); the remaining studies came from book chapters and reports. A large majority of the studies came from the U.S.

OR "spending" OR "autonomy" OR "instructional leader" OR "data use" OR "curriculum" OR "textbook" OR "professional development" OR "planning time" OR "peer collaboration" OR "formative assessment" OR "personalized learning" OR "technology" OR "alignment" OR "class size" OR "learning time" OR "configuration" AND "quantitative" OR "regression" OR "difference-in-difference" OR "difference in difference" OR "fixed effect" OR "regression discontinuity" OR "experiment" OR "RCT" or "randomized control".

[2] Although we tracked effect sizes, we opted against conducting a formal meta-analysis. Such an approach is best suited to pooling estimates of similar interventions that use similar outcome measures, and when research design quality is generally high (Cooper et al., 2019). However, for several policy input categories, our search only revealed a small number of studies. In many cases, these studies relied on different dependent variables, making meta-analytic techniques inappropriate (Cooper et al., 2019). Further, since a large portion of included studies are observational (i.e., not causal) in nature, we were reluctant to pool effect sizes which could mislead readers about the extent to which the extant literature has the capacity to provide credible causal estimates of district-level interventions.

(90%) while the remaining 10% came from five other developed countries: Canada, Israel, the Netherlands, Norway, and the United Kingdom.

In terms of research methodologies, observational studies (i.e., non-causal but not purely correlational) were by far the most common (72%). Of research designs that meet a higher bar for supporting causal inferences, difference-in-differences design was most common (21% of all studies). This pattern makes sense, as district-level policymaking is primed for comparison of treated versus untreated districts over time. Only three articles used experimental design, two of which relied on the same randomized experiment, assessing outcomes at different points in time. This again made sense given the challenges of randomizing interventions at the district level. Five studies used regression discontinuity designs, often exploiting variation in treatment status as a result of narrowly winning versus losing a district-wide vote. The vast majority of included studies (87%) examined student test-score outcomes, while far fewer examined relationships between district-level inputs and other student outcomes (i.e., course grades, high school graduation, college-going; 9%) or teacher-level measures (14%). Just four studies looked at links between district inputs and both student and teacher outcomes.

Most studies (76%) focused on just one policy input. Some studies focused on more, with 8% focusing on four or more inputs. Most studies that included multiple policy inputs are analyses of district-wide turnarounds rolled out as a package of reforms. Below, we describe results from these studies along specific policy inputs, while also cautioning readers that the design of these studies often means that the effect of one policy input cannot be fully disentangled from the effect of another.

Synthesis of Study Findings by Policy Category and Type

In Tables 2a-2c, we summarize findings across studies within each of the 26 total policy input categories, which stem from our theoretical review and align with the three major disciplinary perspectives: Table 2a covers "governance and politics"-related inputs, traditionally—but not always—the focus of political scientists; Table 2b covers inputs related to the "recruitment and

retention" of teacher talent, typically studied by economists; and Table 2c covers "instructional resources and development"-related inputs that ground theoretical frameworks of effective districts from organizational behavior and educational leadership scholars. We present the findings in this order, aligned to a generalized temporal order of district-level policymaking. First, the district-level conditions and structures need to be in place. Next, district leaders need to hire teaching talent. Finally, teachers need to have the resources and supports necessary to deliver high-quality instruction. We also acknowledge that the boundaries between categories and tables are not strict. Scholars from a variety of disciplines, including interdisciplinary teams, have examined inputs in each of the categories.

In each table, we provide the number of studies examining a given policy input, citations for each study with superscripts identifying research design, an asterisk for studies examining multiple inputs as a package of reform, and a short synthesized description of results by input. We disaggregate findings by the set of studies most likely to credibly support causal conclusions versus observational studies. We shade table cells to illustrate the broad patterns that emerge (i.e., positive relationships in green, null or mixed findings in yellow, negative relationships in red).

Our analysis reveals that quantitative scholars have studied inputs in all three of the major categories identified in our review of the theoretical literature, but with varying degrees of coverage: 67 studies focused on one of the 11 policy inputs under the governance and politics umbrella; 35 studies across 11 policy inputs related to instructional resources and development-related inputs, often the focus of organizational behavior and educational leadership scholars; and 25 studies focused on the four recruitment and retention-related inputs, often studied by economists. This pattern reflects, in part, the number of policies examined by a given discipline, with two to three times the number of inputs in the former two categories than the latter. Another factor for differences in coverage may be related to methodological and publication norms between disciplines. For example, the policy inputs under the recruitment and retention category generally studied by economists all were examined by at

least one study likely to support causal conclusions, while observational studies were more common in both the governance and politics and instructional resources and development categories. A high bar for causal inference may have resulted in lower coverage of these issues.

Policies Related to Governance and Politics. In Table 2a, we summarize findings from the studies examining governance and politics-related inputs, typically emphasized in the theoretical literature by political science-oriented scholars. In total, our theoretical review revealed 11 policy inputs under this category, all of which had coverage in the empirical literature. The most common input studied under this umbrella was superintendent characteristics (33 studies), with union strength (14 studies), district leaders appointed versus elected (10 studies), school autonomy (nine studies), school board characteristics (seven studies), and portfolio management models (seven studies) also having decent coverage. Among the most highly studied categories, findings were mixed and research designs generally were not able to support causal conclusions.

For example, we coded all 33 studies of superintendent characteristics as observational in nature. While the literature suggests a positive association between student achievement and superintendent longevity, higher expectations, and humility, there is no relationship with spending on district administration or measures of superintendent accountability approach. Similarly, all studies of school board characteristics rely on observational methods, and most find no relationship between various measures of board effectiveness or practices and student achievement, with only a couple of exceptions (e.g., involvement of school staff and parents, board member social capital, higher ratings of board meeting effectiveness; Saatcioglu et al., 2011; Ford & Ihrke, 2016a).

When it comes to unions, the district-level studies provide mixed results on the relationship between unionization, union strength, or collective bargaining agreement restrictiveness and student achievement. Among studies likely to support causal conclusions—all of which use difference-in-differences designs (e.g., Hoxby, 1996; Strunk, 2011)—there is some evidence that unions are

associated with lower achievement. In a more comprehensive review of all literature on union effects—which includes all of the studies in our review and others that we exclude because they exploit within-district variation—Cowen and Strunk (2015) also find suggestive evidence of an insignificant or modestly negative relationship between union strength and student outcomes.

Appointment processes of district leaders (appointed versus elected), school autonomy, and portfolio management models also had moderate coverage in the literature, including some studies likely to support causal conclusions. But, for all three policy inputs, we find mixed conclusions when looking across studies. Allen's (2013) use of a regression discontinuity design to estimate the effect of school autonomy arguably provides the strongest study for drawing causal inferences on this input. This study finds little evidence that providing schools with autonomy is better than maintaining centralization for student performance. Most of the other causal studies in these three policy input categories use difference-in-difference designs to examine the effect of a package of district-wide reforms (e.g., Harris & Larsen, 2016; Henry et al., 2019). As such, the unique effect of each policy input cannot be teased out. One observational study on mayoral control of school systems (Wong et al., 2007) finds a positive relationship between achievement and a mayor's ability to appoint the majority of the school board, but a negative relationship with a mayor's ability to appoint all members.

Five studies examine the presence of district administrators, with three of these focused on district turnaround reforms that involved reducing central office spending and/or the number of district administrators. These reforms are associated with increased student achievement (Harris & Larsen, 2016; Schueler et al., 2017) and teacher turnover (Lincove et al., 2018). We shade these patterns in red to indicate that a greater number of district administrators generally is associated with *negative* outcomes (with the exception of the teacher turnover finding).

Our sample also includes five studies of school choice systems, where findings diverge across research methods employed. Lavy (2010) used a regression discontinuity design, finding positive

effects of choice on student achievement and retention in school. Two additional studies used difference-in-differences designs, finding positive effects on student test scores, but where school choice was part of a package of reforms (Chin et al., 2019; Harris & Larsen, 2016). In an observational study, Lincove et al. (2018) found a negative relationship between turnaround efforts in New Orleans, which included a choice component, and teacher retention.

Our theoretical review identified three additional policy inputs under the governance and politics umbrella: school size, selective admissions, and school closures. There is suggestive evidence of positive relationships between all three policy inputs and student outcomes. However, for each, we found only one study that met our inclusion criteria. For school size (De Haan et al., 2016) and selective admissions (Gray et al., 1984), the studies are observational and unlikely to support causal conclusions; for the latter, the study examined school closures in the context of broader set of turnaround reforms (Chin et al., 2019).

Policies Related to Retention and Recruitment of Teacher Talent. In Table 2b, we summarize results from quantitative studies examining inputs related to recruitment and retention of teachers, typically emphasized by economists. The most common input under this umbrella is teacher salaries and benefits (20 studies) with the other categories having less coverage: teacher hiring practices (seven studies), teacher seniority protections (three studies), and teacher evaluation (three studies).

The literature on teacher salaries and benefits (at least that capitalizing on district-level variation) is remarkably consistent, showing a positive relationship between higher salaries and both student outcomes and teacher retention. This pattern is evident both within the set of studies more likely to support causal conclusions (all of which use difference-in-differences design; e.g., Biasi, 2018; Figlio, 2002; Hendricks, 2014, 2015), as well as the observational studies. We include in this input category flexible pay schemes (in addition to variation in base salary), where we also find a positive

relationship to student achievement. However, in many cases, flexible pay schemes were part of broader district turnaround efforts, so the unique effect of this input cannot be isolated.

In turnaround settings, district-level teacher hiring practices—namely, widespread replacement of teachers—also seem to be associated with higher student achievement. Other hiring practices are an understudied area, with just a pair of existing studies (Balter & Duncombe, 2008; Naper, 2010) suggesting that a greater number of recruitment practices (e.g., public advertising of teaching positions, active recruitment from local colleges) and decentralized hiring may be beneficial. Both studies used observational research designs. There is no evidence of a relationship between seniority protections and measures of teacher quality.

Finally, our search did not uncover any studies that isolated the unique effect of various features of teacher evaluation systems at the district level, though a few studies examined district-wide reforms that included changes to teacher evaluation as part of the reform bundle (Chin et al., 2019; Pham et al., 2018; Schueler et al., 2017). In all three cases, studies used difference-in-difference designs to examine the effect of a reform package that included but was not limited to changes in evaluation processes. While all three studies find positive effects, we cannot state with certainty that the teacher evaluation component was the key driver.

Policies Related to Instructional Resources and Development. Table 2c displays findings for the set of studies that examine inputs under the instructional resources and development umbrella, typically emphasized in the theoretical frameworks of organizational behavior and educational leadership scholars. Among the 11 policy inputs in this category, the most common is data use (11 studies), followed by student-teacher ratios (a proxy for class size; nine studies), and curriculum materials (eight studies). Relative to the literature on governance and politics-related inputs, patterns of findings from these studies are more positive, and there is decent coverage of studies that are likely to support causal conclusions.

The only experimental studies in our sample examine inputs that fall under instructional resources and development, all focusing on the effect of district-wide use of data (on student or teacher performance). Two studies examine effects of the same data-use intervention, assessing student outcomes at different points in time (Carlson et al., 2011; Slavin et al., 2013). These and a third experimental study of a different data-use intervention (May & Robinson, 2007) all find positive effects on student achievement. Several additional studies use difference-in-differences designs, also finding positive effects of district's use of data to guide instructional reform (e.g., Strunk et al., 2014). Evidence from observational studies on this topic is mixed. The data-use interventions examined in our sample often included a principal professional development component (e.g., Carlson et al., 2011; Slavin et al., 2013; Strunk et al., 2014), and so we also infer positive effects of this policy input.

Additionally, teacher professional development and curriculum materials are thought to be critical resources to support high-quality instruction and, in turn, student outcomes. The empirical literature on teacher development is broad, though most studies included in prior reviews (e.g., Kraft et al., 2018; Yoon et al., 2007) do not meet our inclusion criteria because they focus on within-district variation. Like the broader literature, studies in our review suggest mixed patterns, particularly those likely to support causal conclusions. Studies that examine new curriculum adoption in the context of other reforms and use difference-in-differences designs to estimate effects find positive relationships to student achievement (e.g., Gandhi et al., 2018; Phenix et al., 2005). Observational studies that focused more narrowly on individual curricula are mixed, particularly in the most recent analysis covering the largest set of textbooks (Blazar et al., 2020).

We separate out studies that look at the coherence of instructional resources, though note that coherence generally refers to the alignment of additional inputs also discussed in our review: teacher professional development and curriculum (as well as instructional standards, which we do not include in our review given that standards are set at the state level). One study in our review assessed coherence

in the context of a package of reforms, where establishing coherence was a top priority (Phenix et al., 2005). Another study directly measured the degree of instructional coherence across school districts in Canada (Leithwood & Azah, 2017). Both find positive associations with student achievement. Notably, several studies under the teacher development and curriculum categories (Chin et al., 2019; Gandhi et al., 2018; Gill et al., 2007) find positive effects of these inputs alongside a package of reforms and using difference-in-differences. This may be further evidence of the benefits of coherence. However, we did not code these other studies as focusing on coherence, as these studies did not call out this policy input specifically, and instructional coherence is described in the theoretical literature as a purposeful endeavor (Elmore, 1996; Newmann et al., 2001).

Several policy inputs consider the presence of school-based staff, including teachers, counselors, and other school support staff. The literature on student-teacher ratios consistently suggests that smaller classes are associated with higher student achievement, though most studies are observational. Studies on the presence of school support staff provide mixed results. Using a difference-in-differences design, Gandhi et al. (2018) find positive effects of wraparound services focused on family engagement and community partnerships, which necessarily require hiring support staff to lead these programs. However, observational studies that link the number of support staff (e.g., counselors, psychologists, aides) to student outcomes come to mixed conclusions.

Five studies examine the value of extra learning time using district-level variation. Sims (2008) is the only study to assess extra learning time on its own and likely to support causal conclusions, finding that increasing class time prior to testing results in improved student test score performance. Several district turnaround evaluations study reforms that included increased learning time, finding positive effects of the package of reforms. One observational study (Hinrichs, 2011) found no association between earlier school start times and student achievement (possibly because starting the school day earlier may not increase instructional time).

Three studies examine the effect of improved school buildings and facilities using similar regression discontinuity designs that exploit district-wide votes on capital campaigns. Though this research design is likely to support causal conclusions, findings are mixed across studies. One study (Kai & Zimmer, 2016) finds positive effects on student proficiency rates, while two others find small or null effects on student test-score performance (Cellini et al., 2010; Martorell et al., 2016). Two studies find positive relationships between district-wide purchasing of technology and student outcomes, one in the context of a package of reforms (Chin et al., 2019) and the other that examines technology purchases on their own in the context of a policy change in England (Machin et al., 2007). Finally, one study on school-grade configuration finds negative effects of attending middle school in sixth grade, rather than staying in elementary school (Cook et al., 2008).

Discussion

Synthesizing Findings Across Disciplines and Policy Inputs

Consistent with our review of the theoretical literature, our systematic, interdisciplinary review of the quantitative empirical research reveals varied ways in which school districts are likely to influence student outcomes. Analyses of teacher human resource policies—often the focus of economists of education—identify that higher teacher salaries and strategic hiring and replacement of teachers consistently are associated with higher teacher retention and improved student test scores. Instructional resources including lower student-teacher ratios and data use—generally district levers that have been the focus of scholars from organizational behavior and educational leadership—also are associated with improvements on these outcomes. Though findings are less consistent from the governance and politics literature, some studies suggest that school choice is related to measures of district effectiveness. Thus, our review contributes to the understanding of school districts as key agents in the educational change process and suggests that, to better inform decision making,

researchers must be attuned to the varied routes by which district leaders can impact teacher quality and student outcomes beyond those favored by a single academic discipline.

Our review also provides suggestive evidence that district-level policymaking, in some instances, may be more efficient than policymaking at the school or classroom level. To illustrate, we describe results from studies focused on data use, which increasingly has been used as a strategy by policy actors at different school system levels: states design assessments that generate data; districts collect data and report information to the state, as well as to schools, teachers, and students; and schools work with teachers to interpret data to drive instruction. The data-use literature also is worthy of closer investigation, as it is the only district-level input examined using an experimental design. Analyzing data from the same experiment but on different outcomes and time points, Carlson et al. (2011) and Slavin et al. (2013) found positive effects of district-wide support to district leaders to interpret and use data to guide reform and resource allocation on student test scores. In the May and Robinson (2007) experiment of a similar data-use intervention, average test scores did not improve. But, aligned to the theory of action, districts improved the rates by which underperforming students retook end-of-year assessments, and increased scores on these retakes.

These findings contrast with a broader literature base that has randomly assigned data-use supports and interventions to schools or teachers. Some studies find positive effects on student outcomes (i.e., Betts et al., 2017; Faber et al., 2017; Supovitz et al., 2018). However, one study identifies a statistically significant negative effect on student outcomes of a school-level data-use intervention (Konstantopoulos et al., 2016), and the majority of studies examining similar school-level interventions find null results (e.g., Gleason et al., 2019; Randel et al., 2016; van der Scheer & Visscher, 2018; West et al., 2016). The literature on data use is not yet sufficiently large to determine definitively whether these interventions are more beneficial when implemented at the district versus school or classroom level. However, results across studies hint at such a pattern.

What is Missing from the Quantitative Literature?

At the same time that our review identifies several features of effective district-level policymaking, our mapping of the theoretical frameworks against the quantitative empirical literature reveals gaps that limit the ability for researchers to provide a clear roadmap for district leaders.

Causal Evaluations. First, only a small subset of quantitative studies (21%) employed causal research designs, with a disproportionate number applied to studying recruitment and retention-related policies most often prioritized for evaluation by economists. Roughly half of these studies used difference-in-differences designs to examine a package of reforms that included between two and 12 different policy inputs. Thus, it is difficult to isolate the effect of each. Further, while we exclude purely correlational studies, it is not clear that the control variables (e.g., district size, demographic characteristics) used in multivariate regression models of the remaining observational studies (79%) sufficiently minimize threats to internal validity. Studies without an experimental or natural-experimental design make it impossible to determine the direction of any causal arrow. For example, does peer collaboration among teachers improve student outcomes, or are teachers simply more interested in collaborating with colleagues in higher-performing districts?

The consequences of research design quality can be seen for several policy topics where findings differ between observational studies and those more likely to support causal conclusions. Returning to the data use example, randomized control trials show positive effects (Carlson et al., 2011; May & Robinson, 2007; Slavin et al., 2013), while findings from observational studies are more mixed (e.g., Leithwood, 2017; Lee et al., 2012). For school choice, we observe positive relationships with student achievement for causal studies (Chin et al., 2019; Harris & Larsen, 2016; Lavy, 2010), but some evidence from observational studies that school choice is associated with higher teacher turnover (Lincove et al., 2018). Mixed patterns, potentially related to research design, also are evident amongst

studies examining union strength, portfolio management model, curriculum materials, professional development, extra learning time, support staff, and district administrators.

Under-studied Policy Inputs. A second gap in the literature relates to under-studied inputs. While all of the district-level policy categories that surfaced from our theoretical review had some coverage in the empirical, quantitative literature, in many instances there were only a handful of studies and often only observational ones. Examples include teacher evaluation systems, where all three studies examine evaluation alongside a reform package; school size, selective admissions, school closures, technology, and school-grade configurations all had just one or two studies. While additional published studies on these topics exist, our review revealed that they generally focus on variation within rather than across districts that are unlikely to capture general equilibrium effects. These policies are therefore ripe for future analysis.

We see a particular need for additional research related to instructional program coherence. The theoretical literature on effective districts consistently calls for coherent instructional systems that craft alignment between curriculum materials, professional development, and student assessment (Elmore, 1996; Honig & Hatch, 2004; Newmann et al., 2001). Over several decades, the qualitative case study literature has provided illustrative examples of districts thought to have succeeded in crafting coherence (i.e., Boston, Chicago, District #2 in New York City; Bryk, 2010; Elmore & Burney, 1998). More recently, survey measures have been developed to quantitatively capture variation in instructional program coherence, though analyses generally examine variation across schools (e.g., Newmann et al., 2001). To our knowledge, only one study (Leithwood & Azah, 2017) has used such a survey to quantitatively predict variation in performance across districts. A second study (Phenix et al., 2005) looks at coherence as part of a reform package that includes other inputs (e.g., school autonomy), making it difficult to isolate coherence and leaving room for future research.

A related gap in the literature is a lack of attention to testing the relative efficacy or complementarity of policy inputs across the dominant theories of each discipline. For example, do recruitment and retention practices matter more or less than governance features or instructional supports? Are the benefits of recruitment strategies or instructional supports moderated by unique governance structures (e.g., increased autonomy to schools)? Our study was designed to bring together distinct disciplinary perspectives. As researchers reviewing the literature rather than working with the raw data, we can compare patterns of findings across policy categories but cannot directly estimate mediating and moderating relationships. Understanding mediating and moderating relationships is a challenge in causal research (Murnane & Willett, 2010), and so we recognize that filling this gap alongside a call for causal studies will be difficult. It is unlikely that any single study will be able to examine the multifaceted ways in which varied district inputs are linked to student outcomes *and* draw causal conclusions. That said, the tension between methodological rigor and interconnectedness of policy inputs is too important to leave aside. We provide some guidance below.

Limited Set of Outcomes. A third gap relates to the outcome measures examined. The vast majority of studies in our review relate district-level policy inputs to student test scores, which is important but potentially limited. In addition to improving test scores, districts are thought to play an important role in supporting students' likelihood of graduating from high school, enrolling in college, and gaining skills to support career readiness. In fact, districts that succeed in improving test scores may do so in a way that has negative consequences on other outcomes (Hibel & Penn, 2020). Scholars looking to detect features of effective districts should triangulate estimates against multiple outcomes.

Further, we observed limited attention to teacher-level outcomes, even though all three disciplinary perspectives describe theoretical frameworks in which district-level policies impact students only insomuch as they first improve teaching quality. Where district-level policies do not benefit student outcomes, it is important to understand where and how the theory of action broke

down. Outcome measures at the teacher level are one way to do so. Limited analyses linking district-level inputs to measures of teacher quality and, in turn, student outcomes likely is a consequence of the data generally available to researchers. When exploiting variation across districts, it generally is necessary to rely on existing state or national data sources; primary collection of data across a large number of districts is quite burdensome. While state data systems tend to collect a broad set of student-level outcome measures (e.g., test scores, absences, suspensions, grade progression, graduation), large-scale data on teacher quality across school districts can be hard to come by. However, there has been some progress in recent years, which we turn to below in making recommendations for future research.

Guidance for Future Research

We see multiple ways for future research projects to fill in gaps that we describe in the previous section, and we provide two concrete recommendations. Starting first with a need for a broader set of outcome measures, the scale-up of statewide data systems for accountability purposes and increased push to look beyond test scores (ESSA 2015) provides researchers with several measures that may better capture district performance than looking at test scores alone. Often available measures include: absences, suspensions, on-time grade progression, high school graduation, college-going, and even survey-based measures of social-emotional constructs.

To expand the literature linking district-level policies to teacher quality, one approach would be to use value-added measures that consider the impact of teachers on student outcomes. These teacher-level measures also can be estimated with existing administrative data and generally are thought to better capture the underlying construct of interest (i.e., teacher quality) than proxy measures also available in administrative data (e.g., certification, education, licensure scores; Kane et al., 2013). The scale-up of teacher observation and evaluation systems also provides a new source of data that could be used for future research studies. Until recently, data on classroom practices generally came

only from primary data collection efforts that are quite expensive and tend to cover only small to moderate samples (e.g., Hill et al., 2015). Teacher evaluation systems that are now commonplace in school districts have established a possible new source of administrative data on how teachers engage with students in classrooms. While teacher evaluation and observation processes often are set at the district level, states increasingly are providing oversight and guidance including recommending use of a subset of observation instruments (Reform Support Network, n.d.), suggesting that these measures may be comparable across districts.

We also provide guidance on how researchers may use and analyze these data in a way that can both support more robust causal conclusions and examine the interconnectedness of policy inputs. As described above, we recognize that randomly assigning policy inputs at the district level is challenging if not impossible: such studies would require a very large sample of districts to ensure sufficient statistical power, would require buy-in from a very large number of stakeholders with varied goals and interests, and each experiment would likely focus on a narrow set of input(s). We therefore do not expect that a push for district-level randomization will represent the primary mechanism for advancing the study of effective district-level policymaking.

In lieu of random assignment, researchers should look for opportunities to analyze data using natural-experimental designs. Regression discontinuity designs exploiting thresholds in district-wide votes represent a promising approach. Difference-in-differences designs may be the best way to study district-level policymaking at scale. Our review shows how this design already has been used to study the effect of individual policy inputs, as well as district turnarounds with multiple policy inputs. Above, we note that studies examining a package of reforms make it difficult to tease out the unique effect of each. At the same time, as the turnaround literature continues to grow, patterns are likely to emerge around which package is more versus less effective. In fact, the district turnaround literature may be most closely aligned with the notion of district-level policymaking as a multi-pronged endeavor,

requiring policies related to the recruitment and retention of teacher talent, governance, and allocation of instructional resources. Variations on the standard difference-in-differences design may also allow researchers to model district-level outcomes as a function of multiple policy changes, over time and across districts (see, for example, Johnson & Jackson, 2019). By including multiple policy changes in the same model and analysis, one could also begin to examine mediating and moderating pathways. The Stanford Education Data Archive (SEDA) provides one useful source of data allowing for cross-state comparisons over time at the district level (Reardon et al., 2021).

Conclusion

The extant theoretical and empirical literature provides strong support for the idea that school districts are a critical policy agent for improving student outcomes. Our review of the quantitative evidence on district effectiveness provides some insight into the policy inputs and leverage points that are associated with increased district-wide performance (e.g., higher teacher salary, strategic teacher hiring practices, lower student-teacher ratios, data use) and, in particular, the fact that such differences stem from three broad areas of policymaking related to recruitment and retention of teacher talent, instructional resources and development, and governance and politics. At the same time, much more causal research—looking across disciplinary perspectives—is needed to inform district-level decision-making and to provide a roadmap for leaders working to advance educational excellence and equity in the organizational unit with the primary authority for developing education policy.

References

* = studies that are included in the formal research synthesis

*Adams, J. P. (1987). *Superintendents and effective schools.* [Unpublished doctoral dissertation]. University of California, Santa Barbara.

*Alexander, K. L., & Griffin, L. J. (1976). School district effects on academic achievement: A reconsideration. *American Sociological Review*, *41*(1), 144-152.

*Allen, R. (2013). Measuring foundation school effectiveness using English administrative data, survey data and a regression discontinuity design. *Education Economics*, *21*(5), 431-446.

Anderson, E., & Young, M. D. (2018, October). If they knew then what we know now, why haven't things changed? An examination of district effectiveness research. In *Frontiers in Education* (Vol. 3, p. 87). Frontiers Media SA.

*Balter, D., & Duncombe, W. D. (2008). Recruiting highly qualified teachers: Do district recruitment practices matter?. *Public Finance Review*, *36*(1), 33-62.

*Berlau, D. C. (2011). *Superintendent longevity and its relationship to student performance* [Unpublished doctoral dissertation]. Drake University.

Betts, J. R., Hahn, Y., & Zau, A. C. (2017). Can testing improve student learning? An evaluation of the mathematics diagnostic testing project. *Journal of Urban Economics*, 100, 54-64.

*Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness. *Educational Evaluation and Policy Analysis*, *34*(4), 391-412.

*Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review*, *34*, 107-121.

*Biasi, B. (2018). *The labor market for teachers under different pay schemes* (No. w24813). National Bureau of Economic Research.

*Bidwell, C. E., & Kasarda, J. D. (1975). School district organization and student achievement. *American Sociological Review*, 55-70.

*Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S., Goldhaber, D., Harris, D. N., Hitch, R., Holden, K., & Kurlaender, M. (2020). Curriculum reform in the Common Core era. *Journal of Policy Analysis and Management, 39*(4), 966-1019.

Blazar, D., Litke, E., & Barmore, J. (2016). What does it mean to be ranked a "high" or "low" value-added teacher? *American Educational Research Journal, 53*(2), 324-359.

*Booker, K., & Glazerman, S. (2009). *Does the Missouri teacher career ladder program raise student achievement?*. Mathematica Policy Research, Inc.

Bowers, A. J. (2010). Toward addressing the issues of site selection in district effectiveness research: A two-level hierarchical linear growth model. *Educational Administration Quarterly*, *46*(3), 395-425.

Bowers, A. J., Ni, X., & Esswein, J. (2018). Using hierarchical growth modeling to promote district systematic improvement in Ohio and Texas. *Leading holistically: How schools, districts, and states improve systemically*, 77-100.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2004). The draw of home. *Journal of Policy Analysis and Management*, 24(1), 113—132.

*Brewer, D. J. (1996). Does more school district administration lower educational productivity? *Economics of Education Review*, *15*(2), 111-124.

Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan*, *91*(7), 23-30.

Bulkley, K., Henig, J., & Levin, H. (2010). *Between public and private: Politics, governance, and the new portfolio models for urban school reform*. Harvard Education Press.

*Burnett, R. D. (1990). *The effects of superintendents' leadership behaviors in curriculum and instruction upon student achievement in South Carolina public school districts.* [Unpublished doctoral dissertation]. University of South Carolina.
*Byrd, J. K. (2002). *Effective superintendent leadership strategies and management techniques for improving student performance as perceived by superintendents in selected school districts in Texas.* [Unpublished doctoral dissertation]. Texas A&M University.
*Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378-398.
*Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments. *The Quarterly Journal of Economics*, *125*(1), 215-261.
Childress, S., Doyle, D., & Thomas, D. (2009). *Leading for equity*. Harvard Education Press.
*Chin, M., Kane, T, Kozakowski, W., Schueler, B., & Staiger, D. (2019). School district reform in Newark. *ILR Review, 72*(2), 323-354.
Chingos, M. M., Whitehurst, G. J., & Gallaher, M. R. (2015). School districts and student achievement. *Education Finance and Policy, 10*(3), 378-398.
Chubb, J. & Moe, T. (1990). *Politics, markets and America's schools*. Brookings Institution Press.
*Clore, W. P. (1992). *The relationship of superintendent instructional leadership behavior and school district demographics to student achievement.* [Unpublished doctoral dissertation]. University of Texas.
Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*(4), 778-820.
Clotfelter, C.T., Ladd, H.F., & Jacob L. Vigdor, (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy, 6*, 399-438.
*Cohen-Vogel, L., Feng, L., & Osborne-Lampkin, L. T. (2013). Seniority provisions in collective bargaining agreements and the "Teacher Quality Gap." *Educational Evaluation and Policy Analysis*, *35*(3), 324-343.
Cohen, D. K., & Hill, H. C. (2008). *Learning policy*. Yale University Press.
Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A. & Weinfeld, F. (1966). *Equality of educational opportunity study.* U.S. Department of Health, Education, and Welfare.
*Cook, P. J., MacCoun, R., Muschkin, C., & Vigdor, J. (2008). The negative impacts of starting middle school in sixth grade. *Journal of Policy Analysis and Management*, *27*(1), 104-121.
Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.
*Cotter, M. E. (2002). *Strategic leadership for student achievement.* [Unpublished doctoral dissertation]. Johnson and Wales.
Cowen, J. M., & Strunk, K. O. (2015). The impact of teachers' unions on educational outcomes: What we know and what we need to learn. *Economics of Education Review*, *48*, 208-223.
*De Haan, M., Leuven, E., & Oosterbeek, H. (2016). School consolidation and student achievement. *The Journal of Law, Economics, and Organization*, *32*(4), 816-839.
Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267-297.
DiMaggio & Powell. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*(2), 147—160.
*Duvall, S. (2005). *Superintendent evaluation and other influences on the school board and superintendent relationship.* [Unpublished dissertation]. Eastern Michigan University.
*Eberts, R. W., & Stone, J. A. (1984). *Unions and Public Schools: The Effect of Collective Bargaining on American Education*. Lexington, MA: Lexington Books.

*Eberts, R. W., & Stone, J. A. (1987). Teacher unions and the productivity of public schools. *ILR Review*, *40*(3), 354-363.
*Edwards, C. (2007). *An analysis of the relationship of superintendent instructional leadership behaviors and district performance outcomes.* [Unpublished dissertation]. Tarleton State.
Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1-27.
Elmore, R., & Burney, D. (l998). *Continuous improvement in Community District #2, New York City.*
*Endeman, J. L. (1992). *Visionary leadership in superintendents and its effect on organizational outcomes.* [Unpublished doctoral dissertation]. University of La Verne.
*Engel, A. (2016). *Superintendent characteristics and their relationship to student achievement in Reading and mathematics.* [Unpublished dissertation]. University of Wisconsin.
Faber, J. M., Luyten, H., & Visscher, A. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation. *Computers & Education*, 106, 83-96.
Fahle, E. M., & Reardon, S. F. (2018). How much do test scores vary among school districts? New estimates using population data, 2009–2015. *Educational Researcher*, *47*(4), 221-234.
*Figlio, D. (2002). Can public schools buy better-qualified teachers? *ILR Review*, *55*(4), 686-699.
Firestone, W. A., Mangin, M. M., Martinez, M. C., & Polovsky, T. (2005). Leading coherent professional development. *Educational Administration Quarterly*, *41*(3), 413-448.
*Ford, M. (2013). *The impact of school board governance on academic achievement in diverse states.* [Unpublished doctoral dissertation]. University of Wisconsin, Milwaukee.
*Ford, M. R., & Ihrke, D. M. (2016a). Do school board governance best practices improve district performance? *International Journal of Public Administration*, *39*(2), 87-94.
*Ford, M. R., & Ihrke, D. M. (2016b). Differences in school boards with hired and elected superintendents. *International Review of Public Administration*, *21*(4), 292-304.
*Gandhi, A. G., Slama, R., Park, S. J., Russo, P., Winner, K., Bzura, R. & Williamson, S. (2018). Focusing on the whole student: An evaluation of Massachusetts's wraparound zone initiative. *Journal of Research on Educational Effectiveness*, *11*(2), 240-266.
*Gibbons, S., Machin, S., & Silva, O. (2008). Choice, competition, and pupil achievement. *Journal of the European Economic Association*, *6*(4), 912-947.
*Gill, B., Zimmer, R., Christman, J., & Blanc, S. (2007). State takeover, school restructuring, private management, and student achievement in Philadelphia. *RAND Corporation*.
Glazerman, S., & Max, J. (2011). *Do low-income students have equal access to the highest-performing teachers?* National Center for Education Evaluation and Regional Assistance.
Gleason, P., Crissey, S., Chojnacki, G., Zukiewicz, M., Silva, T., Costelloe, S., & O'Reilly, F. (2019). *Evaluation of support for using student data to inform teachers' instruction*. NCEE 2019-4008. National Center for Education Evaluation and Regional Assistance.
*Goldhaber, D., Lavery, L., & Theobald, R. (2016). Inconvenient truth? Do collective bargaining agreements help explain the mobility of teachers within school districts? *Journal of Policy Analysis and Management*, *35*(4), 848-880.
Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal*, *55*(1), 171-201.
*Goodman, G. S., & Young, I. P. (2006). The value of extracurricular support in increased student achievement. *Educational Research Quarterly*, *30*(1), 3-13.
*Gray, J., Jesson, D., & Jones, B. (1984). Predicting differences in examination results between local education authorities. *Oxford Review of Education*, *10*(1), 45-68.
*Grissom, J. A. (2014). Is discord detrimental? *Journal of Public Administration Research and Theory*, *24*(2), 289-315.

Grissom, J. A., & Strunk, K. O. (2012). How should school districts shape teacher salary schedules? *Educational Policy*, *26*(5), 663-695.

Halloran, C., Jack, R., Okun, J. C., & Oster, E. (2021). *Pandemic schooling mode and student test scores: Evidence from us states* (No. w29497). National Bureau of Economic Research.

*Hanks, J. M. (2010). *The influence of the superintendent of schools on student academic performance*. Texas A&M University.

Hanushek, E. (2010), Education production functions. In P. Peterson, E. Baker, B. McGaw, (Editors), *International Encyclopedia of Education*, *2,* 407-411. Oxford: Elsevier.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources, 39*(2), 326—354.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267-71.

*Harris, D. N., & Larsen, M. F. (2016). *The effects of the New Orleans post-Katrina market-based school reforms on student achievement, high school graduation, and college outcomes*. Education Research Alliance for New Orleans.

*Hart, A. W. (1983). *An exploration of the effects of superintendents on the instructional performance of school districts.* [Unpublished doctoral dissertation]. University of Utah.

*Hart, A. W., & Ogawa, R. T. (1987). The influence of superintendents on the academic achievement of school districts. *Journal of Educational Administration, 25*(1), 72-84.

*Hart, W., Rock Hill, S., Schramm-Possinger, M., & Hoyle, S. (2019). Superintendent Longevity and Student Achievement in North Carolina Public Schools. *Editorial Review Board*.

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two-and three-level cluster-randomized experiments in education. *Evaluation Review*, *37*(6), 445-489.

*Henderson, D. J., Simar, L., & Wang, L. (2017). The three I s of public schools: irrelevant inputs, insufficient resources and inefficiency. *Applied Economics*, *49*(12), 1164-1184.

*Hendricks, Matthew D. Does it pay to pay teachers more? Evidence from Texas. *Journal of Public Economics* 109 (2014): 50-63.

*Hendricks, M. D. (2015). Towards an optimal teacher salary schedule: Designing base salary to attract and retain effective teachers. *Economics of Education Review*, *47*, 143-167.

Henig, J. (2009). Mayors, governors, and presidents: The new education executives and the end of educational exceptionalism. *Peabody Journal of Education*, 84(3).

Henig, J. R. (2013). *The end of exceptionalism in American education: The changing politics of school reform*. Harvard Education Press.

*Henry, G. T., Pham, L. D., Kho, A., & Zimmer, R. (2020). Peeking into the black box of school turnaround. *Educational Evaluation and Policy Analysis*, *42*(2), 232—256.

Hibel, J., & Penn, D. (2020). Bad apples or bad orchards? An organizational analysis of educator cheating on standardized accountability tests. *Sociology of Education, 93*(4), 331-352.

Hightower, A. M. (2002). *San Diego's big boom: District bureaucracy supports culture of learning*.

Hill, H. C., Blazar, D., & Lynch, K. (2015). Resources for teaching: Examining personal and institutional predictors of high-quality instruction. *AERA Open*, *1*(4).

*Hinrichs, P. (2011). When the bell tolls: The effects of school starting times on academic achievement. *Education Finance and Policy*, *6*(4), 486-507.

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, *37*(6), 351-360.

*Hofman, R. H. (1995). Contextual influences on school effectiveness: The role of school boards. *School Effectiveness and School Improvement*, *6*(4), 308-331.

Honig, M. (2012). District central office leadership as teaching. *Educational Administration Quarterly*, *48*(4), 733-774.
Honig, M., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy*, *22*(4), 578-608.
Honig, M. & Hatch, T. (2004). Crafting coherence. *Educational Researcher*, 33(8), 16-30.
*Hoover, G. A. (2008). Elected versus appointed school district officials: Is there a difference in student outcomes? *Public Finance Review*, 36(5), 635-647.
*Houck, E. A., Rolle, R. A., & He, J. (2010). Examining school district efficiency in Georgia. *Journal of Education Finance*, 331-357.
*Hough, K. L. (2014). Internal accountability and district achievement: How superintendents affect student learning. *Journal of School Leadership*, *24*(1), 32-60.
*Hoxby, C. M. (1996). How teachers' unions affect education production. *The Quarterly Journal of Economics*, *111*(3), 671-718.
*Hoyle, J. R., Ealy, C., Skrla, L. E., & Hogan, D. (2001). Superintendent performance evaluation and its relationship to district student performance. *NCPEA Yearbook 2001*.
Jackson, C. K. (2018a). *Does school spending matter? The new literature on an old question* (No. w25368). National Bureau of Economic Research.
Jackson, C. K. (2018b). What do test scores miss? The importance of teacher effects on non—test score outcomes. *Journal of Political Economy*, *126*(5), 2072-2107.
Jacobsen, R., & Saultz, A. (2012). The polls—Trends: Who should control education? *Public Opinion Quarterly*, *76*(2), 379-390.
*Jacques, C., & Brorsen, B. W. (2002). Relationship between types of school district expenditures and student performance. *Applied Economics Letters*, *9*(15), 997-1002.
*Johnson, K. (1997). *The relationship of superintendent tenure to school performance in Arkansas* [Unpublished doctoral dissertation]. University of Arkansas, Fayetteville.
Johnson, R. C., & Jackson, C. K. (2019). Reducing inequality through dynamic complementarity. *American Economic Journal: Economic Policy*, *11*(4), 310-49.
Johnson, S. M., Marietta, G., Higgins, M., Mapp, K., & Grossman, A. (2015). *Achieving coherence in district improvement*. Cambridge, MA: Harvard Education Press.
*Kai, H., & Zimmer, R. (2016). Does investing in school capital infrastructure improve student achievement?. *Economics of Education Review*, *53*, 143-158.
Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers?* MET Project. Bill & Melinda Gates Foundation.
*Koedel, C., Li, D., Polikoff, M. S., Hardaway, T., & Wrabel, S. L. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open*, *3*(1).
*Koski, W. S., & Horng, E. L. (2007). Facilitating the teacher quality gap? *Education Finance and Policy*, *2*(3), 262-300.
Konstantopoulos, S., Miller, S., van der Ploeg, A., & Li, W. (2016). Effects of interim assessments on student achievement. *Journal of Research on Educational Effectiveness, 9*(1), 188-208.
Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement. *Review of Educational Research*, *88*(4), 547-588.
Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37—62.
*Lavy, V. (2010). Effects of free choice among public schools. *The Review of Economic Studies*, *77*(3), 1164-1191.
*Lavy, V., & Boiko, A. (2017). *Management quality in public education: Superintendent value-added, student outcomes and mechanisms* (No. w24028). NBER.

*Lee, M., Seashore Louis, K., & Anderson, S. (2012). Local education authorities and student learning. *School Effectiveness and School Improvement*, *23*(2), 133-158.
Leithwood, K., (2010). A review of evidence about the characteristics of high performing school districts. *Leadership and Policy in Schools, 9*(3), 245 — 291.
*Leithwood, K., & Azah, V. N. (2017). Characteristics of high-performing school districts. *Leadership and Policy in Schools*, *16*(1), 27-53.
Leithwood, K., Sun, J., & McCullough, C. (2019). How school districts influence student achievement. *Journal of Educational Administration*.
*Libka, R. J. (2012). *The relationship between Illinois school district superintendent longevity and high school student achievement*. Western Illinois University.
*Lin, T. C., & Quayes, S. (2006). The impact of local taxes on public school performance: the case of Pennsylvania. *Applied Economics Letters*, *13*(7), 423-426.
*Lin, T. C. (2010). Teacher salaries and student achievement: The case of Pennsylvania. *Applied Economics Letters*, *17*(6), 547-550.
*Lincove, J. A., Barrett, N., & Strunk, K. O. (2018). Lessons from Hurricane Katrina: The employment effects of the mass dismissal of New Orleans teachers. *Educational Researcher*, *47*(3), 191-203.
Little, J. W. (1989). District policy choices and teacher's professional development opportunities. *Educational Evaluation and Policy Analysis*, *11*(2), 165-179.
*Lovenheim, M. F. (2009). The effect of teachers' unions on education production. *Journal of Labor Economics*, *27*(4), 525-587.
*Machin, S., McNally, S., & Silva, O. (2007). New technology in schools: Is there a payoff?. *The Economic Journal*, *117*(522), 1145-1167.
*Martorell, P., Stange, K., & McFarlin Jr, I. (2016). Investing in schools: capital spending, facility conditions, and student achievement. *Journal of Public Economics*, *140*, 13-29.
*May, H., & Robinson, M. A. (2007). *A randomized evaluation of Ohio's personalized assessment reporting system (PARS)*. Unpublished paper.
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340—363.
McGuinn, P. (2012). *The state of teacher evaluation reform: State education agency capacity and the implementation of new teacher evaluation systems*. Center for American Progress.
*Mensah, Y., Schoderbek, M., & Sahay, S. (2013). The effect of administrative pay and local property taxes on student achievement scores. *Economics of Education Review*, *34*, 1-16
*Moe, T. M. (2009). Collective bargaining and the performance of the public schools. *American Journal of Political Science*, *53*(1), 156-174.
Moe, T. (2011). *Special interest: Teachers unions and America's public schools*. Washington, DC: Brookings Institution Press.
Monk, D. H. (1989). The education production function: Its evolving role in policy analysis. *Educational Evaluation and Policy Analysis*, *11*(1), 31-45.
*Morgan, G. W. (1990). *School district effectiveness and the leadership of the superintendent of schools*. [Unpublished doctoral dissertation]. Rutgers University.
*Muller, R. W. (1990). *Instructional leadership superintendent competencies related to student achievement*. [Unpublished doctoral dissertation]. University of Texas, Austin.
*Murnane, R. J., & Olsen, R. J. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *Journal of Human Resources*, 106-124.
Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
Murphy, J., & Hallinger, P. (1988). Characteristics of instructionally effective school districts. *The Journal of Educational Research*, *81*(3), 175-181.

*Myers, S. (2011). Superintendent length of tenure and student achievement. *Administrative Issues Journal*, *1*(2), 6.
*Naper, L. R. (2010). Teacher hiring practices and educational efficiency. *Economics of Education Review*, *29*(4), 658-668.
Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence. *Educational Evaluation and Policy Analysis*, *23*(4), 297-321.
O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, *72*(3), 293-329.
Opfer, V. D., Henry, G. T., & Mashburn, A. J. (2008). The district effect: Systemic responses to high stakes accountability policies in six southern states. *American Journal of Education*, *114*(2), 299-332.
Orton, J.D. & Weick, K.E. (1990). Loosely coupled systems: A reconceptualization. *Academy of Management Review*, 15, 203-223.
Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market. *Journal of Public Economics*, *130*, 105-119.
*Parker-Chenaille, R. (2012). *The impact of superintendent turnover on student achievement in rural districts.* [Unpublished doctoral dissertation]. St. John Fisher College.
*Partridge, M., & Sass, T. R. (2011). The productivity of elected and appointed officials: the case of school superintendents. *Public Choice*, *149*(1-2), 133.
Peterson, P. E. (2016). James S. Coleman: Education's North Star. *Education Next*, *16*(2), 5-6.
*Pham, Lam, Henry, G. T., Zimmer, R., & Kho, A. (2018). *School turnaround after five years.* Tennessee Education Research Alliance.
*Phenix, Siegel, Zaltsman & Fruchter. (2005). A forced march for failing schools: Lessons from the New York City Chancellor's District. *Education Policy Analysis Archives*, *13*, 1-24.
*Plotts, T. (2011). *A multiple regression analysis of factors concerning superintendent longevity and continuity relative to student achievement.* [Unpublished dissertation]. Seton Hall.
*Plotts, T., & Gutmore, D. (2014). The superintendent's influence on student achievement. *AASA Journal of Scholarship & Practice*, *11*(1), 26.
*Podolsky, A., Darling-Hammond, L., Doss, C., & Reardon, S. (2019). *California's positive outliers: Districts beating the odds.* Learning Policy Institute.
Polikoff, M. (2018). The challenges of curriculum materials as a reform lever. *Economic Studies at Brookings: Evidence Speaks Reports*, *2*(58).
Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *Journal of Educational Research*, 109(5), 491-502.
Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., Jang, H., & Chavez, B. (2021). Stanford Education Data Archive (Version 4.1). Retrieved from http://purl.stanford.edu/db586ns4974.
Reform Support Network. (n.d.). *Evaluations of teacher effectiveness: State requirements for classroom observations.*
*Register, C. A., & Grimes, P. W. (1991). Collective bargaining, teachers, and student achievement. *Journal of Labor Research*, *12*(2), 99-109.
*Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 74-95.
Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Rorrer, A. K., Skrla, L., & Scheurich, J. J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly*, *44*(3), 307-357.
*Rose, H., & Sonstelie, J. (2010). School board politics, school district size, and the bargaining power of teachers' unions. *Journal of Urban Economics*, *67*(3), 438-450.
Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, *3*(2), 135-146.
*Rudolph, B. R. (2016). *The Role of Superintendent Longevity in Predicting 2014-2016 APR Scores for Missouri School Districts*. Baker University.
*Saatcioglu, A., Moore, S., Sargut, G., & Bajaj, A. (2011). The role of school board social capital in district governance. *Leadership and Policy in Schools*, *10*(1), 1-42.
*Schueler, B. E., Goodman, J. S., & Deming, D. J. (2017). Can states take over and turn around school districts? *Educational Evaluation and Policy Analysis*, *39*(2), 311-332.
*Shi, Y., & Singleton, J. D. (2019). *Expertise and independence on governing boards: Evidence from school districts*. IZA Discussion Paper No. 12414.
*Simpson, J. (2013). Superintendent Tenure and Student Achievement. *AASA Journal of Scholarship & Practice*, *9*(4).
*Sims, D. P. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, *27*(1), 58-68.
*Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, *50*(2), 371-396.
*Sojourner, A. J., Mykerezi, E., & West, K. L. (2014). Teacher pay reform and productivity panel data evidence from adoptions of q-comp in Minnesota. *Journal of Human Resources*, *49*(4), 945-981.
Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, *18*(2), 141-179.
Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity. *Educational Evaluation and Policy Analysis*, *19*(2), 185-203.
*Steele, F., Vignoles, A., & Jenkins, A. (2007). The effect of school resources on pupil attainment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(3), 801-824.
Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? *Education Finance and Policy*, *10*(4), 535-572.
*Strunk, K. O. (2011). Are teachers' unions really to blame? *Education Finance and Policy*, *6*(3), 354-398.
*Strunk, K. O., & McEachin, A. (2011). Accountability under constraint. *American Educational Research Journal*, *48*(4), 871-903.
*Strunk, K., McEachin, A., & Westover, T. (2014). The use and efficacy of capacity-building assistance for low-performing districts. *Journal of Policy Analysis and Management*, *33*(3), 719-751.
*Strunk, K. O., & McEachin, A. (2014). More than sanctions: Closing achievement gaps through California's use of intensive technical assistance. *Educational Evaluation and Policy Analysis*, *36*(3), 281-306.
Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Harvard Education Press.
Supovitz, J. A., Ebby, C. B., Remillard, J., & Nathenson, R. A. (2018). Experimental impacts of the ongoing assessment project on teachers and students. *CPRE Research Report 2018-1*.
*Theobald, N. D. (1990). An examination of the influence of personal, professional, and school district characteristics on public school teacher retention. *Economics of Education Review*, *9*(3), 241-250.

*Theobald & Gritz (1996). The effects of school district spending priorities on the exit paths of beginning teachers leaving the district. *Economics of Education Review*, *15*(1), 11-22.

Thompson, C. L., Sykes, G., & Skrla, L. (2008). *Instructionally-focused district leadership: Toward a theoretical account.* The Education Policy Center at Michigan State University. Retrieved from: https://education.msu.edu/epc/forms/Thompson_et_al_2008_Coherent_Instructionally_Focused_District_Leadership.pdf

*Tran, H. (2017). Does the pay stance of South Carolina public school districts influence their math and science achievement scores? *Journal of Education Finance*, *43*(2), 105-122.

Trujillo, T. (2013). The reincarnation of the effective schools research: Rethinking the literature on district effectiveness. *Journal of Educational Administration*, 51(4), 426-452.

van der Scheer, E. A., & Visscher, A. J. (2018). Effects of a data-based decision-making intervention for teachers on students' mathematical achievement. *Journal of Teacher Education, 69*(3), 307-320.

*Vaughan, N. K. (2003). *The relationship between student performance and the leadership behavior of superintendents in Texas public school districts*. [Unpublished doctoral dissertation]. Texas A&M University.

*Veltri, P. J. (2002). *The relationship between school districts' planning practices, student achievement and, the implementation of the Correlates of Effective Schools*. [Unpublished doctoral dissertation]. University of Idaho.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*(1), 89-122.

Weatherly, R. & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review*, 47(2), 171-197.

Weick, K.E. (1976). Educational organizations as loosely coupled systems. *Administration Science Quarterly*, 21, 1-19.

West, M., Morton, B., & Herlihy, C. (2016). Achievement Network's investing in innovation expansion: Impacts on educator practice and student achievement. *Grantee Submission.*

*Wiswell, C. L. (2011). *Superintendent emotional intelligence and student achievement: A quantitative study*. [Unpublished doctoral dissertation]. Texas Technology University.

*Wooderson-Perzan, M. K. (2001). *The relationship of superintendent leadership styles to student achievement and school district financial and demographic factors in Texas*. [Unpublished doctoral dissertation]. Sam Houston State University.

Wong, K. & Shen, F. (2003). Measuring the effectiveness of city and state takeover as a school reform strategy. *Peabody Journal of Education*, 78(4), 89-119.

Wong, K. & Shen, F. (2013). Mayoral governance and student achievement. *Center for American Progress.*

*Wong, K. K., Shen, F. X., Anagnostopoulos, D., & Rutledge, S. (2007). Evaluating the Effects of Mayoral Control. In G. W. Boychuk, K. Mossberger & M. C. Rom (Eds). *The Education Mayor: Improving America's Schools (pp. 52-75)*. Georgetown University Press.

Yoon, K., Duncan, T., Lee, S., Scarloss, B. & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement.* REL 2007-No. 033.

Tables

Table 1. Describing the Sample of Studies (N=110)

Coding Category	Specific Code	N	%
Publication Type			
	Peer-Reviewed Journal	71	64.5%
	Non-Peer Reviewed Report	12	10.9%
	Book Chapter	3	2.7%
	Dissertation	24	21.8%
Country			
	Canada	1	0.9%
	Israel	2	1.8%
	Netherlands	2	1.8%
	Norway	1	0.9%
	United Kingdom	5	4.5%
	United States	99	90.0%
Research Design			
	Randomized Control Trial	3	2.7%
	Regression Discontinuity	5	4.5%
	Difference-in-Differences	23	20.9%
	Observational	79	71.8%
Number of Policy Inputs			
	One	84	76.4%
	Two or Three	17	15.5%
	Four of More	9	8.2%
Outcome(s)			
	Student Achievement	96	87.3%
	High School Graduation	6	5.5%
	Other K-12 Student Measures	4	3.6%
	Teacher Turnover	9	8.2%
	Teacher Quality Measures	6	5.5%

Notes: All categories are mutually exclusive except for outcomes, as some studies look at relationships between a district policy and multiple student- or teacher-level measures. Thus, for this category, percentages do not sum to 100%. "Other K-12 School Behaviors" includdes course grades, attendance, suspensions, and retained in grade. Teacher quality measures include college/major, teaching experience, licensure/exam scores, certification, and value-added to student test scores.

Table 2a. Findings by Type of District Input: Governance and Politics

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings	
			Causal Studies	Observational Studies
Superintendent characteristics	33	Adams, 1987[4]; Berlau, 2011[4]; Brewer, 1996[4]; Burnett, 1990[4]; Byrd, 2002[4]; Clore, 1992[4]; Cotter, 2002[4]; Duvall, 2005[4]; Edwards, 2007[4]; Endeman, 1992[4]; Engel, 2016[4]; Hanks, 2010[4]; Hart, 1983[4]; Hart & Ogawa, 1987[4]; Hart, Schramm-Possinger, & Hoyle, 2019[4]; Henderson, Simar, & Wang, 2017[4]; Hough, 2014[4]; Hoyle et al., 2001[4]; Johnson, 1997[4]; Lavy & Boiko, 2017[4]; Libka, 2012[4]; Morgan, 1990[4]; Muller, 1990[4]; Myers, 2011[4]; Parker-Chenaille, 2012[4]; Plotts, 2011[4]; Plotts & Gutmore, 2014[4]; Rudolph, 2016[4]; Simpson, 2013[4]; Vaughan, 2003[4]; Veltri, 2002[4]; Wiswell, 2011[4]; Wooderson-Perzan, 2001[4]	N/A	Superintendent value-added to test scores, higher expectations, instructional leadership, internal hires, emotional intelligence, and humility associated with higher student achievement. Positive or null relationship between superintendent longevity, leadership strategies, spending on district-level administration, superintendent accountability behaviors, vision, and student achievement.
Union strength	15	Eberts & Stone, 1984[4]; Eberts & Stone, 1987[4]; Figlio, 2002[3]; Harris & Larsen, 2016[3]*; Henry, et al., 2019[3]*; Hoxby, 1996[3]; Lincove, Barrett, & Strunk, 2018[4]*; Lovenheim, 2009[3]; Moe, 2009[4]; Pham et al., 2018[3]*; Register & Grimes, 1991[4]; Rose & Sonstelie, 2010[4]; Schueler, Goodman, & Deming, 2017[3]*; Strunk, 2011[4]; Strunk & McEachin, 2011[4]	Unionization associated with lower rates of student achievement. Some evidence of null effects.	Mixed results on relationship between unionization / union strength / collective bargaining agreement restrictiveness and student achievement. Suggestive evidence unionization decreases teacher qualifications and turnover.
District leaders elected or appointed	10	Gill et al., 2007[3]*; Harris & Larsen, 2016[3]*; Henry et al., 2019[3]*; Hoover, 2008[4]; Ford & Ihrke, 2016b[4]; Lincove, Barrett, & Strunk, 2018[4]*; Partridge & Sass, 2011[3]; Pham et al., 2018[3]*; Schueler, Goodman, & Deming, 2017[3]*; Wong & Shen, 2007[4]	Mixed results on relationship between state takeover, appointed (vs. elected) superintendent, and student achievement.	No relationship between whether either a school board or superintendent is elected or appointed and student achievement. Ability for mayor to appoint majority of school board associated with higher student achievement while ability to appoint all members associated with lower achievement.
School autonomy	9	Allen, 2013[2]; Chin et al., 2019[3]*; Harris & Larsen, 2016[3]*; Henry et al., 2019[3]*; Lincove, Barrett, & Strunk, 2018[4]*; Naper, 2010[4]; Pham et al., 2018[3]*; Phenix et al, 2005[3]*; Schueler, Goodman & Deming, 2017[3]*	School autonomy associated with higher student achievement, although some mixed results.	Decentralized hiring associated with greater efficiency (grades/hours per student). Some evidence autonomy associated with teacher turnover.
School board characteristics	7	Duvall, 2005[4]; Ford, 2013[4]; Ford & Ihrke, 2016a[4]; Grissom, 2014[4]; Hofman, 1995[4]; Saatcioglu et al., 2011[4]; Shi & Singleton, 2019[4]	N/A	Involvement of school-level staff in board decision-making, higher board member social capital, board-superintendent collaboration, pluralistic and data-driven boards, board strategic planning, and adherence to board best practices associated with higher student achievement. No relationship between educator representation on boards and student achievement.
Portfolio management model	7	Chin et al., 2019[3]*; Gill et al., 2007[3]*; Harris & Larsen, 2016[3]*; Henry et al., 2019[3]*; Lincove, Barrett & Strunk, 2018[4]*; Pham et al., 2018[3]*; Schueler, Goodman & Deming, 2017[3]*	Mixed results on association between portfolio model and student achievement.	Some evidence portfolio models associated with higher teacher turnover.
District administrators	5	Brewer, 1996[4]; Harris & Larsen, 2016[3]*; Henderson, Simar and Wang, 2017[4]; Lincove, Barrett, & Strunk, 2018[4]*; Schueler, Goodman & Deming, 2017[3]*	More district administrators associated with lower student achievement.	Number of district administrators not associated with student achievement. Fewer district administrators associated with teacher turnover.
Parental school choice	5	Chin et al., 2019[3]*; Gibbons, Machin & Silva, 2007[4]; Harris & Larsen, 2016[3]*; Lavy, 2010[2]; Lincove, Barrett & Strunk, 2018[4]*	School choice offerings associated with higher student achievement.	No relationship between school choice offerings associated and student achievement. Some evidence choice associated with higher teacher turnover.
School size	1	de Haan, Leuven & Oosterbeek, 2016[4];	N/A	Larger school size requirements associated with higher student achievement.
Selective admissions	1	Gray, Jesson & Jones, 1984[4];	N/A	Selective admissions not associated with higher student achievement.
School closures	1	Chin et al., 2019[3]*	Closure of low-performing schools associated with higher district-wide academic performance.	N/A

Notes: Each citation ha a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; in the summary of findings, we denote designs 1 through 3 as causal and 4 as observational. * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. Cells are shaded to describe the overall pattern of relationships: green = positive relationships, yellow = null or mixed relationships, and red = negative relationships.

Table 2b. Findings by Type of District Input: Recruitment and Retention of Teachers

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings	
			Causal Studies	Observational Studies
Teacher salaries and benefits	20	Biasi, 2018[3]; Booker & Glazerman, 2009[4]; Chin et al., 2019[3*]; Figlio, 2002[3]; Harris & Larsen, 2016[3*]; Hendricks, 2014[3]; Hendricks, 2015[3]; Henry et al., 2019[3*]; Houck, Rolle & He, 2010[4]; Lin, 2010[4]; Lin & Quayes, 2006[4]; Lincove, Barrett, & Strunk, 2018[4*]; Murnane & Olsen, 1990[4]; Pham et al., 2018[3*]; Podolsky et al., 2019[4]; Schueler, Goodman & Deming, 2017[3*]; Sojourner, Mykerezi & West, 2014[3]; Theobald & Gritz, 1996[4]; Theobald, 1990[4]; Tran, 2017[3]	Higher teacher salaries associated with higher student achievement, higher teacher quality and higher teacher retention. Flexible pay schemes (e.g., merit-based, career ladder) associated with higher teacher quality and higher student achievement.	Higher teacher salaries associated with higher student achievement and higher teacher quality. Mixed results for teacher retention. Flexible pay schemes (e.g., career ladder) associated with higher teacher quality and higher student achievement.
Teacher hiring practices	7	Balter & Duncombe, 2008[4]; Harris & Larsen, 2018[3*]; Henry et al., 2019[3*]; Naper, 2010[4]; Pham et al., 2018[3*]; Schueler, Goodman & Deming, 2017[3*]	Teacher replacements associated with higher student achievement.	Use of more recruitment practices associated with higher teacher qualifications. Decentralized hiring associated with greater efficiency (grades/hours per student). Teacher replacements associated with lower teacher retention.
Teacher seniority protections	3	Cohen-Vogel, Feng & Osborne-Lampkin, 2013[4]; Goldhaber, Lavery & Theobald, 2016[3]; Koski & Horng, 2007[3]	No evidence of association between voluntary transfer provisions and between-district teacher mobility. Seniority preference associated with higher teacher qualifications. Mixed evidence on relationship between seniority protections and both average teacher quality and within-district inequity in teacher quality.	No relationship between seniority protections and within-district inequity in teacher quality.
Teacher evaluation systems	3	Chin et al., 2019[3*]; Pham et al., 2018[3*]; Schueler, Goodman & Deming, 2017[3*]	Changes to teacher evaluation system associated with higher student achievement. No studies that isolate evaluation system from other districtwide reforms.	N/A

Notes: Each citation ha a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; in the summary of findings, we denote designs 1 through 3 as causal and 4 as observational. * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. Cells are shaded to describe the overall pattern of relationships: green = positive relationships, yellow = null or mixed relationships, and red = negative relationships.

Table 2c. Findings by Type of District Input: Instructional Resources and Development

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings	
			Causal Studies	Observational Studies
Data use	11	Carlson, Borman & Robinson, 2011[1]; Chin et al., 2019[3*]; Gandhi et al., 2018[3*]; Lee, Louis & Anderson, 2012[4]; Leithwood & Azah, 2017[4]; May & Robinson, 2007[1]; Phenix et al., 2005[3*]; Schueler, Goodman & Deming, 2017[3*]; Slavin et al., 2013[1]; Strunk & McEachin, 2014[3*]; Strunk, McEachin & Westover, 2018[3*]	Data use associated with higher student achievement.	Evidence use associated with higher student achievement. Data use associated with worse instructional practices.
Student-teacher ratio	9	Alexander & Griffin, 1976[4]; Bidwell & Kasarda, 1975[4]; Hofman, 1995[4]; Lin & Quayes, 2006[4]; Mensah, Schoderbek & Sahay, 2013[3]; Phenix et al., 2005[3*]; Podolsky et al., 2019[4]; Steele, Vignoles & Jenkins, 2007[4]; Theobald, 1990[4]	Suggestive evidence that lower ratios associated with higher student achievement, though not always consistent across models.	Lower ratios generally associated with higher student achievement, though not always statistically significant.
Curriculum materials	8	Bhatt & Koedel, 2012[4]; Bhatt, Koedel & Lehmann, 2013[4]; Blazar et al., 2020[4]; Chin et al., 2019[3*]; Gandhi et al., 2018[3*]; Koedel et al., 2017[4]; Phenix et al., 2005[3*]; Reys et al., 2003[4]	Suggestive evidence curricular changes associated with higher student achievement.	Evidence of variation in effectiveness of particular curricular materials from single-state studies but less evidence variation is consistent across time and states.
Principal professional development	8	Carlson, Borman & Robinson, 2011[1]; Chin et al., 2019[3*]; Leithwood & Azah, 2017[4]; Phenix et al., 2005[3*]; Schueler, Goodman & Deming, 2017[3*]; Strunk & McEachin, 2014[3*]; Strunk, McEachin, & Westover, 2018[3*]	Professional development associated with higher student achievement.	Professional development associated with higher student achievement.
Teacher professional development	7	Chin et al., 2019[3*]; Gandhi et al., 2018[3*]; Gill et al., 2007[3*]; Jacques & Brorsen, 2002[4]; Lee, Louis & Anderson, 2012[4]; Leithwood & Azah, 2017[4]; Phenix et al., 2005[3*]	Professional development generally associated with higher student achievement though one study finds no association.	Professional development associated with better instructional practices and higher student achievement.
Extra learning time	5	Chin et al., 2019[3*]; Hinrichs, 2011[4]; Phenix et al., 2005[3*]; Schueler, Goodman & Deming, 2017[3*]; Sims, 2008[3]	Extra learning time associated with higher student achievement.	No association between school start times and student achievement.
School support staff (e.g., counselors)	5	Bidwell & Kasarda, 1975[4]; Brewer, 1996[4]; Gandhi et al., 2018[3*]; Goodman & Young, 2006[4]; Jacques & Brorsen, 2002[4]	Wraparound services that require hiring support staff associated with higher student achievement.	Mixed results on relationship between number of support staff (e.g., counselors, psychologists) and student achievement.
Facilities	3	Cellini, Ferreira, & Rothstein, 2010[2]; Kai & Zimmer, 2016[2]; Martorell, Stange, & McFarlin, 2016[2]	No consistent association between facilties (measured through capital campaigns) and student achievement or teacher mobility.	N/A
Coherence of instructional resources	2	Leithwood & Azah, 2017[4]; Phenix et al., 2005[3*]	Coherence of instructional program associated with higher student achievement.	Coherence of instructional program associated with higher student achievement.
Technology	2	Chin et al., 2019[3*]; Machin, McNally, & Silva, 2007[3]	Technology associated with higher student achievement.	N/A
School grade configurations	1	Cook et al., 2008[3]	Attending middle school, instead of elementary school, in 6th grade is associated with lower student achievement.	N/A

Notes: Each citation ha a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; in the summary of findings, we denote designs 1 through 3 as causal and 4 as observational. * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. Cells are shaded to describe the overall pattern of relationships: green = positive relationships, yellow = null or mixed relationships, and red = negative relationships.