JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2024, VOL. 17, NO. 4, 647–686
https://doi.org/10.1080/19345747.2023.2269923

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Effective School District Policies and Practices: Synthesizing Theoretical Frameworks and Empirical Findings across Disciplines

David Blazar[a] and Beth Schueler[b]

[a]Teaching and Learning, Policy and Leadership, University of Maryland at College Park, College Park, Maryland, USA; [b]Education Leadership, Foundations, and Policy, University of Virginia, Charlottesville, Virginia, USA

ABSTRACT

What guidance does research provide about how to improve school district performance in the United States? 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 for researchers, we first review the theoretical literatures on how districts are thought to affect educational outcomes, arguing that an expanded set of disciplinary perspectives—organizational behavior, political science, economics—have distinct theories about the types of district-level policies that might improve district-wide performance. Using these frameworks as a guide, we next conduct a review of quantitative studies that estimate the relationship between district-level inputs and educational outcomes, finding benefits of policies that cross disciplinary perspectives: higher teacher salaries, data use, and school autonomy and parental choice in the context of district-wide turnarounds. Our review also reveals the need for significant additional causal evidence and provides an inter-disciplinary map of theorized pathways through which district-level policies could influence student outcomes that are ripe for rigorous testing.

ARTICLE HISTORY
Received 25 May 2022
Revised 27 June 2023
Accepted 31 August 2023

KEYWORDS
District effectiveness; education policy; multi-disciplinary; systematic review

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 (Spillane & Thompson, 1997; Supovitz, 2006). District-wide policymaking may provide a

CONTACT David Blazar dblazar@umd.edu Teaching and Learning, Policy and Leadership, University of Maryland at College Park, College Park, Maryland, USA.

more efficient approach to reform than school- or classroom-level change through economies of scale, while also attending to local needs and contexts in a way that may be less feasible at the state level. For example, 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 researchers have for district leaders about how to run effective school systems? Our work follows a "production function" framework that aims to draw generalizable conclusions about "what works" in terms of district-level and district-wide policymaking, and the set of policy alternatives that may be responsible for improved educational outcomes. The production function literature in education research stems back to the Coleman report (1966) and was expanded substantially in the following decades with pioneering work by Hanushek (1979), Monk (1989), and others who quantitatively linked schooling resources to student achievement gains and other desired educational outcomes. At the same time, reviews of this line of work have focused primarily on school- and classroom-level variation in inputs, rather than policy alternatives generally set at the district level (Fryer, 2017; Hanushek, 2003)—despite questions of effective district-level policymaking being so fundamental to U.S. education policy.

This is not to say that the existing literature base on district effectiveness is uninformative for researchers providing guidance to local leaders. Looking beyond the education production function literature, scholars from other traditions have, for decades, described the importance of district-level policymaking and have hypothesized key dimensions of effective districts (e.g., Anderson & Young, 2018; Murphy & Hallinger, 1988; Opfer et al., 2008). Drawing primarily on qualitative case study designs, educational leadership scholars, for example, describe how district-level leaders and district-level policymaking are critical for building capacity for improved instruction and reorienting the organization toward reform (Leithwood, 2010; Rorrer et al., 2008; Trujillo, 2013). 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 classrooms (e.g., professional development) (Elmore & Burney, 1998; Honig & Hatch, 2004; O'Day, 2002).

A primary aim of our work, then, is to bring together multiple perspectives and lines of inquiry to provide a more comprehensive picture of school district policies and practices that can improve district-wide performance. Cross- and inter-disciplinary approaches are important to all areas of educational research—which is an academic field that has been studied by scholars from varied perspectives and academic traditions—but are likely particularly valuable in the study of effective district policies and practices for several reasons. First, existing frameworks on district effectiveness are

relatively narrow in scope because they tend to be dominated by educational leaderships scholars interested in policies related to instructional improvement and development (e.g., Anderson and Young, 2018; Leithwood, 2010; Trujillo, 2013). 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 from political science—often focused on district governance arrangements (e.g., Henig, 2009; Weatherley & Lipsky, 1977)—and economics—often focused on the supply of and demand for teacher talent (e.g., Boyd et al., 2005; Murnane & Olsen, 1990). Any one perspective on its own provides a limited understanding of the potential set of effective district policies and practices.

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, production-function style analyses that link inputs to outcomes and are more common in other traditions such as economics. In a recent review from the educational leadership perspective relying on the specific search term phrases, "district practices" and "district structures," 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 review to cross-disciplinary perspectives, we find 99 quantitative, production function-style 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 their ability to isolate practices that result in improved outcomes. Despite the strengths of deep-dive qualitative case study and exploratory quantitative work for hypothesis building and theory generation, we argue that quantitative analyses that pay specific attention to causality are critical for identifying effective district policies and practices.

To expand upon existing district effectiveness frameworks and empirical syntheses, we begin by reviewing the theoretical literatures on how districts are thought to affect educational outcomes, looking across the three primary disciplinary approaches to the topic: (i) organizational behavior with a close link to educational leadership, (ii) political science, and (iii) economics. Using this expanded set of frameworks as a guide, we then review the empirical literature testing these theorized pathways. We limit studies in our review to U.S.-based analyses that use quantitative research designs, and where policies under investigation are implemented and studied at the district level. We further limit to studies that link policy inputs to educational outcomes, and that attempt in some way to address selection and omitted variables bias. Given our interest in guiding research on the causal effects of various district-level policy inputs, we further code studies and describe patterns of findings along a hierarchy of evidence quality for the purpose of drawing causal inference. Though not an inclusion criterion per se, we also note for readers that we exclude 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 a discussion of this topic, see Jackson, 2020).

We find that studies most likely to support causal claims reveal benefits of district-level and district-wide policies that cross disciplinary perspectives: higher teacher salaries

(a focus of labor market theory); data use (a focus of organizational behavior and educational leadership perspectives); and school autonomy and school choice systems (a focus of political science theory), though this evidence comes primarily from studies of district turnarounds with a broad package of reforms limiting the ability to isolate the effectiveness of any single component of reform. 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 effective district policies and practices. A majority of the quantitative articles in our sample are unlikely to support causal conclusions (73%). Those that can support causal claims focus on a narrow set of policy inputs (e.g., teacher salaries, data use, facilities). Further, studies generally define outcomes narrowly through student test-score performance, even though theory describes a much more dynamic logic model linking district-level policies to a broad set of performance measures. For example, 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 identifies just 15% of included studies that examine 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.

In our discussion section, we provide recommendations for future research design quality to support causal conclusions, as well as to feasibly expand a set of outcomes within existing data system constraints. We also situate our findings on the *effectiveness* of specific district-level policies and practices with equally important questions on *efficiency*. In other words, how do the effects of district-level policies examined in our review compare to the effects of similar or related policy inputs that may be implemented at other levels of the education system (e.g., state, school, classroom)? After all, a fundamental premise of district-based authority in the U.S. school system is that the district is attune to local needs while also providing economies of scale for reform (Goldin & Katz, 2008; Supovitz, 2006). While it is beyond the scope of our paper to systematically review and then compare effects of all possible policy inputs at all possible levels of implementation, we believe there is some evidence to suggest that policies—such as the use of educational data to guide reform—may be more effective when implemented at the district level versus other levels (e.g., at schools). Future research should pick up our findings with additional attention to this important question.

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 within-school resources and school-level policies that drive student outcomes. A critical finding from this line of work is the outsized effect that individual teachers have on improving student performance (Hanushek & Rivkin, 2010), with some suggestion that school-level policies that aim to increase teacher quality—such as peer collaboration and the formation of instructional teams—can affect both teacher and student outcomes (e.g., Jackson & Bruegmann, 2009; Ronfeldt et al., 2015).

The study of effective districts also stems, in part, from this earlier work in that scholars have explored districts as an additional source of variation in student outcomes (e.g., Bowers, 2010; Bowers et al., 2018), hypothesizing that district-level policymaking may be a key avenue for allocating resources to and supporting schools (e.g., Supovitz, 2006). (We discuss these specific hypotheses in detail in the following section.) However, this literature base on districts 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) examine 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%) is expected, given that districts are farther removed from students than schools. That said, in similar analyses of North Carolina data, Chingos et al. (2015) argue that documented variation in student test scores across districts (roughly 2%) is substantively meaningful. A 1 standard deviation (SD) difference in district effectiveness is associated with a 0.11 SD difference in student test scores, roughly equivalent to nine weeks of schooling for the average student. Following closely with the Coleman tradition, Fahle and Reardon (2018) also document 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.

Mirroring findings from the school-level variance decomposition literature, a likely mechanism for between-district variation relates to teachers. Goldhaber et al. (2018) find 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 occurs across districts and schools, rather than within schools. These patterns also are seen in other states (e.g., Glazerman & Max, 2011). Similarly, Blazar et al. (2016) find very large differences in measures of teachers' instructional practice across several urban school districts (upwards of 1 SD). Using the same dataset, Hill et al. (2015) find that districts explain much more of the variation in math-specific teaching practices than schools; the reverse is true for content-neutral teaching practices (e.g., classroom

Table 1. Describing the sample of studies ($N = 99$).

Coding Category	Specific Code	*N*	%
Publication Type			
	Peer-Reviewed Journal	62	62.6%
	Non-Peer Reviewed Report	10	10.1%
	Book Chapter	3	3.0%
	Dissertation	24	24.2%
Research Design			
	Randomized Control Trial	3	3.0%
	Regression Discontinuity	3	3.0%
	Difference-in-Differences	21	21.2%
	Observational	72	72.7%
Number of Policy Inputs			
	One	78	78.8%
	Two or Three	12	12.1%
	Four of More	9	9.1%
Types of Policy Inputs			
	Governance (Political Science)	82	49.7%
	Recruitment & Retention of Teachers (Economics)	30	18.2%
	Instructional Resources & Development (Education Leadership)	53	32.1%
Outcome(s)			
	Student Achievement	86	86.9%
	High School Graduation	6	6.1%
	Other K-12 Student Measures	4	4.0%
	Teacher Turnover	9	9.1%
	Teacher Quality Measures	6	6.1%

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, and types of policy inputs, as some studies examined multiple inputs. 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.

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.

Though the extant literature describing the magnitude of district effects is small relative to the literature on school 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 more formal review of the quantitative studies examining this topic.

Synthesis of Frameworks on District Capacity to Influence Educational Outcomes

To begin answering our driving question regarding effective district policies and practices, 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 limited information on the set of district-level policy inputs that can improve district performance.

Table 2. Findings by Type of district input: Governance and politics.

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings: Causally-Oriented Studies: Policy Input Uniquely Isolated	Summary of Findings: Causally-Oriented Studies: Multiple Policy Inputs not Uniquely Isolated	Summary of Findings: Observational Studies
Superintendent characteristics	32	(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 et al., 2019[4]; Henderson et al., 2017[4]; Hough, 2014[4]; Hoyle et al., 2001[4]; Johnson, 1997[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	N/A	32 studies: 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	14	(Eberts & Stone, 1984[4]; Eberts & Stone, 1987[4]; Figlio, 2002[3]; Harris & Larsen, 2016[3]*; Hoxby, 1996[3]; Lincove et al., 2018[4]*; Lovenheim, 2009[3]; Moe, 2009[4]; Pham et al., 2020[3]*; Register & Grimes, 1991[4]; Rose & Sonstelie, 2010[4]; Schueler et al., 2017[3]*; Strunk, 2011[4]; Strunk & McEachin, 2011[4])	3 studies: Mixed effects of unionization on high school dropout and on the quality of new teacher hires.	3 studies: Unionization associated with lower rates of student achievement.	8 studies: 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	9	(Gill et al., 2007[3]*; Harris & Larsen, 2016[3]*; Hoover, 2008[4]; Ford & Ihrke, 2016b[4]; Lincove et al., 2018[4]*; Partridge & Sass, 2011[3]; Pham et al., 2020[3]*; Schueler, Goodman, & Deming, 2017[3]*; Wong et al., 2007[4])	1 study: Inconsistent effects on student test scores for district leader appointed vs. elected.	4 studies: Mixed results on relationship between state takeover, appointed (vs. elected) superintendent, and student achievement.	4 studies: 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.
Portfolio management model	6	(Chin et al., 2019[3]*; Gill et al., 2007[3]*; Harris & Larsen, 2016[3]*; Lincove et al., 2018[4]*; Pham et al., 2020[3]*; Schueler, Goodman & Deming, 2017[3]*)	N/A	5 studies: Mixed results on association between portfolio model and student achievement.	1 study: Some evidence portfolio models associated with higher teacher turnover.

(*continued*)

Table 2. Continued.

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings: Causally-Oriented Studies: Policy Input Uniquely Isolated	Summary of Findings: Causally-Oriented Studies: Multiple Policy Inputs not Uniquely Isolated	Summary of Findings: Observational Studies
School autonomy	6	(Chin et al., 2019[3]*; Harris & Larsen, 2016[3]*; Lincove et al., 2018[4]*; Pham et al., 2020[3]*; Phenix et al., 2005[3]*; Schueler, Goodman & Deming, 2017[3]*)	N/A	5 studies: School autonomy associated with higher student achievement.	1 study: Some evidence autonomy associated with higher teacher turnover.
School board characteristics	6	(Duvall, 2005[4]; Ford, 2013[4]; Ford & Ihrke, 2016a[4]; Grissom, 2014[4]; Saatcioglu et al., 2011[4]; Shi & Singleton, 2019[4])	N/A	N/A	6 studies: 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.
District administrators	5	Brewer, 1996[4]; Harris & Larsen, 2016[3]*; Henderson, Simar and Wang, 2017[4]; Lincove et al., 2018[4]*; Schueler, Goodman & Deming, 2017[3]*	N/A	2 studies: More district administrators associated with lower student achievement.	3 studies: Number of district administrators not associated with student achievement. Fewer district administrators associated with teacher turnover.
Parental school choice	3	(Chin et al., 2019[3]*; Harris & Larsen, 2016[3]*; Lincove et al., 2018[4]*)	N/A	2 studies: School choice offerings associated with higher student achievement.	1 study: Some evidence choice associated with higher teacher turnover.
School closures	1	(Chin et al., 2019[3]*)	N/A	1 study: Closure of low-performing schools associated with higher district-wide academic performance.	N/A
School admissions	0	N/A	N/A	N/A	N/A
School size	0	N/A	N/A	N/A	N/A

Notes. Each citation has a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. In the summary of findings, we denote designs 1 through 3 as causal if the policy input is uniquely isolated from other inputs (superscript bolded), or causally-oriented but where the policy input is not uniquely separated from other inputs (superscript in italics); and 4 as observational. Cells are shaded to describe the overall pattern of relationships: light gray = positive relationships, dark gray = null or mixed relationships, and black = negative relationships.

Table 3. 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: Causally-Oriented Studies: Policy Input Uniquely Isolated	Summary of Findings: Causally-Oriented Studies: Multiple Policy Inputs not Uniquely Isolated	Summary of Findings: Observational Studies
Teacher salaries and benefits	19	(Biasi, 2021[3]; Booker & Glazerman, 2009[4]; Chin et al., 2019[3]*; Figlio, 2002[3]; Harris & Larsen, 2016[3]*; Hendricks, 2014[3]; Hendricks, 2015[3]; Houck et al., 2010[4]; Lin, 2010[4]; Lin & Quayes, 2006[4]; Lincove et al., 2018[4]*; Murnane & Olsen, 1990[4]; Pham et al., 2020[3]*; Podolsky et al., 2019[4]; Schueler et al., 2017[3]*; Sojourner et al., 2014[3]; Theobald & Gritz, 1996[4]; Theobald, 1990[4]; Tran, 2017[3])	6 studies: Higher teacher salaries results in higher teacher quality, and higher teacher retention. Flexible pay schemes (e.g., merit-based, career ladder) results in with higher teacher quality and higher student achievement.	4 studies: 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.	9 studies: 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	5	(Balter & Duncombe, 2008[4]; Harris & Larsen, 2016[3]*; Pham et al., 2020[3]*; Schueler, Goodman & Deming, 2017[3]*; Lincove et al., 2018)	N/A	3 studies: Teacher replacements associated with higher student achievement.	2 studies: Use of more recruitment practices associated with higher teacher qualifications. Teacher replacements associated with lower teacher retention.
Teacher seniority protections	3	(Cohen-Vogel et al., 2013[4]; Goldhaber et al., 2016[3]; Koski & Horng, 2007[3])	2 studies: Seniority protections result in greater percentage of credentialed teachers and greater within-district retention of effective teachers in disadvantaged schools. No effect on between-district teacher mobility.	N/A	1 study: No relationship between seniority protections and within-district inequity in teacher quality.
Teacher evaluation systems	3	(Chin et al., 2019[3]*; Pham et al., 2020[3]*; Schueler, Goodman & Deming, 2017[3]*)	N/A	3 studies: 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 has a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. In the summary of findings, we denote designs 1 through 3 as causal if the policy input is uniquely isolated from other inputs (superscript bolded), or causally-oriented but where the policy input is not uniquely separated from other inputs (superscript in italics); and 4 as observational. Cells are shaded to describe the overall pattern of relationships: light gray = positive relationships, dark gray = null or mixed relationships, and black = negative relationships.

By looking across these perspectives, we generate a fairly exhaustive list of the major district-level policies and inputs scholars have theorized should influence performance—which we list in Tables 2–4—and that we subsequently search for in the empirical literature that quantitatively links inputs to outcomes. As we reviewed the theoretical literatures, we tagged keywords describing the policy inputs, as used by the study authors. Through an iterative process, we then grouped similar keywords and narrowed the list until we felt that each policy input was mutually exclusive from others. Because disciplinary traditions overlap in some policy inputs discussed but vary in terms of language and definitions, we describe details of the policy input categories in our presentation of results from the review of the empirical literature.

Organizational Behavior and Educational Leadership

Prior reviews and treatments of "district effectiveness" stem largely from scholars in the traditions of organizational behavior and organizational sociology and the closely related field of educational leadership. Thus, these prior reviews primarily focus 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 is sometimes 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 et al., 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 are thought to develop school capacity for reform, including through efforts to apply evidence to decision-making (Honig, 2012; Honig & Coburn, 2008). Coherent systems start their work with strategies for teaching and learning, but do not end there. High-performing systems also are theorized to hold a performance-oriented culture that emphasizes data and results; strategically allocate resources (e.g., curricular materials, technology); build stakeholder support across a variety of constituents including teachers' unions, parents, school boards, community and advocacy groups; and address environmental factors that can influence student wellbeing related to health, nutrition, and safety (Childress et al., 2006).

Much of the organizational theory and educational leadership framework is 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 for turning around literacy and then mathematics

Table 4. Findings by Type of district input: Instructional resources and development.

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings: Causally-Oriented Studies: Policy Input Uniquely Isolated	Summary of Findings: Causally-Oriented Studies: Multiple Policy Inputs not Uniquely Isolated	Summary of Findings: Observational Studies
Data use	9	(Carlson et al., 2011[1]; Chin et al., 2019[3]*; Gandhi et al., 2018[3]*; Lee et al., 2012[4]; May & Robinson, 2007[1]; Phenix et al., 2005[3]*; Schueler et al., 2017[3]*; Slavin et al., 2013[1]; Strunk et al., 2014[3]*)	3 studies: Data use results in higher student achievement.	5 studies: Data use associated with higher student achievement.	1 study: Data use associated with stronger teacher professional community but less focused instructional practices.
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])	N/A	3 studies: Suggestive evidence curricular changes associated with higher student achievement.	5 studies: Evidence of variation in effectiveness of particular curricular materials from single-state studies but less evidence variation is consistent across time and states.
Student-teacher ratio/ Class size	7	(Alexander & Griffin, 1976[4]; Bidwell & Kasarda, 1975[4]; Lin & Quayes, 2006[4]; Mensah et al., 2013[3]; Phenix et al., 2005[3]*; Podolsky et al., 2019[4]; Theobald, 1990[4])	1 study: Suggestive evidence that lower ratios results in higher student achievement, though not always consistent across models.	1 study: Lower ratios associated with higher student achievement in math but not reading.	5 studies: Lower ratios generally associated with higher student achievement, though not always statistically significant.
Wellbeing supports	7	(Bidwell & Kasarda, 1975[4]; Brewer, 1996[4]; Gandhi et al., 2018[3]*; Goodman & Young, 2006[4]; Jacques & Brorsen, 2002[4]; Lacoe & Steinberg, 2018[3]; Ruffini, 2022[3])	2 studies: No effect on student achievement for offering free meals to all students. No effect on suspensions for rollback of zero-tolerance discipline policy.	1 study: Wraparound services that require hiring support staff associated with higher student achievement.	4 studies: Mixed results on relationship between number of support staff (e.g., counselors, psychologists) and student achievement.
Teacher professional development	6	(Chin et al., 2019[3]*; Gandhi et al., 2018[3]*; Gill et al., 2007[3]*; Jacques & Brorsen, 2002[4]; Lee et al., 2012[4]; Phenix et al., 2005[3]*)	N/A	4 studies: Teacher professional development generally associated with higher student achievement though one study finds no association.	2 studies: Teacher professional development associated with better instructional practices and higher student achievement.

(*continued*)

Table 4. Continued.

Policy Input (sorted by N of studies)	N of studies	Citations	Summary of Findings: Causally-Oriented Studies: Policy Input Uniquely Isolated	Summary of Findings: Causally-Oriented Studies: Multiple Policy Inputs not Uniquely Isolated	Summary of Findings: Observational Studies
Principal professional development	5	(Carlson et al., 2011[1]; Chin et al., 2019[3]*; Phenix et al., 2005[3]*; Schueler et al., 2017[3]*; Strunk et al., 2014[3]*)	1 study: Principal professional development in the context of a data-use intervention results in higher student achievement.	4 studies: Principal professional development associated with higher student achievement.	N/A
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])	1 study: Extra learning time results in higher math student achievement, as well as reading achievement for students in the upper end of the test-score distrubution.	3 studies: Extra learning time associated with higher student achievement.	1 study: No association between school start times and student achievement.
Facilities	3	(Cellini et al., 2010[2]; Kai & Zimmer, 2016[2]; Martorell et al., 2016[2])	3 studies: Findings mixed across studies. No consistent effect of capital campaigns for facilities on student achievement or teacher mobility.	N/A	N/A
Coherence of instructional resources	1	(Phenix et al., 2005[3]*)	N/A	1 study: Coherence of instructional program associated with higher student achievement in math but not reading.	N/A
School grade configurations	1	(Cook et al., 2008[3])	1 study: Attending middle school, instead of elementary school, in 6th grade results in lower student achievement.	N/A	N/A
Technology	1	(Chin et al., 2019[3]*)	N/A	1 study: Technology associated with higher student achievement.	N/A

Notes. Each citation has a superscript identifying the specific research design: 1 = randomized control trial, 2 = regression discontinuity design, 3 = difference-in-differences design, 4 = observational; * = studies that evaluate the effect of a package of districtwide reforms, making it more difficult to isolate the impacts of specific intervention components. In the summary of findings, we denote designs 1 through 3 as causal if the policy input is uniquely isolated from other inputs (superscript bolded), or causally-oriented but where the policy input is not uniquely separated from other inputs (superscript in italics); and 4 as observational. Cells are shaded to describe the overall pattern of relationships: light gray = positive relationships, dark gray = null or mixed relationships, and black = negative relationships.

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).

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 this 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 tend to focus on district governance structures, often arguing that these factors can create barriers to district-wide improvement. As such, they theorize 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 (Weatherley & Lipsky, 1977). This perspective informs, in part, the work of educational leadership scholars who often argue 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) 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 where the district oversees schools that operate under a variety of managers such as the central office, private organizations, and charter operators (Bulkley et al., 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, as well as increase engagement across other key constituents (e.g., parents, community groups).

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 traditionally 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;

each state administered its own assessment, making cross-state outcome comparisons historically 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.

Economics

Economists theorizing about how school districts influence student outcomes often focus on the relatively decentralized nature of school systems, particularly in the U.S. They argue that this distinct feature generally promotes positive student outcomes through increased responsiveness and greater accountability to local citizens. Between-district competition is thought to drive higher performance as local leaders attempt to generate satisfaction among, attract, and retain residents (e.g., Hayek, 1939; Tiebout, 1956). Goldin and Katz (2008) argue that, historically, the existence of numerous, small, fiscally independent districts in the U.S. helped pave the way for the high school expansion movement in the 20th Century. Despite a lack of nationwide majority support for public school expansion at the time, secondary school enrollment rose as residents with an appetite for public high schools selected to live in localities providing those services and nearby districts were pressured to also expand their school systems to compete.

Notions of competition and sorting across districts also play out in the subfield of labor economics, where labor market theory identifies districts as key agents in the human capital pipeline. District inputs related to recruitment and retention of teachers can serve 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. To build coherent instructional systems, districts need to identify, recruit, hire, and retain teaching talent. At the same time, the labor economics literature focuses much more on teachers' preferences rather than their day-to-day work with students. For example, empirical research from this tradition indicates that teachers' decisions about where to teach are influenced by contextual factors at the district level, including geography or proximity (Boyd et al., 2005) 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 focuses 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 nonrandom 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 examines the efficacy of district-wide turnaround reforms that often include but are not limited to human resource policies. 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 the instructional resources that come with these changes. To do so, researchers compare 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 more often use single case studies without formal comparison groups to generate theory on district effects. Further, when researchers speculate about sources of turnaround success, they often focus 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 Inter-Disciplinary Synthesis

Our review of these three theoretical literatures identifies multiple ways district-level policymaking *is hypothesized* to benefit district-wide performance, while also identifying key gaps in our knowledge base. As noted above, prior reviews cover the organizational behavior and educational leadership research traditions, and focus on the ways 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 educational 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 review often suffer from the use of "outlier" designs. That is, districts are 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 like economics typically are more attuned to drawing causal conclusions. At the same, this approach has required economists 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 Review of the Quantitative Empirical Evidence

To fill gaps in the current literature on district effectiveness, we next describe the methods we use 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

We restrict the set of quantitative empirical studies for our review in three primary ways, related to the: (i) level of variation of the policy input, (ii) outcome measure(s) examined, and (iii) research design. We describe these specific criteria and our motivation for these criteria next. In addition, we include only studies conducted in the U.S. given the level of educational (de)centralization varies quite a bit from country to country, as does the definition of what constitutes a school district and its relationship to the other levels of government. We do not impose any restrictions on publication type. Neither do we place a lower bound on publication date, though we do limit 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. In some instances, multiple articles are published that examine effects of the same district-level policy on the same sample of individuals. In these instances, we only include multiple articles if the outcome measures are different or are captured at different points in time.

District-Level Variation in Policy Inputs

First, we include studies where policy inputs under investigation vary at the district level, as opposed to other levels and sources of variation such as states, schools, classrooms, teachers, or students. We make this restriction given our aim to identify effective district-level policies that are implemented and are theorized to have impacts district wide. While some policy inputs might be implemented and vary at different levels of the education system, it is district-level and district-wide implementation that is the focus of this study. A second, practical reason for this inclusion criteria is that inclusion of every study of a policy that *could potentially be implemented and studied* at the district level—but where authors exploit variation at the state, school, classroom, teacher, or student level—would explode our sample and make the task unmanageable.

This restriction means, for example, that we exclude studies of school closure that use student- or school-level variation only. Knowing the effect of school closure on the outcomes of students who attend 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. We also exclude studies of state takeover of districts, where the primary level of implementation and source of variation is the state rather than the district. 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.) This restriction may mean that our review does not capture certain effective polices or programs districts could implement that have not yet been studied at the district level.

Focus on Student or Teacher Outcomes

Second, we focus only on studies that quantitatively examine 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 prioritize studies that examine links between district-level policy inputs and varied student outcomes: test-based achievement, course grades, school behaviors (e.g., absences, suspensions), grade retention, high school graduation, and college-going.

We also include studies that examine 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 are shown to impact student outcomes including classroom experience (Papay & Kraft, 2015) and turnover (Ronfeldt et al., 2013). We also include studies that use 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

Finally, we require that included studies attempt, 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. Additionally, researchers should be concerned about isolating the causal effect of a particular policy reform from other contemporaneous policy changes or temporal trends. We decided on a moderate to low bar for supporting causal claims, as very few studies account fully for nonrandom selection of district-level policies using experimental designs. Further, if we focused only on "quasi-experiments" (i.e., regression discontinuity and difference-in-differences), we would be left with studies of district-level inputs generally of interest to the economists historically most likely to wield these analytic tools. Therefore, we require that studies have a comparison group of treated and non-treated districts and control in some way for nonrandom selection of district-level policies. In other words, studies could not be purely correlational nor have simple pre/post designs, both thought to fall well short of allowing researchers to draw causal claims (Murnane & Willett, 2010).

Because we set a moderate to low bar for supporting causal claims, we then group studies into three 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, under certain conditions and assumptions: 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 experience a policy shock, relative to changes in outcomes for comparison units (districts) that do not experience this shock (Murnane & Willett, 2010). We do not categorize studies as difference-in-differences designs if they examine how changes in a policy input related to changes in outcomes absent a policy shock that is arguably exogenous.

A primary assumption of these research designs—and difference-in-differences designs in particular—is that the policy is uniquely isolated from other contemporaneous shocks. However, in our review, we identify and include many studies that apply difference-in-differences designs where the policy shock is a district-wide turnaround that includes a package of reforms. We separate these studies into a second group, which we refer to as "causally oriented" but where the effects of multiple policy inputs cannot be uniquely isolated. The third broad research design category includes all observational studies, which we define as those studies that attempt in some way to limit bias due to nonrandom selection but generally do not capitalize on exogenous variation and therefore fall short of supporting causal claims on their own. This set of designs includes multivariate regression, propensity score matching, 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 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 meet our inclusion criteria. A full list of search terms is provided in the Appendix. Second, to cross-check our search process, we reviewed the reference lists from prior reviews (e.g., Anderson & Young, 2018; Leithwood, 2010; Trujillo, 2013) and from all studies that meet 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. This generated a list of over 300,000 articles that we screened. From that pool, we reviewed 312 articles against our inclusion criteria and ultimately included 99 studies in our analytic sample.

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 2–4). 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.[1] 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 Review of the Quantitative Empirical Evidence

Characteristics of Included Studies

Our search identified a total of 99 quantitative studies that meet our inclusion criteria. In Table 1, we describe that the majority of studies in our sample (63%) come from peer-reviewed journals, though the second largest category is unpublished dissertations (24%), all but one of which examined superintendent characteristics; the remaining studies come from book chapters and reports.

In terms of research methodologies, observational studies (i.e., non-causal but not purely correlational) are by far the most common (73%). Of research designs that meet a higher bar for supporting causal inferences, difference-in-differences design in the presence of a policy shock is 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. At the same time, roughly half of these difference-in-differences studies examine the effects of a package of reforms in the context of district-wide turnarounds, where the unique effect of a given policy input cannot be teased out. Only three articles use experimental designs, two of which rely on the same randomized experiment, assessing outcomes at different points in time. This again makes sense given the challenges of randomizing interventions at the district level. Another three studies use regression discontinuity designs, all of which examine the effects of expanding district facilities as a result of narrowly winning versus losing a district-wide capital campaign vote. Most studies (79%) focus on one policy input. Some studies focus on more, with 9% focusing on four or more inputs; most studies that include multiple policy inputs are analyses of district-wide turnarounds rolled out as a package of reforms. The vast majority of included studies (87%) examine student test-score outcomes, while far fewer examine relationships between district-level inputs and other student outcomes (i.e., course grades, absences,

[1]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 reveals a small number of studies. In many cases, these studies rely on different dependent variables, making meta-analytic techniques inappropriate. 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.

suspensions, grade retention, high school graduation; 9%) or teacher-level measures (14%). Just four studies look at links between district inputs and both student and teacher outcomes.

Synthesis of Study Findings by Policy Category and Type

In Tables 2–4, we summarize findings across studies within each of the 26 total policy input categories, which we generate from our theoretical review of the three major disciplinary perspectives on the topic of district effectiveness: Table 2 covers "governance and politics"-related inputs theorized primarily by political science traditions; Table 3 covers inputs related to the "recruitment and retention" of teacher talent stemming primarily from labor market theory; and Table 4 covers "instructional resources and development"-related inputs from theoretical frameworks on educational leadership and organizational behavior traditions. 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 are not strict. Scholars from a variety of disciplines, including interdisciplinary teams, examine 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 (i) most likely to credibly support causal conclusions of the unique effect of a given policy input (superscript bolded), (ii) causally-oriented studies that examine the effect of a package of reforms and so are not able to isolate the unique effect of a given policy input (superscript italicized), and (iii) observational studies. We shade cells to illustrate the broad patterns that emerge (i.e., positive relationships in light gray, null or mixed findings in dark gray, negative relationships in black). While we do not include the magnitudes of effects in these tables, we discuss these findings in the narrative text, focusing primarily on the magnitudes of estimates generated from causal analyses.

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: 58 studies focus on one of the 11 policy inputs under the governance and politics umbrella; 30 studies across 11 policy inputs related to instructional resources and development-related inputs; and 23 studies focused on the four recruitment and retention-related inputs. 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, where a high bar for causal inference in economics may result in lower coverage of policy inputs typically studied by scholars in this tradition.

Policies Related to Governance and Politics

In Table 2, we summarize findings from the studies examining governance and politics-related inputs. In total, our theoretical review reveals 11 policy inputs under this category, all but two of which have coverage in the empirical literature. The most common input studied under this umbrella is superintendent characteristics (32 studies), with union strength (15 studies), district leaders appointed versus elected (10 studies), portfolio management models (seven studies), school autonomy (seven studies), and school board characteristics (six studies) also having decent coverage. Among the most highly studied categories, findings are mixed and research designs often are not able to support causal conclusions.

Superintendent characteristics and approaches to system structures and management is the most studied policy input category across our entire review. At the same time, all 32 studies are observational in nature, including a large number of studies from the 1980s and 1990s before the causal revolution in education research, as well as more recent analyses. 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**, which includes boards' approaches to engaging diverse stakeholders, 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). Findings also are mixed for **appointment process of district leaders**, though there is one study that can credibly isolate the unique effect of this policy input. Using a difference-in-differences design that exploits changes in appointment processes, Partridge and Sass (2011) estimate inconsistent effects on student test scores for district leaders appointed versus elected: appointed superintendents have no effect on math scores, a 0.07 SD increase in short-run reading scores, but a negative effect on longer-run reading scores. In district turnaround contexts that examine a package of reforms, more **district administrators** is associated with lower student achievement; in other observational studies on this topic, there is no relationship to student outcomes.

When it comes to **unions**, we identify three district-level studies that can credibly isolate the unique effect of this policy input, all of which use difference-in-difference designs. However, findings are mixed across studies and outcomes. Exploiting the timing of collective bargaining agreements across districts, Hoxby (1996) finds negative effects of unions on high-school graduation rates of roughly 2 percentage points. At the same time, Lovenheim (2009) finds no effect on high-school dropout using a similar strategy, but does find positive effects of unions on teacher employment. Potentially helping to explain this inconsistency, Figlio (2002) finds null effects of unionization on the characteristics and qualifications of incoming teachers (e.g., SAT scores, majoring in the subject she/he teachers). Findings from four district-wide turnaround studies, where teachers' unions roles were greatly reduced or eliminated, generally show positive effects on student test-score performance (Harris & Larsen, 2016; Pham et al., 2020; Schueler et al., 2017).

Four policy inputs under the governance and politics umbrella are examined primarily or exclusively in the context of district-wide turnaround efforts: **portfolio management models**, **school autonomy**, **school choice**, and **school closures**. Because these policy inputs often are explored as a group—and generally include additional policy inputs related to appointment processes of school district leaders, unions, hiring of personnel, and instructional resources—we cannot tease out their independent contributions to educational outcomes, even with the use of difference-in-differences strategies. Nonetheless, we infer several patterns when looking across turnaround cases. First, portfolio management, as a primary feature of the turnaround model, does not appear to be associated with improved student outcomes in Philadelphia, relative to comparison districts across Pennsylvania (Gill et al., 2007) or in the Achievement School District in Tennessee that consolidated low-performing schools in Memphis, Nashville, Chattanooga, and Knoxville (Pham et al., 2020).

Second, new governance arrangements that facilitate innovation often return more positive results. For example, in Tennessee, an additional set of low-performing schools organized into a new Innovation Zone characterized by school autonomy saw very large test-score gains in reading, math, and science between 0.15 and 0.25 SD (Pham et al., 2020). Schueler et al (2017) find similar positive effects in math—but not reading—of turnaround in Lawrence, Massachusetts with some similar policy features. On the flipside, when New York City pulled back autonomy from low-performing schools and created a newly constituted Chancellor's District, effects on student performance were small or null (Phenix et al., 2005); this suggests that reducing autonomy is unsuccessful and so is consistent with other studies on this topic.

New Orleans arguably is the most intensive case of a market-based school accountability system, which includes portfolio management, school autonomy, school choice, as well as four other policy inputs related to unions and personnel. These combined efforts resulted in positive effects on test scores upwards of 0.4 SD, high school graduation of 3–9 percentage points, and college attendance of 8–15 percentage points, compared to a matched comparison group of Louisiana school districts that did not experience these policy changes (Harris & Larsen, 2016). At the same time, Lincove et al (2018) find that turnaround in New Orleans is associated with higher teacher turnover relative to other districts that suffered hurricane damage. Finally, while turnaround in Newark, New Jersey includes similar governance-altering strategies as other turnaround cases, analyses suggests that a primary feature of its success is shifting enrollment to higher-performing schools (Chin et al., 2019). That said, the Newark effects on reading performance of 0.07 SD are smaller than in Tennessee, New Orleans, and Lawrence, and there are null effects on math performance. Because these successful district turnaround cases also include policy inputs related to hiring of personnel and instructional resources—topics we turn to below—it may be that new governance arrangements create conditions for other types of innovations that, in turn, improve educational outcomes.

Our theoretical review identifies two additional policy inputs under the governance and politics umbrella but that did not have any coverage in the empirical literature that meet our inclusion criteria: **school admissions** (e.g., centralized lotteries) and **school size** (e.g., consolidating schools, subdividing buildings into small schools).

Policies Related to Retention and Recruitment of Teacher Talent

In Table 3, we summarize results from studies examining inputs related to recruitment and retention of teachers, typically emphasized by labor economists. The most common input under this umbrella is teacher salaries and benefits (19 studies) with the other categories having less coverage: teacher hiring practices (five 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 within the set of studies most likely to support causal conclusions that all use difference-in-differences design to exploit a stand-alone salary change across districts and over time (e.g., Biasi, 2021; Figlio, 2002; Hendricks, 2014, 2015), as well as difference-in-differences analysis of district-wide turnaround where changes in salary and benefits accompanied other policy inputs (e.g., Chin et al., 2019; Harris & Larsen, 2016; Schueler et al., 2017) and a set of observational studies (e.g., Booker & Glazerman, 2009; Houck et al., 2010). For example, leveraging panel data from Texas, Hendricks finds that a 1% increase in teacher pay results in a 0.04 to 0.08 percentage point increase in the proposition of targeted teachers hired and a 0.16 percentage point decrease in teacher turnover. 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. For example, Biasi (2021) finds that student math scores increase by 0.02 SD in districts that change from seniority to flexible pay schemes, relative to districts that maintain seniority pay over time. Similarly, Sojourner et al. (2014) find that changing to a pay-for-performance salary scheme results in increases district-wide reading achievement by 0.03 SD.

In turnaround settings, district-level **teacher hiring practices**—namely, widespread replacement of teachers—also is associated with higher student achievement (e.g., Pham et al., 2020). 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 use observational research designs. We do not identify any studies that isolate the causal effect of novel hiring practices separated from district turnaround reform. There is mixed evidence of the effect of **teacher seniority protections**. Both using difference-in-differences analysis that exploit cross-district and cross-time changes in seniority protections within collective bargaining agreements, Koski and Horng (2007) and Goldhaber et al. (2016) find positive effects on the percentage of credentialed teachers and greater within-district retention of effective teachers in disadvantaged schools. However, there is no effect on between-district teacher mobility.

Finally, our search did not uncover any studies that isolate the unique effect of various features of **teacher evaluation systems** at the district level, though a few studies examine district-wide reforms that include changes to teacher evaluation as part of the reform bundle (Chin et al., 2019; Pham et al., 2020; Schueler et al., 2017). In all three cases, studies use difference-in-difference designs to examine the effect of a reform package that includes but is not limited to changes in evaluation processes. While all three studies find positive effects, we cannot state with certainty that the teacher evaluation component is the key driver.

Policies Related to Instructional Resources and Development

Table 4 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 (nine studies), followed by curriculum materials (eight studies), student-teacher ratios (a proxy for class size; seven studies), and wellbeing supports (seven studies). Patterns of findings from these studies often are positive and there is decent coverage of studies likely to support causal conclusions, including the only experimental and regression discontinuity designs in our sample.

All three experimental studies examine the effect of district-wide **use of data** on student or teacher performance. These interventions aim to standardize student assessment and assignment practices, build processes for collecting these data from and then sharing them back with schools, and providing supports on how to interpret these data to guide instruction. 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 upwards of 0.3 SD, depending on the grade level and subject area, with the largest effects coming several years after the start of the widespread data use. In the May and Robinson (2007) study, positive effects are concentrated amongst students who re-take exams, as the theory of action suggests: widespread use of data can be used for progress monitoring and for re-allocating resources for students who require additional supports. 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). At the same time, one observational study finds a negative relationship between the extent of data use in a district and a measure of high-quality instruction (Lee et al., 2012). The data-use interventions examined in our sample often include 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, at least in the context of data use.

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., 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. Turnaround contexts that emphasize teacher professional development return small positive results in some instances (Chin et al., 2019; Phenix et al., 2005) but null effects in others (Gill et al., 2007). Two observational studies find that more teacher professional development is associated with improved instructional practices and student achievement (Jacques & Brorsen, 2002; Lee et al., 2012). 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). However, observational studies that focus more narrowly on individual curricula are mixed. In the most recent analysis covering the largest set of textbooks across six U.S. states, new adoptions of high-quality materials in math is not associated with student achievement gains (Blazar et al., 2020).

We create a separate category for **coherence of instructional resources**, though note that coherence generally refers to the alignment of 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), finding a small positive effect in math but not in reading. To our knowledge, only one study directly measures the degree of instructional coherence across school systems, finding a positive association to student achievement gains (Leithwood & Azah, 2017); however, we exclude this study from our empirical review because it examines school systems in Canada. The literature on **student-teacher ratios** is mixed. Exploiting a policy shock in school district funding that immediately impacted personnel hiring, Mensah et al. (2013) find some positive effects of lower student-teacher ratios on student achievement; however, findings are not always consistent across models. Five observational studies on this topic return consistently positive associations.

We define a relatively broad policy input category called **wellbeing supports** that consider the ways that district-level policies and practices can improve health and safety, with possible trickle-down effects on academic performance. Observational studies that link a broad measure of the number of support staff (e.g., counselors, psychologists, aides) to student outcomes come to mixed conclusions. However, Gandhi et al. (2018) find positive effects on math and reading achievement between 0.15 and 0.24 SD of targeted wraparound services focused on family engagement and community partnerships (as well as curriculum, professional development, and data use). Ruffini (2022) exploits that staggered rollout across school districts of the Community Eligibility Provision to offer free meals to all students, finding no effect on district-wide math or reading performance. In a similar difference-in-differences setup, Lacoe and Steinberg (2018) find no effect on student suspensions of rolling back a zero-tolerance discipline policy; there are some negative effects on attendance and math achievement.

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 include increased learning time, generally finding positive effects of the package of reforms (e.g., Chin et al., 2019; Schueler et al., 2017). One observational study (Hinrichs, 2011) finds 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 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). One study on **school-grade configuration** that uses a difference-in-differences design finds negative effects of attending middle school in sixth grade, rather than staying in elementary school (Cook et al., 2008). Finally, one district turnaround study that includes district-wide purchasing

of **technology** finds positive effects on student outcomes, but only in the context of a package of reforms (Chin et al., 2019).

Discussion

Synthesizing Findings across Disciplines and Policy Inputs

Consistent with our review of the theoretical literature, our interdisciplinary review of the quantitative empirical research reveals varied ways in which school districts are likely to influence student outcomes. We may be limited in our ability to identify promising district policies that have not yet been the subject of empirical research. That said, 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 result in higher teacher retention and improved student test scores. Instructional resources including district-wide policies on data use paired with principal professional development on how to use these data to guide improvement, as well as extra learning time—generally district levers that have been the focus of scholars from organizational behavior and educational leadership—also lead to improvements in educational outcomes. Though findings are less consistent from the governance and politics literature, studies of district-wide turnarounds suggest that the combination of school autonomy, school choice, and school closures (alongside other policy inputs) can improve 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 educational 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.

Similarly, another recent meta-analysis of turnaround evaluations provides suggestive evidence that efforts to improve low-performing schools have been more effective when implemented at the district than at the school level (Schueler et al., 2022). Future research may be designed specifically around questions of efficiency and scalability by more systematically comparing the effects of similar policy inputs implemented at the district versus school (or even state, classroom, or teacher) levels.

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 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 (27%) employed causal research designs, with a disproportionate number applied to a small number of policy inputs (i.e., teachers' unions, teacher salaries and benefits, teacher seniority protections, data use, facilities). Roughly half of the studies using difference-in-differences designs examine the effects of a package of reforms that include between two and 12 different policy inputs. Thus, it is difficult to isolate the unique 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 (73%) sufficiently minimize threats to internal validity. Studies without an experimental or quasi-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?

Under-Studied Policy Inputs

A second gap in the literature relates to under-studied inputs. While all but two of the district-level policy categories that surfaced from our theoretical review have some coverage in the empirical, quantitative literature, in many instances there are only a handful of studies and often only observational ones. Understudied policy areas—at least those that exploit district-level implementation and district-wide performance outcomes—include teacher evaluation systems, school size, selective admissions, school closures, technology, and school-grade configurations, which 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) uses a survey to quantitatively predict variation in performance across school systems in Canada. Another study that fits our inclusion criteria (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 is 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 observe 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 Causal 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 may 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 establish 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 quasi-experimental designs. Regression discontinuity designs exploiting thresholds in district-wide votes represent a promising approach. Difference-in-differences designs that leverage policy shocks 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 (though only a limited set), 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. Further, 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).

One caveat with this recommendation is that use of difference-in-differences designs to estimate the causal effect of policy inputs are best employed in the context of an exogenous shock, such as a new law that allows districts to develop new compensation schemes for teachers (Biasi, 2021), unionization of teachers in some school districts but not others (Lovenheim, 2009), a new wraparound services initiative that affects only high-poverty districts (Gandhi et al., 2018), etc. Simply looking at how changes in a given policy input relate to changes in a given educational outcome is unlikely to support causal claims. However, some policy inputs—particularly some under the instructional resources category such as instructional coherence—may not flow from a top-down policy shock because they are, instead, ground-up practices. Estimating the causal effect of these sorts of district-level policy inputs on district-wide performance may be a focus of future experimental analyses.

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 salaries, data use, extra learning time, school autonomy and parental choice in the context of district turnaround) 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.

Open Research Statements

Study and Analysis Plan Registration

There is no study and analysis plan registration associated with this manuscript.

Data, Code, and Materials Transparency

The data, code, and materials underlying the results reported in this manuscript are not publicly available.

Design and Analysis Reporting Guidelines

Not applicable.

Transparency Declaration

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Replication Statement

This manuscript reports an original study.

Disclosure Statement

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

References

*studies that are included in the formal research synthesis

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

*Alexander, K. L., & Griffin, L. J. (1976). School district effects on academic achievement: A reconsideration. *American Sociological Review*, *41*(1), 144–152. https://doi.org/10.2307/2094378

Anderson, E., & Young, M. D. (2018). 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. https://doi.org/10.3389/feduc.2018.00087

*Balter, D., & Duncombe, W. D. (2008). Recruiting highly qualified teachers: Do district recruitment practices matter. ?*Public Finance Review*, *36*(1), 33–62. https://doi.org/10.1177/1091142106293949

*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. https://doi.org/10.1016/j.jue.2017.04.003

*Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness. *Educational Evaluation and Policy Analysis*, *34*(4), 391–412. https://doi.org/10.3102/0162373712440040

*Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review*, *34*, 107–121. https://doi.org/10.1016/j.econedurev.2013.01.014

*Biasi, B. (2021). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, *13*(3), 63–102. https://doi.org/10.1257/pol.20200295

*Bidwell, C. E., & Kasarda, J. D. (1975). School district organization and student achievement. *American Sociological Review*, *40*(1), 55–70. https://doi.org/10.2307/2094447

*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. https://doi.org/10.3102/0002831216630407

*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. https://doi.org/10.1177/0013161X10375271

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*,

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The draw of home. *Journal of Policy Analysis and Management*, *24*(1), 113–132. https://doi.org/10.1002/pam.20072

*Brewer, D. J. (1996). Does more school district administration lower educational productivity? *Economics of Education Review*, *15*(2), 111–124. https://doi.org/10.1016/0272-7757(95)00032-1

Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan*, *91*(7), 23–30. https://doi.org/10.1177/003172171009100705

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. https://doi.org/10.3102/0162373711412765

*Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments. *Quarterly Journal of Economics*, *125*(1), 215–261. https://doi.org/10.1162/qjec.2010.125.1.215

Childress, S., Elmore, R., & Grossman, A. (2006). How to manage urban school districts. *Harvard Business Review*, *84*(11), 55.

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. https://doi.org/10.1177/0019793918774318

Chingos, M. M., Whitehurst, G. J., & Gallaher, M. R. (2015). School districts and student achievement. *Education Finance and Policy*, *10*(3), 378–398. https://doi.org/10.1162/EDFP_a_00167

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. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy*, *6*(3), 399–438. https://doi.org/10.1162/EDFP_a_00040

*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. https://doi.org/10.3102/0162373713482765

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. https://doi.org/10.1002/pam.20309

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.

DiMaggio & Powell. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160. https://doi.org/10.2307/2095101

*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 Books.

*Eberts, R. W., & Stone, J. A. (1987). Teacher unions and the productivity of public schools. *ILR Review*, *40*(3), 354–363. https://doi.org/10.1177/001979398704000303

*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. https://doi.org/10.17763/haer.66.1.g73266758j348t33

Elmore, R., & Burney, D. (1998). Continuous improvement in Community District #2.

*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. https://doi.org/10.1016/j.compedu.2016.12.001

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. https://doi.org/10.3102/0013189X18759524

*Figlio, D. (2002). Can public schools buy better-qualified teachers? *ILR Review*, *55*(4), 686–699. https://doi.org/10.2307/3270629

Firestone, W. A., Mangin, M. M., Martinez, M. C., & Polovsky, T. (2005). Leading coherent professional development. *Educational Administration Quarterly*, *41*(3), 413–448. https://doi.org/10.1177/0013161X04269605

*Ford, M. (2013). *The impact of school board governance on academic achievement in diverse states* [Unpublished doctoral dissertation]. University of Wisconsin.

*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. https://doi.org/10.1080/01900692.2014.982293

*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. https://doi.org/10.1080/12294659.2016.1266181

Fryer, R. G. Jr, (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments* (Vol. 2, pp. 95–322). North-Holland.

*Gandhi, A. G., Slama, R., Park, S. J., Russo, P., Winner, K., Bzura, R., Jones, W., & 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. https://doi.org/10.1080/19345747.2017.1413691

*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.

Goldin, C., & Katz, L. (2008). *The race between education and technology*. Harvard University Press.

*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. https://doi.org/10.1002/pam.21914

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. https://doi.org/10.3102/0002831217733445

*Goodman, G. S., & Young, I. P. (2006). The value of extracurricular support in increased student achievement. *Educational Research Quarterly*, *30*(1), 3–13.

*Grissom, J. A. (2014). Is discord detrimental? *Journal of Public Administration Research and Theory*, *24*(2), 289–315. https://doi.org/10.1093/jopart/mus042

Grissom, J. A., & Strunk, K. O. (2012). How should school districts shape teacher salary schedules? *Educational Policy*, *26*(5), 663–695. https://doi.org/10.1177/0895904811417583

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

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, *14*(3), 351–388. https://doi.org/10.2307/145575

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, *113*(485), F64–F98. https://doi.org/10.1111/1468-0297.00099

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *The Journal of Human Resources*, *39*(2), 326–354. https://doi.org/10.2307/3559017

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*(2), 267–271. https://doi.org/10.1257/aer.100.2.267

*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. https://doi.org/10.1108/eb009926

*Hart, W., Rock Hill, S., Schramm-Possinger, M., & Hoyle, S. (2019). Superintendent longevity and student achievement in North Carolina Public Schools. *AASA Journal of Scholarship and Practice, 15*(4), 4–13.

Hayek, F. A. (1939). The economic conditions of interstate federalism. *New Commonwealth Quarterly*, *5*(2), 131–149.

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. https://doi.org/10.1177/0193841X14529126

*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. https://doi.org/10.1080/00036846.2016.1213363

*Hendricks, M. D. (2014). Does it pay to pay teachers more? Evidence from Texas. *Journal of Public Economics*, *109*, 50–63. https://doi.org/10.1016/j.jpubeco.2013.11.001

*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. https://doi.org/10.1016/j.econedurev.2015.05.008

Henig, J. (2009). Mayors, governors, and presidents: The new education executives and the end of educational exceptionalism. *Peabody Journal of Education*, *84*(3), 283–299. https://doi.org/10.1080/01619560902973449

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

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), 233285841561770. https://doi.org/10.1177/2332858415617703

*Hinrichs, P. (2011). When the bell tolls: The effects of school starting times on academic achievement. *Education Finance and Policy*, *6*(4), 486–507. https://doi.org/10.1162/EDFP_a_00045

Ho, A. D. (2008). The problem with "proficiency": Limitations of statistics and policy under No Child Left Behind. *Educational Researcher*, *37*(6), 351–360. https://doi.org/10.3102/0013189X08323842

Honig, M. (2012). District central office leadership as teaching. *Educational Administration Quarterly*, *48*(4), 733–774. https://doi.org/10.1177/0013161X12443258

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. https://doi.org/10.1177/0895904807307067

Honig, M., & Hatch, T. (2004). Crafting coherence. *Educational Researcher*, *33*(8), 16–30. https://doi.org/10.3102/0013189X033008016

*Hoover, G. A. (2008). Elected versus appointed school district officials: Is there a difference in student outcomes? *Public Finance Review*, *36*(5), 635–647. https://doi.org/10.1177/1091142107311219

*Houck, E. A., Rolle, R. A., & He, J. (2010). Examining school district efficiency in Georgia. *Journal of Education Finance*, *35*(4), 331–357. https://doi.org/10.1353/jef.0.0022

*Hough, K. L. (2014). Internal accountability and district achievement: How superintendents affect student learning. *Journal of School Leadership*, *24*(1), 32–60. https://doi.org/10.1177/105268461402400102

*Hoxby, C. M. (1996). How teachers' unions affect education production. *The Quarterly Journal of Economics*, *111*(3), 671–718. https://doi.org/10.2307/2946669

*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. (2020). Does school spending matter? The new literature on an old question. In L. Tach, R. Dunifon, & D. L. Miller (Eds.), *Confronting inequality: How policies and practices shape children's opportunities* (pp. 165–186). American Psychological Association.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108. https://doi.org/10.1257/app.1.4.85

Jacobsen, R., & Saultz, A. (2012). The polls—Trends: Who should control education? *Public Opinion Quarterly*, *76*(2), 379–390. https://doi.org/10.1093/poq/nfs017

*Jacques, C., & Brorsen, B. W. (2002). Relationship between types of school district expenditures and student performance. *Applied Economics Letters*, *9*(15), 997–1002. https://doi.org/10.1080/13504850210148161

*Johnson, K. (1997). *The relationship of superintendent tenure to school performance in Arkansas* [Unpublished doctoral dissertation]. University of Arkansas.

Johnson, R. C., & Jackson, C. K. (2019). Reducing inequality through dynamic complementarity. *American Economic Journal: Economic Policy*, *11*(4), 310–349. https://doi.org/10.1257/pol.20180510

Johnson, S. M., Marietta, G., Higgins, M., Mapp, K., & Grossman, A. (2015). *Achieving coherence in district improvement*. Harvard Education Press.

*Kai, H., & Zimmer, R. (2016). Does investing in school capital infrastructure improve student achievement? *Economics of Education Review*, *53*, 143–158. https://doi.org/10.1016/j.econedurev.2016.05.007

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), 233285841769051. https://doi.org/10.1177/2332858417690511

*Koski, W. S., & Horng, E. L. (2007). Facilitating the teacher quality gap? *Education Finance and Policy*, *2*(3), 262–300. https://doi.org/10.1162/edfp.2007.2.3.262

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*(sup1), 188–208. https://doi.org/10.1080/19345747.2015.1116031

*Lacoe, J., & Steinberg, M. P. (2018). Rolling back zero tolerance: The effect of discipline policy reform on suspension usage and student outcomes. *Peabody Journal of Education*, *93*(2), 207–227. https://doi.org/10.1080/0161956X.2018.1435047

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. https://doi.org/10.3102/01623737024001037

*Lee, M., Seashore Louis, K., & Anderson, S. (2012). Local education authorities and student learning. *School Effectiveness and School Improvement*, *23*(2), 133–158. https://doi.org/10.1080/09243453.2011.652125

Leithwood, K. (2010). A review of evidence about the characteristics of high performing school districts. *Leadership and Policy in Schools*, *9*(3), 245–291. https://doi.org/10.1080/15700761003731500

*Leithwood, K., & Azah, V. N. (2017). Characteristics of high-performing school districts. *Leadership and Policy in Schools*, *16*(1), 27–53. https://doi.org/10.1080/15700763.2016.1197282

*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. https://doi.org/10.1080/13504850500397635

*Lin, T. C. (2010). Teacher salaries and student achievement: The case of Pennsylvania. *Applied Economics Letters*, *17*(6), 547–550. https://doi.org/10.1080/13504850802167223

*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. https://doi.org/10.3102/0013189X18759542

Little, J. W. (1989). District policy choices and teacher's professional development opportunities. *Educational Evaluation and Policy Analysis*, *11*(2), 165–179. https://doi.org/10.3102/01623737011002165

*Lovenheim, M. F. (2009). The effect of teachers' unions on education production. *Journal of Labor Economics*, *27*(4), 525–587. https://doi.org/10.1086/605653

*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. https://doi.org/10.1016/j.jpubeco.2016.05.002

*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. https://doi.org/10.1086/226550

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. https://doi.org/10.1016/j.econedurev.2013.01.005

*Moe, T. M. (2009). Collective bargaining and the performance of the public schools. *American Journal of Political Science*, *53*(1), 156–174. https://doi.org/10.1111/j.1540-5907.2008.00363.x

Moe, T. (2011). *Special interest: Teachers unions and America's public schools*. 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. https://doi.org/10.3102/01623737011001031

*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.

*Murnane, R. J., & Olsen, R. J. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *The Journal of Human Resources*, *25*(1), 106–124. https://doi.org/10.2307/145729

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. https://doi.org/10.1080/00220671.1988.10885819

*Myers, S. (2011). Superintendent length of tenure and student achievement. *Administrative Issues Journal*, *1*(2), 6. https://doi.org/10.5929/2011.1.2.4

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence. *Educational Evaluation and Policy Analysis*, *23*(4), 297–321. https://doi.org/10.3102/01623737023004297

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. https://doi.org/10.1086/521242

Orton, J. D., & Weick, K. E. (1990). Loosely coupled systems: A reconceptualization. *The Academy of Management Review*, *15*(2), 203–223. https://doi.org/10.2307/258154

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market. *Journal of Public Economics*, *130*, 105–119. https://doi.org/10.1016/j.jpubeco.2015.02.008

*Parker-Chenaille, R. (2012). *The impact of superintendent turnover on student achievement in rural districts* [Unpublished doctoral dissertation].

*Partridge, M., & Sass, T. R. (2011). The productivity of elected and appointed officials: the case of school superintendents. *Public Choice*, *149*(1-2), 133–149. https://doi.org/10.1007/s11127-011-9832-9

Peterson, P. E. (2016). James S. Coleman: Education's North Star. *Education Next*, *16*(2), 5–6.

*Pham, L. D., Henry, G. T., Kho, A., & Zimmer, R. (2020). Sustainability and maturation of school turnaround: A multiyear evaluation of Tennessee's achievement school district and local

innovation zones. *AERA Open*, *6*(2), 233285842092284. https://doi.org/10.1177/2332858420922841

*Phenix, D., Siegel, D., Zaltsman, A., & Fruchter, N. (2005). A forced march for failing schools: Lessons from the New York City Chancellor's District. *Education Policy Analysis Archives*, *13*, 1–24. https://doi.org/10.14507/epaa.v13n40.2005

*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), 1–11.

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. *The Journal of Educational Research*, *109*(5), 491–502. https://doi.org/10.1080/00220671.2014.992581

Reform Support Network. (n.d.). Evaluations of Teacher Effectiveness: State Requirements for Classroom Observations. https://files.eric.ed.gov/fulltext/ED611524.pdf

*Register, C. A., & Grimes, P. W. (1991). Collective bargaining, teachers, and student achievement. *Journal of Labor Research*, *12*(2), 99–109. https://doi.org/10.1007/BF02685375

*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*, *34*(1), 74–95. https://doi.org/10.2307/30034700

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475–514. https://doi.org/10.3102/0002831215585562

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36. https://doi.org/10.3102/0002831212463813

Rorrer, A. K., Skrla, L., & Scheurich, J. J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly*, *44*(3), 307–357. https://doi.org/10.1177/0013161X08318962

*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. https://doi.org/10.1016/j.jue.2010.01.001

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, *3*(2), 135–146. https://doi.org/10.1093/oxfordjournals.oep.a041827

*Rudolph, B. R. (2016). *The Role of Superintendent Longevity in Predicting 2014-2016 APR Scores for Missouri School Districts*. Baker University.

*Ruffini, K. (2022). Universal Access to Free School Meals and Student Achievement Evidence from the Community Eligibility Provision. *Journal of Human Resources*, *57*(3), 776–820. https://doi.org/10.3368/jhr.57.3.0518-9509R3

*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. https://doi.org/10.1080/15700760903511780

*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. https://doi.org/10.3102/0162373716685824

Schueler, B. E., Asher, C., Larned, K., Mehrotra, S., & Pollard, C. (2022). Improving low-performing schools: A meta-analysis of impact evaluation studies. *American Educational Research Journal*, *59*(5), 975–1010. https://doi.org/10.3102/00028312211060855

*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), 10–23.

*Sims, D. P. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, *27*(1), 58–68. https://doi.org/10.1016/j.econedurev.2006.05.003

*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. https://doi.org/10.3102/0002831212466909

*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. https://doi.org/10.1353/jhr.2014.0029

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, *18*(2), 141–179. https://doi.org/10.1207/S1532690XCI1802_01

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity. *Educational Evaluation and Policy Analysis*, *19*(2), 185–203. https://doi.org/10.2307/1164209

*Strunk, K. O. (2011). Are teachers' unions really to blame? *Education Finance and Policy*, *6*(3), 354–398. https://doi.org/10.1162/EDFP_a_00039

*Strunk, K. O., & McEachin, A. (2011). Accountability under constraint. *American Educational Research Journal*, *48*(4), 871–903. https://doi.org/10.3102/0002831211401006

*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. https://doi.org/10.1002/pam.21658

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. https://doi.org/10.1016/0272-7757(90)90005-P

*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. https://doi.org/10.1016/0272-7757(95)00022-4

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

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, *64*(5), 416–424. https://doi.org/10.1086/257839

*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. https://doi.org/10.1108/09578231311325640

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. https://doi.org/10.1177/0022487117704170

*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. https://doi.org/10.3102/00346543073001089

Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review*, *47*(2), 171–197. https://doi.org/10.17763/haer.47.2.v870r1v16786270x

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, *21*(1), 1–19. https://doi.org/10.2307/2391875

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. 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.

Appendix

The 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 "student outcome" 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 "stakeholder engagement" OR "family engagement" OR "community engagement" OR "mayoral control" OR "teacher union" OR "collective bargain" OR "portfolio model" 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" OR "wellbeing" OR "support staff" OR "transportation" OR "meal" OR "lunch" OR "discipline" 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."