

The Relationship Between Measures of Preservice Principal Practice and Future Principal Job Performance

Christopher Doss
Melanie A. Zaber
Benjamin K. Master
Susan M. Gates
RAND Corporation
Laura S. Hamilton
ETS

Principals are the second-largest school-based contributor to student achievement. Interventions focused early in the “pipeline” for identifying and developing effective principals might be a promising strategy for promoting principal effectiveness, yet no prior research has examined measures of principal performance during preservice preparation. We analyze 31 measures of principal practices developed by New Leaders and integrate into their year-long, preservice Aspiring Principals program. We link these measures to administrative data in nine districts to understand how they predict student and principal outcomes after candidate placement. We find associations with gains in student achievement on standardized tests, gains in student attendance, and higher rates of principal retention. We compare our results with studies of measures from licensure exams and evaluation systems.

Keywords: *administration, educational policy, leadership, principals, econometric analysis, factor analysis, measurements, policy analysis, secondary data analysis*

PRINCIPALS play an essential role in the functioning and improvement of schools. It is estimated that principals are one of the largest school-based influences on student achievement, second only to teachers (Grissom et al., 2021; Louis et al., 2010). The proposition that principals and other school leaders can affect student achievement both has face validity and has been a subject of interest in scholarly research for four decades (e.g., Brewer, 1993; Dhuey & Smith, 2014; Eberts & Stone, 1988). Principals have a multidimensional job that includes leading the hiring and firing of staff, directing the school’s instructional goals and resources to support the attainment of those instructional goals, and setting the overall vision and culture of the school (Darling-Hammond et al., 2007; Grissom et al.,

2021; Hallinger & Heck, 1998; Leithwood & Jantzi, 2005; Leithwood et al., 2004; Waters et al., 2003).

Measuring performance in this multidimensional role that spans both management and leadership is challenging. With the advent of large-scale administrative data, a body of literature has developed that leverages value-added models to directly estimate principal effects on student test scores (e.g., Branch et al., 2012; Grissom et al., 2015). This approach dovetails with a policy-driven emphasis on school accountability that focuses on instructional leadership, or the principal’s role in improving teaching and learning.

A parallel movement has arisen to identify and systematically measure principal competencies in

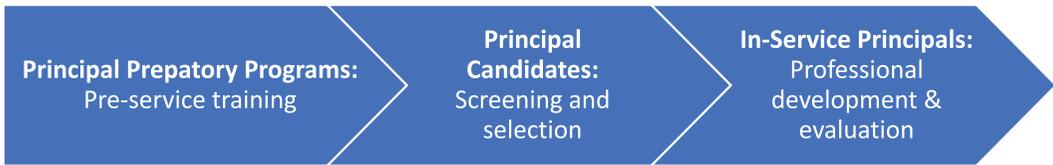


FIGURE 1. *Timeline to the principalship.*

carrying out each dimension (i.e., domain) of the job, and then understand how these competencies relate to policy-relevant outcomes such as principal retention and student achievement. To date, this movement has mostly concentrated on in-service principals due to federal policies such as the Race to the Top competition under the Obama administration, which led states such as Tennessee, Pennsylvania, and New Jersey to establish school leader evaluation systems (Grissom et al., 2018; Herrmann & Ross, 2016; McCullough et al., 2016). This effort continues under the *Every Student Succeeds Act* (ESSA) that allows the use of grant funds for activities that support principals and other school leaders as well as for the measurement and evaluation of school leadership effectiveness.

One ultimate goal of this approach and policy push is to be able to evaluate principals on multiple dimensions of their job. However, measuring school leadership effectiveness is very challenging. Conceptions of effectiveness and principals' potential contributions to school outcomes have given rise to evaluation instruments that may overlap but are not aligned. In addition, although there is a proliferation of measures, some may not be well differentiated, valid, or reliable. Data collection is also challenging because measures must be deployed at scale and connected to outcomes, with only a subset, such as student achievement, broadly available. Furthermore, principal turnover can hamper measuring associations with outcomes because it may take time for a principal to affect outcomes (Corcoran et al., 2009; George W. Bush Institute, 2016).

In addition, although it is generally easier to evaluate the contributions of in-service principals after they have spent significant time leading a school, measuring principal practice at earlier parts of the principal pipeline (as depicted in Figure 1)—namely, during preservice preparation and at the screening and selection stages—may

be even more valuable. First, replacing in-service principals (in the third stage) is costly. It has been estimated that after taking into account the cost of preservice training, hiring, and in-service development, losing a principal typically costs about US\$75,000 (School Leaders Network, 2014). However, with effective early measurement of leadership quality, identifying and cultivating promising candidates and hiring the most promising applicants can be more cost-effective. In this ideal scenario, preparation programs can use measures of practice to assess the needs of the candidates during the training period. They can then modify their approach to address those needs. At the end of the training period, programs can use the measures to ensure that candidates leave the program having demonstrated mastery of the practices to the program's satisfaction. Schools and districts could then select candidates, in part, based on mastery of these practices. In line with this theory, recent evidence suggests that aligning training, selection, evaluation, and professional development of principals around a set of core standards of practice can result in improved student achievement (Gates, Baird, Master, & Chavez-Herrerias, 2019).

Unfortunately, there are a dearth of studies that have measured practice or even the competencies or principal skills thought to underly that practice at the preservice stages of the principal pipeline. We are aware of only one other preservice study that focused on the screening and selection stage of the principal pipeline (Grissom et al., 2017). That study explored how knowledge and skills (measured by licensure test scores) predict improvements in principal value-added measures (VAMs) of student achievement once placed in schools. However, a study about assessment at the licensure stage (after training is complete) does not shed light on the value of assessment during preservice training or the potential of such assessment to contribute to

continuous program improvement and tailored support for individual candidates.

This study contributes to this burgeoning literature on the measurement of leadership practice across the principal pipeline by analyzing efforts to cultivate and measure specific principal practices during the *preservice* portion of the pipeline. We focus on the New Leaders Aspiring Principals program and the practices, called “competencies,” that they identified and measured. We analyze measures of 31 competencies across five standards, or domains of principal practice, that New Leaders consistently used in assessing principal candidates during the 2013–2014 through 2015–2016 school years. In addition to understanding how these domains fit in with scholarly models of school leadership, we use these data to answer three research questions:

Research Question 1: How are New Leaders measures of preservice skills associated with student outcomes such as gains in test scores and attendance?

Research Question 2: How are New Leaders measures of preservice skills associated with principal outcomes such as endorsement by New Leaders, placement into partner districts, and retention?

Research Question 3: How sensitive are results to issues unique to preservice measures of principal skills?

While New Leaders conceptualized competencies spanning five standards, our exploratory factor analysis suggests three underlying constructs across their preservice performance measures. One of those constructs, which we call the Human Capital construct, is composed of competencies that measure a principal’s ability to guide and improve instruction in the school and develop the skills of the school’s staff and is most predictive of student outcomes. A 1 standard deviation (*SD*) increase in scores on this construct is associated with a 0.033 *SD* increase in English language arts (ELA) achievement ($p < .05$), a 0.044 *SD* increase in math achievement ($p < .05$), and a 0.6 percentage point increase in attendance ($p < .01$). A second construct, which we call the Cultural Capital construct, is composed of competencies that measure

a candidate’s ability to foster a culture focused on equity and a productive working and learning environment and is associated with principal retention in the second year. A 1 *SD* increase in this construct is associated with a 9 percentage point increase in the probability of staying in the same school for a second year and 10 percentage point increase in the probability of staying in the same district for a second year ($p < .01$). The third construct, which we call the Personal Leadership construct, was composed of competencies related to the vision and mission of the school, managing change, and reflective practice. A one *SD* increase in that construct is associated with a 7 to 9 percentage point increase in the probability of being placed as a principal ($p < .10$ or less).

This study capitalizes on a unique and rigorous performance assessment system to provide some of the first evidence on how ratings of principals in the preservice context are predictive of policy-relevant principal performance and retention outcomes. The measures included in this study are of particular interest as they are aligned with the preservice program’s curriculum, presenting a unique opportunity for the program to use the measures to support continuous improvement of program content. Findings from this study can help to inform practitioners’ and policymakers’ evaluation efforts to enhance practice-based principal professional development and to screen or certify the effectiveness of future school leaders hired from preservice preparation programs.

Literature Review

Previous Conceptions of School Leadership

A rich line of research has examined the complex ways through which effective school leaders improve outcomes for staff and students. On one level, effective principals are effective organizational leaders. They provide a vision and direction for the school and gather buy-in from the staff and community around that vision. On another level, effective principals manage membership in that community, relationships in the community, and the human capital of the community. This view of the principalship suggests that effective principals engage in a set of practices—as both manager and leader—that

lead to improved outcomes for staff and students (Hallinger & Heck, 1998; Leithwood & Jantzi, 2005; Leithwood et al., 2004).

The five standards that New Leaders developed to judge their candidates are one in a long line of efforts to formalize the dimensions of principal practice. Table 1 illustrates how each New Leaders standard is related to five other conceptualizations of the principalship: (a) Grissom and colleagues' (2021) analysis of principal research since 2000; (b) the Vanderbilt Assessment of Leadership in Education (VAL-Ed; Murphy et al., 2007); (c) Leithwood et al.'s (2004) work on how effective principals inspire staff to new levels and commitment to overcome challenges and reach goals; (d) Hallinger and Murphy's (1985, 2005) work on how effective principals produce the conditions needed for effective and rigorous instruction; and (e) Robinson et al.'s (2008) review of the empirical literature. As Table 1 illustrates, the New Leaders standards and each of the past conceptualizations all have dimensions that substantially overlap with each other, although each may differ on the particular aspects of each dimension and how they are best operationalized in schools.

All conceptualizations of leadership touch on a principal's responsibility to communicate and manage the vision and mission of the school. Some conceptualizations (Hallinger, 2005; Leithwood et al., 2004) directly identify this responsibility while others include it as part of managing the school climate (Grissom et al., 2021), or as establishing goals and expectations (Robinson et al., 2008). New Leaders combined this dimension along with a candidate's ability to reflect on their own practice to form the Personal Leadership standard. Robinson et al. (2008) found a correlation of 0.42 *SD* of their version of this dimension with student outcomes.

All conceptualizations of the principalship also focused on ensuring that the core mission of teaching and learning in schools is occurring. Responsibilities in this dimension include observing and supervising teachers, managing the curriculum, and other interactions that support instruction. However, some conceptualizations like Leithwood et al. (2004) characterize this responsibility as part of general staff development. Grissom et al. (2021) note that not all instruction-related activities are productive

ones and posit that those that support or improve classroom instruction have the most leverage. Robinson et al. (2008) found a moderately large correlation of 0.42 *SD* between their version of this dimension and student outcomes. The New Leaders Instructional Leadership standard is defined as a candidate's ability to guide pedagogy and instructional practices, use data to drive instruction, and observe and supervise schools.

Prior conceptualizations of leadership have also looked at the overall culture and school climate of the school and how well they foster a working environment for teachers and a learning environment for students. Robinson et al. (2008) found that their version of this dimension had a smaller 0.27 *SD* correlation with student outcomes. New Leaders labeled this dimension the Cultural Leadership standard and measured how principal candidates fostered a working environment through systems, routines, behaviors, and by engaging the community—all with an emphasis on equity and cultural competence.

Prior conceptualizations of leadership have considered how principals establish structures conducive to promoting collaboration and improving the human capital of the staff in an effort to improve the teaching and learning in schools. Some frameworks explicitly focus on these behaviors (Grissom et al., 2021; Leithwood et al., 2004; Robinson et al., 2008), whereas others fold these responsibilities into promoting a productive school culture (Hallinger & Murphy, 1985; Murphy et al., 2007). New Leaders labeled this dimension the Adult and Team Leadership standard and included a candidate's performance management, leadership development, and professional development. Robinson et al. (2008) found their largest correlation of 0.84 *SD* between their version of the dimension and student outcomes.

Less often included in formulations of principal practice is the ability to strategically leverage the physical capital and resources of the school, which New Leaders calls the Operational Leadership standard. This measure is missing from Murphy et al.'s (2007) VAL-Ed instrument and from Hallinger and Murphy's (1985) formulations. Furthermore, this dimension of practice is most aligned with Leithwood et al.'s (2004) "Transactional and Managerial" domain,

TABLE 1

New Leaders Aspiring Principal Program Measures and Relationships With Prior Conceptualizations of Principal Practice

New Leaders standard	New Leaders concept	Number of competencies	New Leaders definition	Review of studies since 2000	VAL-Ed instrument	Leithwood et al. (2004)	Hallinger and Murphy (1985), Hallinger (2005)	Literature review
Personal leadership	Reflective practice and continuous improvement	2	Ability to reflect on his or her own practice and articulate and manage a vision for the school	Grissom et al. (2021)	Murphy et al. (2007)	Setting directions	Defining the school's mission	Establishing goals and expectations
	Communication and interpersonal relationships	2						
	Vision and mission	2						
	Managing change	2						
Instructional leadership	Pedagogy and instructional strategies	2	Ability to guide instructional practice in the school and ensure the staff are working to provide students with high-quality, standards-aligned instruction	Engaging in instructional focused interactions with teachers	Rigorous curriculum; quality instruction; systemic performance accountability	Developing people; transactional and managerial	Managing the instructional program	Planning, coordinating, and evaluating teaching and the curriculum
	Data-driven instruction	2						
	Observation and supervision of instruction	2						
Cultural leadership	Urgency and building schoolwide efficacy	3	Ability to foster a culture focused on equity and a productive working and learning environment	Building a productive school climate	Culture of learning and professional behavior; connections to external communities	Redesigning the organization	Promoting a positive school climate	Ensuring an orderly and supportive environment
	Equity and cultural competence	2						
	Systems, routines, behaviors, and code of conduct	3						
Adult and team leadership	Family engagement	1						
	Performance management	2						
	Leadership development	2	Ability to manage the professional development and leadership development in a school	Engaging in instructionally focused interactions with teachers; facilitating collaboration and professional learning communities	Culture of learning and professional behavior	Developing people	Promoting a positive school climate	Promoting and participating in teacher learning and development
	Professional development	1						
Operational leadership	Diagnostic and strategic planning	3	Ability to leverage the physical capital and resources in the building in a strategic manner	Managing personnel and resources strategically	NA	Transactional and managerial	NA	Strategic resourcing

which they posit has little relationship with student outcomes. Robinson et al.'s (2008) category of "Strategic Resourcing," however, had a 0.31 *SD* correlation with student outcomes.

Recent Efforts to Measure Principal Skills

More recently, the multidimensional conception of school leadership has been integrated into large-scale third-party assessments of school leadership. These assessments are a response to a federal and state policy landscape that has increasingly tied school and district funding to evaluation of both teachers and principals, leading to broader adoption of evaluation systems in K–12 schools and districts (Donaldson et al., 2021; Superville, 2014). Various provisions of ESSA permit the allocation of grant funds for activities supporting principals and other school leaders, as well as for measuring teacher and principal effectiveness (Haller et al., 2016). Race to the Top encouraged teacher and principal evaluation, and six states implemented full principal evaluation systems by the 2012–2013 school year (when teacher quality funding expanded to include school leader evaluation), providing useful context for analyzing their benefits and consequences. A review of principal evaluations notes that all states now require that principals receive a summative rating based on multiple measures of performance and most states require districts to assign one of four performance ratings, with clear consequences to performing below standard (Donaldson et al., 2021). A 2015 study reported that 67% of states allowed, recommended, or mandated that the results be used in making personnel decisions (Fuller et al., 2015).

To our knowledge, to date, three states have developed large-scale principal assessment measures that have been analyzed by researchers: Tennessee (Grissom et al., 2018), Pennsylvania (McCullough et al., 2016), and New Jersey (Herrmann & Ross, 2016). There are also efforts underway to formalize and standardize these measures. For example, the National Professional Standards for Educational Leaders (PSEL) identified a set of 10 research-grounded, core practice areas spanning the above categories (National Policy Board for Educational Administration, 2015). The PSEL have informed revisions to

standards in a number of states (Scott, 2018) as well as standards for novice administrators or preparation program graduates (University Council for Educational Administration, 2018). However, much like the literature that precedes these assessments, to date each state's or organization's measures or conceptions of leadership varies in the specific categories they delineated and the skills that compose each category.

Challenges in Measuring Principal Effects

Efforts to identify the core domains of a principal's job that are most consequential to the functioning of a school, and subsequent efforts to assess principal performance on these domains at scale, require a way to isolate the effects of principals on school outcomes. This is not a straightforward proposition. Some challenges are akin to those in measuring teacher contributions to learning. Foremost, models must account for the sorting of students and educators. VAMs leveraging standardized tests have been the most popular way to account for this sorting and selection, however; these models can be sensitive to the test metric used and test measurement error can introduce error into the VAMs (Boyd et al., 2013).

Other factors specific to principals complicate the endeavor further. Whereas teachers are expected to affect student learning immediately, some aspects of principals' efforts, such as changing the school environment or hiring different types of personnel, may take longer. Some studies suggest it can take up to 3 years for a principal to affect student outcomes through these channels (Corcoran et al., 2009; George W. Bush Institute, 2016). Even when principals affect student outcomes, effects may be smaller compared with teacher effects because principals are further removed from the classroom. This fact, coupled with the smaller number of principals relative to teachers, can make effects hard to detect (Gates et al., 2014). Finally, there are many factors relevant to student learning that are outside the principal's control that may not be completely accounted for by traditional covariate controls. For example, Grissom et al. (2015) identify the importance of partially uncontrollable factors like the strength of the local hiring pool and the degree of support from the surrounding community.

Recent scholarship has made inroads to some of the issues in accounting for school characteristics with the introduction of principal VAMs (Fuller & Hollingworth, 2014). There are three general approaches to principal VAMs. Each have their advantages and drawbacks, with the nuances of each discussed and compared in Grissom et al. (2015). The first follows the teacher VAM literature and uses lagged achievement, student and school characteristics, and a series of fixed effects to account for sorting. The downside to this approach is that it assumes an immediate impact of principals on student outcomes. A second approach uses principal turnover in a school to compare the relative effectiveness of different principals in the same building. This approach may be unappealing because the strength of one principal in this model is dependent on the strength of the principal who precedes him or her. Finally, the third looks at improvement over a principal's tenure, allowing for the accumulation of effects, although the data needs for this approach are high. There is a growing body of literature that estimates these measures, most often with the first and second approaches (Branch et al., 2012; Chiang et al., 2016; Coelli & Green, 2012; Dhuey & Smith, 2014; Grissom et al., 2015) and all which detect meaningful variation in principal effects.

Associating Measures of Principal Practice With Student Outcomes

Building upon our increasing capacity to identify individual principals' contributions to learning, researchers have begun to analyze how measures of principal knowledge and practice are associated with student outcomes by incorporating the measures into value-added models to understand correlations in a model that plausibly teases out principal effects. This more recent line of work improves upon earlier efforts that relied on correlations or structural equation modeling (e.g., Waters et al., 2003). These studies have been fielded in using supervisor ratings of principals in Tennessee (Grissom et al., 2018), principal licensure test results in Tennessee (Grissom et al., 2017), and ratings of principals in Pennsylvania (McCullough et al., 2016) and New Jersey (Herrmann & Ross, 2016). Principal licensure scores were more

correlated with job placement and less on student outcomes. At least a subset of principal in-service ratings, however, were associated with student outcomes in Tennessee, Pennsylvania, and New Jersey.

These efforts have revealed two key challenges in measuring associations between principal ratings and student outcomes: Ratings can be correlated to principal characteristics or school characteristics. Grissom and colleagues found that the principal licensure scores were systematically lower for minority test takers and scores were negatively associated with the proportion of minority students and economically disadvantaged students in the school in which principals were eventually placed (Grissom et al., 2018). This correlation with minority and economically disadvantaged students was also seen with ratings in New Jersey (Herrmann & Ross, 2016). Biased measures would undermine their utility because they would be less related to skills the principals can improve and more related to contexts in which they happen to work—potentially discouraging principals from working in challenging environments.

Studies of Principal Practice at the Preservice Stage

The bulk of studies that have contributed to the knowledge base of domains of principal practice, how to measure them, and how to estimate principal effects has largely come from studies of in-service principals. The dearth of studies looking at preservice practice is perhaps not a surprise given that even wholesale evaluations of principal preparation programs are rare (George W. Bush Institute, 2016). Primary metrics tracked by principal pipeline programs are candidate placement and tenure rather than preservice practices (e.g., Davis & Darling-Hammond, 2012), with some programs utilizing licensing exam scores as a measure of competencies. In a more recent study, Grissom et al. (2019) used administrative data from Tennessee to analyze the variation in outcomes of graduates from 12 principal preparation programs. They analyzed licensure exam scores, job placement, and in-service evaluations. They found that programs did not have stable relative rankings across metrics, but their data did not allow them to disentangle why a

particular program performed better on a particular metric.

New Leaders conceptualized a set of principal competencies based on the previous work of scholars. We connect these measures to administrative data in nine districts over 3 years to estimate associations with principal outcomes and with student outcomes based on the principal value-added methods recently devised. To our knowledge, this is the first time measures that were consistently and purposively designed and deployed in the preservice context have been associated with outcomes once principals are placed in schools. Our results can provide insight as to which skills are feasible to evaluate and important to develop in an early phase of the principal pipeline where measures of effectiveness may be particularly valuable for informing both training and hiring of new principals. Finally, this study also presents the opportunity to examine possible bias in these measures.

New Leaders Context

Overview of New Leaders Program

New Leaders was established in 2000 with the mission of improving school performance by developing effective school leaders to serve in urban schools. The New Leaders program recruits aspiring principals and trains them through a year-long residency program in partnering districts. The completion of the Aspiring Principals program can culminate with an endorsement from New Leaders. An endorsement indicates the candidate met a threshold of performance set forth by New Leaders, but does not have any formal impact on a candidate's ability to apply for jobs because candidates can satisfy licensure requirements without obtaining an endorsement. An endorsement does not guarantee a resident a job in the partnering district. All residents must undergo the same application, screening, and placement process in the district as would outside applicants. Residents do not have to serve in the partnering districts and may apply for any position in any district.¹ Two evaluations have shown that students in schools with New Leaders-trained principals outperformed students in schools led by principals trained through other avenues by up to 0.089 *SD* in math

and 0.057 *SD* in ELA (Gates, Baird, Doss, et al., 2019; Gates et al., 2014).

A central component of the New Leaders program for Aspiring Principals is their year-long residency in which participants hold an official position in a partnering district, usually an Assistant Principal position.² While performing the duties of their job, New Leaders residents participate in learning experiences through mentorships, simulations, role-playing, and feedback. Each resident must identify and work toward personalized goals. Most pertinent to this study, New Leaders devised and integrated a series of performance assessments into the year-long residency program. New Leaders has identified specific principal practices, called competencies, across five standards, on which residents must demonstrate proficiency.

Structure of New Leaders Aspiring Principals Program Measures

Based on a review of the principal leadership literature, New Leaders developed a set of preservice performance assessments that span the previously detailed five standards, or dimensions of principal practice in which they determined a successful principal should be proficient. Each of the five standards is divided into more fine-grained concepts that delineate more specific categories of practice that principals are expected to display. These concepts are further divided into the specified competencies on which residents are assessed. Figure 2 provides a sample structure of the Personal Leadership standard, concepts, and competencies. These standards and performance assessments were integrated into the residency program beginning in 2012 (for Cohort 12 of the program).

We conducted an exploratory factor analysis to gauge the extent to which the performance measures collected by New Leaders in fact measured distinct skill sets.³ In practice, we found that, as assessed, the five New Leaders standards mapped to three distinct higher-level constructs. The first construct is mainly composed of the competencies within the Instructional Leadership and Adult and Team Leadership standards. We named this construct the "Human Capital" construct because these competencies are focused on the delivering of instruction, evaluations of how

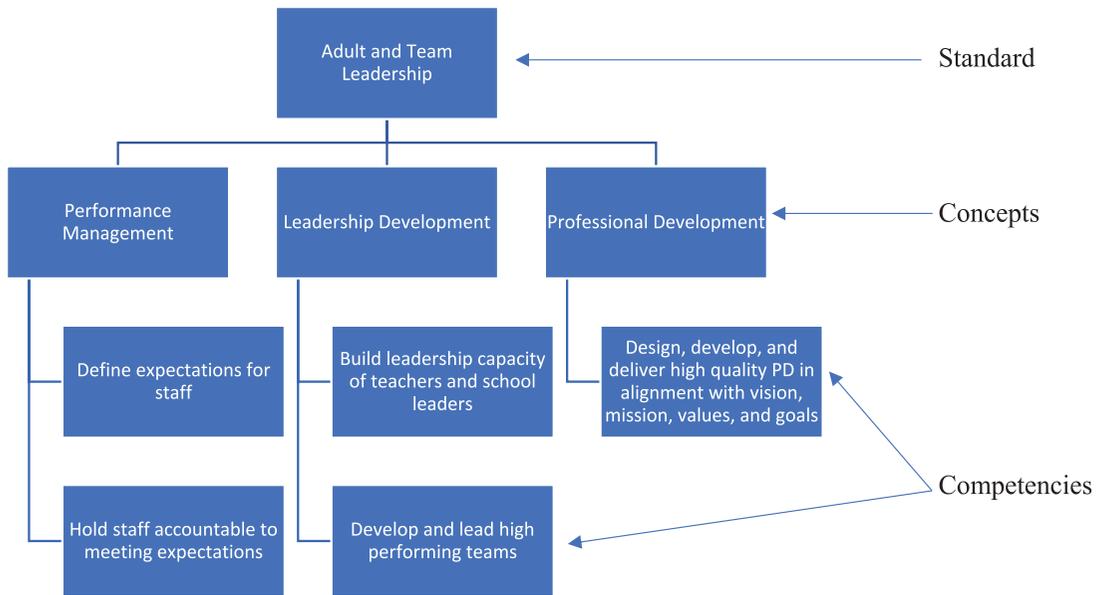


FIGURE 2. *Sample structure of standards, concepts, and competencies.*

well staff delivers instruction, and the professional development of the staff. The second construct is mainly composed of competencies in the Cultural Leadership standard, which we therefore named the Cultural Capital construct. Finally, the third construct is mainly composed of the competencies within the Personal Leadership standard, which we called the Personal Leadership construct.

Deployment of New Leaders Measures

New Leaders integrated these measures into several aspects of the residency year, using them as the basis for formative assessments of which to tailor feedback and instruction and summative assessments that contributed to New Leaders' endorsement decision. This approach entailed multiple New Leaders staff rating the resident throughout the year on competencies in different contexts that range from observations of practice to 360° surveys (for more detail on these contexts, please see Appendix B in the online version of the journal). Residents are not rated on all competencies at once; rather, competency ratings are staggered throughout the year and integrated into the residency curriculum. New Leaders staff were trained on the assessment process, the rubrics, and the role of ratings in evaluating the

resident, including its role in the endorsement decision. Documents were circulated that reinforced the training and ratings expectations and staff justified their ratings with evidence.

Each year, New Leaders revisits the measures and considers the need for revisions based on the feedback from the previous year. Beginning with Cohort 13 in the 2013–2014 school year, the measures remained relatively stable. No changes were made between Cohorts 13 and 14 and minor changes were made for Cohort 15. In this analysis, our goal was to use measures that were as comparable across years as possible. Therefore, we organized the competencies identically in Cohorts 13 through 15. To do so, we took the 31 competencies that were used in all 3 years and reorganized the competencies of Cohorts 13 and 14 to match those of Cohort 15. The measures delineated in Table 1 represent this consistent set of measures used in the analysis.

In Cohorts 14 and 15, there were two cycles of competency ratings, one before the middle of the year and one afterward. The first set of ratings were used solely for feedback and instruction and did not count toward the final score. The second set of ratings were also used for feedback and instruction and contributed to the final scores. Within each cycle, a competency was measured multiple times, with each instance of

measurement occurring in a different context. For the two later cohorts, New Leaders created a single measurement of the competency by averaging the multiple ratings of the same competency within the second cycle. In Cohort 13, both cycles contributed to the final rating, but the second cycle was weighted twice.

In all years, the final concept ratings were one of several factors that contributed to the final endorsement decision. Thus, each year, New Leaders used the measures to provide guidance on how to improve throughout the year and then as a final determination of skill level at the end of the residency period. As the goal of this study is to understand whether preservice programs can identify and measure principal practice that is predictive of student and principal outcomes, we remain agnostic to year-to-year fluctuations in the ways competencies were measured. Rather, we are careful to ensure that the measures across years are composed of the same underlying competencies and then assess whether these aggregate measures are predictive of outcomes that are consequential for students and principals.

Data

This study leverages programmatic data provided by New Leaders and administrative data from partnering school districts. Programmatic data from New Leaders include competency averages for Cohorts 13, 14, and 15. Members of Cohort 13 were residents in school year 2013–2014, members of Cohort 14 were residents in 2014–2015, and members of Cohort 15 were residents in 2015–2016. New Leaders also provided demographic data on candidates, including recruitment pathway, years of experience as a teacher, gender, race/ethnicity (Black, Hispanic, White, or Other), whether the residency took place in a charter school, and whether the residency took place in a school led by a New Leaders alumnus. We also determined whether the candidate was endorsed by New Leaders for principal placement, and what placement role(s) they reported to New Leaders. New Leaders also provided assessment scores obtained through the relevant recruitment pathway that was used, in part, for selection into the Aspiring Principals program.⁴

We link this programmatic data to school- and student-level data for graduates who were placed in the following nine partnering school districts during the 2014–2015 through 2016–2017 school years: Baltimore City Public Schools, Charlotte-Mecklenburg Public Schools, Chicago Public Schools, New York City Public Schools, Oakland Unified School District, Prince George’s County Public Schools, Shelby County Public Schools, Washington, D.C., Public Schools, and Washington, D.C., Charter Schools. The districts provided data on student achievement on standardized tests of reading and math, student attendance (expressed in the percentage of the school year present), the student’s grade (third grade, fourth grade, etc.), whether a student repeated the grade, race/ethnicity (Black Hispanic, White, or Other), English Language Learner status, and gender.⁵ We obtained data for the 2013–2014 through 2016–2017 school years.

In total, Cohorts 13 through 15 contained ratings for 216 Aspiring Principal residents. Of those residents, 71 were placed into partnering districts in between 2014–2015 and 2016–2017 and matched to the administrative data. As earlier cohorts were in schools for a longer period of time, we have 3 years of data for a resident from Cohort 13, 2 years of data for a resident from Cohort 14, and 1 year of data for a resident from Cohort 15.

Supplemental Table A1 (available in the online version of this article) provides the descriptive statistics for all residents and those who were placed into partner districts and matched to partner districts’ administrative data. We compare the two samples to illustrate the representativeness of the residents matched to schools compared with the overall sample of residents. Both samples of residents scored an average of about 3.3 out of 4 on each of the five standards. Residents had about 7.5 years of prior teaching experience. The majority were female and minority, with a substantial number having done their residency in schools led by New Leaders alumni. *T*-tests comparing the full sample of residents with those matched to the district data show that the matched sample performed slightly higher on the measures and differed on some background characteristics. We include controls for these variables in our models to account for these differences and to mitigate any

sorting of New Leaders to schools based on observables.

Online Supplemental Table A1 also presents the characteristics of students and schools served by the New Leaders principals placed in the partnering schools. On average, students in the analytical sample were about half male (51%), mostly minority (4% White), and 18% were classified as English Language Learners. These demographics mirrored the demographics of the schools which they attended. Schools were mostly elementary schools (66%), with fewer middle schools (24%) and high schools (10%). The schools were also served predominantly Black and Hispanic students, with 64% of students identified as Black and 28% as Hispanic. About 18% of students were English Language Learners and 4% repeated a grade.

In the 3 years in our sample, the 71 principals serve more than 35,700 unique students, which highlights the broad reach of a principal position. Thus, although the count of principals in this sample may seem relatively small, this study investigates effects on a large population of students which, in part, allows us to detect policy-relevant relationships. This statistical power, in combination with the distinct categories of measures created by New Leaders, allows us to understand which dimension of a principal's practice cultivated at the preservice level is associated with which outcomes.

Method

Links to Student Outcomes

We focus on policy-relevant student outcomes like attendance and performance on standardized tests of reading and mathematics. We explore the relationship between student outcomes and the standards, as conceived by New Leaders, as well as the relationship between outcomes and the constructs of the underlying competencies, as represented by factor scores from exploratory factor analysis. It should be noted that all New Leaders principals in this analysis were endorsed by the program. We use models of the following form to estimate the relationship between New Leaders measures and principal contributions to student outcomes:

$$\begin{aligned}
 Y_{ipsdt} = & \beta_0 + \beta_1 M_p + Y_{ipsdt-1} \beta_2 \\
 & + \mathbf{X}_p \beta_3 + \mathbf{C}_{ipsdt} \beta_4 + \mathbf{S}_{sdt} \beta_5 \\
 & + \tau_t + \delta_g + \alpha_c + \gamma_d + \varepsilon_{ipsdt},
 \end{aligned} \tag{1}$$

where Y_{ipsdt} is the math, ELA, or attendance outcome of student, i , where principal, p , is leading school, s , in district, d , and year, t . Math and ELA scores were standardized within district and year while attendance is the proportion of days a student was present. Thus, math and ELA results have an effect size interpretation and attendance outcomes have percentile point interpretation. M_p represents the New Leaders measure of interest. When analyzing standards as conceived by New Leaders, we place each standard in a separate regression. When analyzing factor scores of the underlying competencies, we place all factor scores in one regression because factor scores are orthogonal to each other. $Y_{ipsdt-1}$ is the student's lagged academic or attendance outcome. Because a resident in earlier cohorts can be in a school for more than 1 year, this variable always takes on the value of the year before the resident entered the school as principal. In models that analyze math and ELA outcomes, we include lagged achievement on both subjects. \mathbf{X}_p is a vector of time-invariant New Leaders characteristics, including two possible recruitment pathways, whether principals passed the Emerging Leaders pathway recruitment process, years of experience as a teacher, gender, race/ethnicity (Black, Hispanic, Other), whether their residency took place in a charter school, and whether their residency took place in a school led by a New Leaders alumnus. This vector includes summative rating residents received from their recruitment pathway. These measures are related to each standard used in the Aspiring Principal program, but do not correspond in a one-to-one fashion. We therefore include it as a control for baseline abilities but note that the lack of one-to-one correspondence between these measures and our program measures of interest mean that our estimates do not represent how growth in the program measures predict our outcomes of interest. \mathbf{C}_{ipdst} is a vector of child-level covariates, including grade level, gender, race/ethnicity (White, Black, or Hispanic), and indicators for having repeated the grade, English Language Learner

Status, and being older than expected for the grade. \mathbf{S}_{sdt} is a vector of school-level characteristics that include school-level averages of the above child characteristics, except grade level is replaced by grade band (elementary, middle, high school) and enrollment of the school is included. Finally, τ_p , δ_g , α_c , and γ_d are time, grade, New Leaders cohort, and district fixed effects, respectively.

The inclusion of student- and school-level covariates controls for bias due to student sorting that is not accounted for by the lagged outcomes. Grade fixed effects account for stable differences among grades, cohort fixed effects for differences among New Leaders cohorts such as differences across cohorts in the calculation of competency measures, time fixed effects for common yearly shocks to principal performance and student achievement, and district fixed effects for stable differences among districts, such as the strength of the local labor market.

This model is akin to the value-added model used by Grissom et al. (2018) when analyzing the relationship between principal supervisor ratings and student outcomes and VAMs used to measure teacher effectiveness. The identifying assumption is that the prior year student test scores, fixed effects, and student, principal, and school observable characteristics account for sorting of principals and students to schools such that the remaining variation in student outcomes within a district is associated only with remaining variation in skills across principals, as captured by the New Leaders measures. The unit of observation is the principal-year and we relate their preservice competency measures to yearly outcomes in their school(s). Standard errors are clustered by school to account for the correlation of student outcomes in a school.

One main threat to this identification strategy is whether principals sort to districts or schools on long-term trends in student achievement not fully captured by the 1-year lagged achievement. Another main threat is whether the resident characteristics do not fully control for resident sorting to schools. This second threat may be particularly pertinent to studies that examine ratings of principals and has been seen in other contexts such in principal ratings in Tennessee (Grissom et al., 2018). The limited longitudinal nature of our data set precludes us from assessing

any bias on trends in school student achievement, but we interrogate the robustness of the above model to various district-level time trends in the “Robustness and Sensitivity Checks” section. We also look at the importance of controlling for resident characteristics.

A second issue that may be more relevant to principals who recently graduated from preservice programs is the limited amount of time they are in schools. Our preferred sample contains all observed resident-year observations, but this approach implicitly weights earlier cohorts more because they can be observed as principals in schools for a longer period of time. We therefore look at results for a “balanced sample” that only includes each resident’s first year as principal. However, there is evidence that it may take up to 3 years for a principal to affect student achievement (Corcoran et al., 2009). We therefore compare results from each sample to understand how resident measures may differently predict student achievement in a relatively short time frame.

Links to Principal Endorsement and Placement

School districts may incorporate a variety of information inputs when making hiring decisions, implicitly predicting future job performance given the information at hand—a process called signaling (Goldhaber, 2007; Goldhaber & Hansen, 2010; Grissom et al., 2018). The New Leaders measures of candidate skills can serve as signals in two ways. New Leaders makes its endorsement decisions partially based on performance of the measures. Although this endorsement decision carries no consequences for credentialling, it may act as a signal of their abilities. Candidates rated more highly on the measures may also display skills or dispositions during the hiring process that districts value.

We use linear probability models to analyze the endorsement and placement of New Leaders program participants.⁶ Models take the following form:

$$\text{PrincipalOutcome}_i = \beta_0 + \beta_1 M_p + \mathbf{X}_p \beta_2 + \alpha_c + \varepsilon_i, \quad (2)$$

where $\text{PrincipalOutcome}_i$ is an indicator for one of the following outcomes of interest: *not endorsed* by New Leaders, *first placement* in a school leadership role (assistant or lead principal), *first placement as lead principal*, and *ever*

being placed as lead principal (over potentially multiple observed placements). The elements on the right-hand side of the equation contain the same variables as their respective elements in Equation 1. As this analysis focuses on placement, we do not include school or district fixed effects. The coefficient of interest, β_1 , will provide the relationship between the New Leaders measure and the outcome of interest, when accounting for resident covariates and differences in district context and cohorts. We use 202 program participants when analyzing the relationship between residency measures on endorsement.⁷ The sample used to analyze the relationship with placement outcomes only contains endorsed residents. Thus, in this analysis, the unit of observation is the principal, and we relate preservice measures to the one-time endorsement decision and placement outcomes.

We view these results as correlational, as there can be many unobserved principal characteristics related to both performance on the measures and the probability of endorsement or placement. For example, principal dispositions and interpersonal skills correlated with, but not fully captured by, the measures may make a candidate more attractive to districts. Other credentials, such as educational attainment or field of study, may have the same effect. Furthermore, the measures may be related to characteristics of the schools to which principals apply that can also affect their probability of placement. For example, candidates with higher measures may apply to more “desirable” schools with better performance trends or, conversely, they may feel better equipped to lead a “less desirable” school with worse performance trends.

Links to Principal Retention

Principal turnover is disruptive to the functioning of a school (Grissom et al., 2019) and represents a sunk cost of recruiting, onboarding, and potentially supporting new principals (School Leaders Network, 2014). Turnover in 2 years or less means a prolonged period of leadership instability, which can be detrimental to student achievement (Rangel, 2018). Nationally, 35% of principals are in their schools for less than 2 years and the national turnover rate of principals is 18%, with higher poverty schools

experiencing higher turnover (Levin & Bradley, 2019). Thus, rapid principal turnover is not trivial.

We also use linear probability models to explore the relationship between residency measures and principal retention in their first placed position for 2 years and retention in district as a principal for 2 years. Note that we cannot observe the reason for failures in retention. Departures have been voluntary on the part of the principal or the district may have believed another person was more suitable for the position. Nevertheless, we would expect that failure to retain recently hired principals is in most cases a negative outcome from a district management perspective.

These models take the following form:

$$\text{Retained}_i = \beta_0 + \beta_1 M_p + \mathbf{X}_p \beta_2 + \gamma_d + \varepsilon_i,$$

where Retained_i is an indicator for the principal remaining in the school or in the district for a second year. Models that analyze retention outcomes have the smallest sample size because retention outcomes can only be analyzed after matching to the aforementioned nine partnering districts for which we have administrative data. Furthermore, Cohort 15 must be excluded from the analysis because cohort members had not been in the district for more than 1 year at the conclusion of the 2016–2017 school year, the final year of data in our sample (retention for this group is completely truncated). Due to the limited sample size, we include only race/ethnicity and gender in the vector \mathbf{X}_p , eliminate cohort fixed effects, and present results both with and without district fixed effects γ_d . Finally, in this analysis, the unit of observation is once again the principal and we relate preservice measures to the two definitions of retention (i.e., 2-year retention in their first placement and 2-year retention as a principal in any school in the district). We once again consider these results to be correlational as they can suffer many of the same omitted variable biases detailed in the placement analysis.

Results

The Relationships Between Residency Measures and Student Outcomes

Table 2 illustrates that several residency measures are statistically significantly related to

TABLE 2

Relationship Between New Leaders Measures and Student Outcomes

New Leaders Measure	(1)	(2)	(3)
	ELA	Math	Attendance
Panel A: Constructs of competencies (Single regression)			
Construct 1 (Human capital)	0.033* (0.011)	0.044* (0.015)	0.006** (0.002)
Construct 2 (Cultural capital)	-0.007 (0.011)	-0.023 (0.016)	0.003 (0.002)
Construct 3 (Personal leadership)	-0.004 (0.013)	-0.016* (0.006)	0.002 (0.002)
<i>p</i> -value of <i>F</i> -test (Equality of all constructs)	0.154	0.044	0.050
Panel B: Standards (Separate regressions)			
Standard 1 (Personal leadership)	0.001 (0.013)	-0.011 (0.008)	0.005 (0.003)
Standard 2 (Instructional leadership)	0.022 (0.013)	0.027 (0.015)	0.006* (0.002)
Standard 3 (Cultural leadership)	0.008 (0.010)	0.003 (0.011)	0.004 (0.003)
Standard 4 (Adult & team leadership)	0.041* (0.016)	0.029 [†] (0.013)	0.007* (0.003)
Standard 5 (Operational leadership)	0.022 (0.016)	0.011 (0.012)	0.005 (0.003)
Baseline ELA and Math score	✓	✓	
Baseline attendance			✓
Observations	28,489	28,489	51,803

Note. Standard errors clustered at school level. All models include cohort fixed effects, year fixed effects, and district fixed effects. New Leaders principal covariates include preresidency recruitment pathway, an indicator for passing the Emerging Leaders Program screening, years of experience as a teacher, gender, race/ethnicity, an indicator for the residency occurring in a charter school, an indicator for the residency occurring in a school led by a New Leaders alumnus from a previous cohort, and preresidence recruitment pathway ratings (either overall for factors or standard specific). Student covariates include fixed effects for grade, an indicator for having repeated a grade, classification as an English Language Learner, student race/ethnicity, gender, and an indicator for being old for the grade. School covariates include school enrollment, school level, and school level averages of race/ethnicity, gender, English Language Classification, students repeating a grade, and students old for their grade. Constructs were made from underlying competency data. ELA = English language arts.

[†] $p < .1$. * $p < .05$. ** $p < .01$.

consequential student outcomes. Panel A presents the results on the underlying constructs of the measures. The Human Capital construct is most robustly related to student outcomes. A 1 *SD* increase in this construct is associated with a 0.033 *SD* increase in student ELA scores ($p < .05$), 0.044 increase in math scores ($p < .05$), and a 0.6 percentage point increase in attendance ($p < .01$). The magnitude of these associations is on par with those seen in Grissom et al. (2018) who found that average supervisor ratings in Tennessee were associated with a 0.01 to 0.05 *SD* gain in ELA and Math depending of the value-added

specification. The one exception is the Personal Leadership construct, where a 1 *SD* increase is associated with a 0.016 *decrease* in math performance ($p < .05$).

Panel B presents the results on the average standard scores, as measured by New Leaders. The standards most robustly associated with student outcomes are the Instructional Leadership and Adult and Team Leadership standards. A 1 *SD* increase in the Instructional Leadership standard is associated with a 0.6 percentage point increase in attendance ($p < .05$). Meanwhile, a 1 *SD* increase in the Adult and Team Leadership

standard is associated with a 0.041 *SD* increase in ELA scores ($p < .05$) and a 0.029 *SD* increase in math scores ($p < .10$). The relationship with attendance is a positive 0.7 percentage point increase ($p < .05$). Recall that the Instructional Leadership and Adult and Team Leadership standards are the primary components of the Human Capital construct. In this sense, the results from Panels A and B complement each other.

Collectively, these results suggest that measures of leadership in instructional goal setting, observations of instructional quality, and professional development are most highly associated with student outcomes. Those measures of leadership in core teaching and staff development domains directly affect student outcomes. These results are broadly consistent with aspects of principal leadership that scholars have long posited matter. Waters et al.'s (2003) review of the literature saw correlations between principal leadership of curriculum, instruction, and assessment and student achievement that range from 0.08 to 0.24 *SD*. Robinson et al.'s (2008) review concluded that planning, coordinating, and evaluation teaching and the curriculum had a 0.42 *SD* correlation with student outcomes and that promoting and participating in teacher learning and development had a 0.84 *SD* correlation with student outcomes. These associations in older studies are larger likely because those models did not account for bias in as rigorous a manner compared with more recent value-added models.

Results from more recent studies that leverage VAMs also comport with our findings. In Pennsylvania's Framework for Leadership (FFL) tool, aspects of the systems and professional community leadership domains focused on establishing expectations, supporting professional growth, and leveraging resources to support students. The associations between those domains and student achievement are about 0.05 to 0.07 *SD* (McCullough et al., 2016), which is of the approximate magnitude of our findings. Most recently, Grissom et al. (2021) posit that activities supporting classroom instruction are the most productive instructional-related activities.

The Relationships Between Residency Measures and Principal Outcomes

To explore the extent to which these measures serve as signals, we explore their relationship

with the probability of not being endorsed, the probability of being placed as an assistant principal or principal in the year after finishing the residency, the probability of being placed as a principal in the year after finishing the residency, and the probability of ever being placed as a principal in the 3 years in our sample. Given the numerous costs of principal turnover, we then explore the relationship between the measures and the probability of retaining the principalship in the second year or remaining in the district as a principal in the second year.

Table 3 presents the relationship between standards or constructs and endorsement or placement after completing the residency. Panels A and B show that all standards and all constructs except the Cultural Capital construct have a strong correlation with endorsement. Residents who score better on these measures are less likely to be declined an endorsement (i.e., they are more likely to be endorsed). Point estimates range from a low of 4 percentage points for the Personal Leadership factor (60% increase) to a high of 10 percentage points (156% increase) for the Instructional Leadership standard. These point estimates translate to large changes in percentage terms because only 6.4% of the sample was not endorsed. These results may be expected because the endorsement decision is made in part by performance on these ratings.

Fewer standards and factors are significantly related to placement outcomes. The Adult and Team Leadership and Cultural Leadership standards are the only measures associated with placement as an assistant principal or principal the year after the residency. The Adult and Team Leadership standard is most robustly associated with all placement outcomes. A 1 *SD* increase in that standard is associated with an 8 percentage point increase in the probability of being placed as an assistant principal or principal the first year after the residency and a 10 percentage point increase being placed as a principal either in the first year or at any time in the school years captured in this study. The Personal Leadership construct and the Personal Leadership standard that composes it are also related to principal placement outcomes. A 1 *SD* increase in that construct or standard is associated with a 7 to 9 percentage point increase in principal placement ($p < .10$ or less). The remainder of the constructs and standards have positive and potentially large point

TABLE 3

Relationships Between New Leaders Measures and Principal Placement

New Leaders Measure	(1)	(2)	(3)	(4)
	Not endorsed	First placement as assistant principal or principal	First placement as principal	Ever placement as principal
Panel A: Constructs of competencies (Single regression)				
Construct 1 (Human capital)	-0.0800** (0.0157)	0.0376 (0.0364)	0.0788 [†] (0.0408)	0.0430 (0.0418)
Construct 2 (Cultural capital)	-0.0151 (0.0157)	0.0311 (0.0327)	-0.0224 (0.0366)	-0.0144 (0.0375)
Construct 3 (Personal leadership)	-0.0384* (0.0151)	0.0196 (0.0328)	0.0653 [†] (0.0368)	0.0872* (0.0377)
Panel B: Standards (Separate regressions)				
Personal leadership	-0.0821** (0.0155)	0.0447 (0.0347)	0.0710 [†] (0.0392)	0.0662 (0.0402)
Instructional leadership	-0.1000** (0.0153)	0.0351 (0.0364)	0.0399 (0.0412)	0.0131 (0.0424)
Cultural leadership	-0.0571** (0.0160)	0.0593 [†] (0.0333)	0.0238 (0.0380)	0.0450 (0.0389)
Adult and team leadership	-0.0851** (0.0158)	0.0833* (0.0357)	0.106** (0.0402)	0.102* (0.0414)
Operational leadership	-0.0661** (0.0159)	0.0538 (0.0337)	0.0431 (0.0382)	0.0467 (0.0392)
Controls				
New Leaders demographics	✓	✓	✓	✓
Cohort fixed effects	✓	✓	✓	✓
Observations	202	191	191	191

Note. Coefficient cells in each column in Panel B are from five independently run models of retention. Demographic controls are indicators for African American, Hispanic, and female. New Leaders standards are standardized; coefficients represent a 1 *SD* increase in the standard. Constructs were made from underlying competency data. Standard errors in parentheses.

[†] $p < .1$. * $p < .05$. ** $p < .01$.

estimates, but rarely reach traditional significance levels.

Finally, when analyzing retention outcomes, all standards except for Instructional Leadership relate at least marginally to retention outcomes, although the only construct with a significant relationship with retention is Cultural Capital. As seen in Table 4, among the standards, Personal Leadership has the strongest relationship, increasing retention in position and in district by approximately 13 percentage points. A 1 *SD* increase in Cultural Leadership is associated with an approximately 10 percentage point increase in position and district retention. After accounting for district-specific effects, Operational Leadership is also associated with

retention. Adult and Team Leadership is weakly associated with in-district retention alone.

Looking across the student and principal outcomes, the Adult and Team Leadership standard is the one standard most consistently related to outcomes, emphasizing the potential importance of leadership in raising the human capital and professional development of the school's staff. Interestingly, the Personal, Cultural, and Operational Leadership standards and factors are more strongly related to principal placement and retention outcomes than student outcomes. This difference may indicate that a broader set of potentially different competencies may be pertinent to the districts' decision to hire principals and to the principal's decision to stay in a

TABLE 4

Relationships Between New Leaders Measures and Principal Retention

New Leaders Measure	(1)	(2)	(3)	(4)
	Same position in Year 2		Same district in Year 2	
Panel A: Constructs of competencies (Single regression)				
Construct 1 (Human capital)	0.0475 (0.0559)	0.0582 (0.0683)	0.0261 (0.0341)	0.0259 (0.0291)
Construct 2 (Cultural capital)	0.0928* (0.0407)	0.0807 [†] (0.0467)	0.100** (0.0249)	0.0933** (0.0199)
Construct 3 (Personal leadership)	0.0208 (0.0486)	0.0349 (0.0623)	0.0275 (0.0296)	0.0167 (0.0266)
Panel B: Standards (Separate regressions)				
Personal leadership	0.128* (0.0602)	0.137 [†] (0.0714)	0.135** (0.0470)	0.133* (0.0513)
Instructional leadership	0.0690 (0.0716)	0.0675 (0.0860)	0.0524 (0.0588)	0.0452 (0.0651)
Cultural leadership	0.105 [†] (0.0604)	0.109 (0.0777)	0.117* (0.0475)	0.109 [†] (0.0569)
Adult and team leadership	0.110 (0.0690)	0.111 (0.0874)	0.112 [†] (0.0552)	0.120 [†] (0.0638)
Operational leadership	0.0928 (0.0725)	0.137 [†] (0.0780)	0.102 [†] (0.0582)	0.141* (0.0557)
Controls				
New Leaders demographics	✓	✓	✓	✓
District fixed effects		✓		✓
Observations	37	37	37	37

Note. Coefficient cells in each column in Panel B are from five independently run models of retention. Demographic controls are indicators for African American, Hispanic, and female. New Leaders standards are standardized; coefficients represent a 1 *SD* increase in the standard. Constructs were made from underlying competency data. Standard errors in parentheses.

[†] $p < .1$. * $p < .05$. ** $p < .01$.

position or district, highlighting the multifaceted nature of the principal's job and career trajectory.

This study is relatively unique in its exploration of principal competencies and their relationship to principal turnover. Only one recent study, Grissom et al. (2017) study of principal licensure exam scores, looked at principal turnover. The authors found that principal scores on the School Leaders Licensure Assessment in Tennessee were not associated with turnover. However, results may differ because the authors only present relationships with the overall score, not specific domains of principal practice, and because knowledge of skills does not necessarily translate to a demonstration of competencies in those skills in practice.

Robustness and Sensitivity Checks

Robustness of Student Outcomes to Balanced Panel

Recall that the current results include all observations of a New Leaders principal after being placed in the partner districts. This sample is our preferred sample because of evidence that it may take up to 3 years for a principal to have detectable effects on student achievement (Corcoran et al., 2009). However, this sample also implicitly weights the earlier cohorts of New Leaders principals more because they have an opportunity to lead schools for a longer period of time.

We therefore rerun our main specification on a sample that only includes each resident's first

TABLE 5

Relationships Between New Leaders Measures and Student Outcomes, Balanced Panel

New Leaders Measure	(1)	(2)	(3)	(4)	(5)	(6)
	All principal-year observations (original results)			First-year principal observation (balanced panel)		
	ELA	Math	Attendance	ELA	Math	Attendance
Panel A: Constructs of competencies (Single regression)						
Construct 1 (Human capital)	0.033* (0.011)	0.044* (0.015)	0.006** (0.002)	0.013 (0.013)	0.025 [†] (0.013)	0.006** (0.002)
Construct 2 (Cultural capital)	-0.007 (0.011)	-0.023 (0.016)	0.003 (0.002)	0.017 (0.012)	0.017 (0.014)	0.002 (0.002)
Construct 3 (Personal leadership)	-0.004 (0.013)	-0.016* (0.006)	0.002 (0.002)	-0.010 (0.009)	-0.044* (0.015)	0.000 (0.002)
Panel B: Standards (Separate regressions)						
Standard 1 (Personal leadership)	0.001 (0.013)	-0.011 (0.008)	0.005 (0.003)	-0.002 (0.014)	-0.028 (0.025)	0.003 (0.003)
Standard 2 (Instructional leadership)	0.022 (0.013)	0.027 (0.015)	0.006* (0.002)	0.017 (0.009)	0.026* (0.008)	0.006* (0.002)
Standard 3 (Cultural leadership)	0.008 (0.010)	0.003 (0.011)	0.004 (0.003)	0.019 (0.013)	0.019 (0.016)	0.003 (0.003)
Standard 4 (Adult & team leadership)	0.041* (0.016)	0.029 [†] (0.013)	0.007* (0.003)	0.013 (0.013)	0.010 (0.013)	0.004 (0.003)
Standard 5 (Operational leadership)	0.022 (0.016)	0.011 (0.012)	0.005 (0.003)	0.007 (0.016)	0.001 (0.009)	0.003 (0.003)
Baseline ELA and Math Score	✓	✓		✓	✓	
Baseline attendance Observations			✓			✓
	28,489	28,489	51,803	16,281	16,281	28,146

Note. Standard errors clustered at school level. All models include cohort fixed effects, year fixed effects, and district fixed effects. New Leaders principal covariates include preresidency recruitment pathway, an indicator for passing the Emerging Leaders Program screening, years of experience as a teacher, gender, race/ethnicity, an indicator for the residency occurring in a charter school, an indicator for the residency occurring in a school led by a New Leaders alumnus from a previous cohort, and preresidency recruitment pathway evaluation scores (overall for factors or standard specific). Student covariates include fixed effects for grade, an indicator for having repeated a grade, classification as an English Language Learner, student race/ethnicity, gender, and an indicator for being old for the grade. School covariates include school enrollment, school level, and school-level averages of race/ethnicity, gender, English Language Classification, students repeating a grade, and students old for their grade. Constructs were made from underlying competency data. ELA = English language arts.

[†] $p < .1$. * $p < .05$. ** $p < .01$.

year in the school as principals such that results are based on a balanced panel. Table 5 shows that the relationships between measures and attendance outcomes are largely stable in this sample. A 1 *SD* increase in the Human Capital construct and the Instructional Leadership standard are both still associated with a 0.6 percentage point increase in attendance ($p < .01$ for the construct and $.05$ for the standard). However, the estimate on the Adult and Team Leadership standard drops

by about half and is not significant. The associations with student test scores are more muted. The relationship between Adult and Team Leadership construct and both test score outcomes are smaller and no longer significant and the relationship between the Human Capital construct and ELA scores are smaller and insignificant. Only the relationships between the Human Capital construct and math remain a marginally significant (0.025) and the negative relationship

between the Personal Leadership construct and math grows to a significant ($-0.044, p < .05$). Finally, the point estimate on the Instructional Leadership standard's relationship with math scores remains relatively stable at 0.026 but is now significant at the 5% level.

The more muted results are consistent with the evolving nature of the effect of principals on student outcomes. After 1 year, principals may have a limited effect on student test scores, thus muting any potential relationship between measures and outcomes. The stronger and more positive relationships between New Leaders measures and student outcomes in the sample that includes observations from principals from multiple years are consistent with the previous findings that principals need time to move student outcomes in a school. This highlights one challenge when analyzing preservice measures of principal skills—relationships with student outcomes can evolve quickly. Studies look at in-service measures (e.g., Grissom et al., 2018) may be less sensitive, as in-service principals include principals more established in their schools. These heterogeneous results suggest more research needs to be done on how associations between measures of principal practice and student outcomes evolve over time.

Sensitivity of Student Outcomes to New Leaders Resident Characteristics

Our current specification relies on controls for resident background characteristics to account for any sorting of residents to schools. These controls would be especially important if ratings of residents are associated with resident characteristics and if resident characteristics are associated with outcomes within the value-added framework. Online Supplemental Tables A2 and A3 show that in the full New Leaders sample, and the sample matched to district outcomes, some resident characteristics are related to the measures. Namely, differences by gender, race, and whether the residence took place charter school have significant relationships with the measures. Furthermore, Online Supplemental Table A4 shows that New Leaders recruitment pathway, teacher years of experience, and race have a relationship with student test score outcomes within the value-added framework.

Characteristics are less related to attendance outcomes in the value-added framework.

To interrogate the sensitivity of our main results to resident characteristics, we re-estimate our main models, but remove the resident characteristics as controls. As seen on Table 6, their exclusion reduces the relationship between the Human Capital construct by about a half and only the relationship with ELA achievement remains marginally significant. Similarly, their exclusion reduces the estimates on the standards by about half and all become insignificant. Attendance relationships are less sensitive. The estimate on the Human Capital construct remains the same in magnitude and significance, and the estimate on the Instructional Leadership construct is reduced by one third and is marginally significant. The more robust attendance estimates are likely due to the muted associations between characteristics and attendance in the value-added framework.

In total, the sensitivity analyses suggest that systematic differences in ratings by resident characteristics, in combination with potential sorting of principals to schools, could bias the relationship between resident measures and outcomes. Our results suggest that this bias is likely downward in direction given that controlling for observed characteristics increases the magnitude of the relationships. Of course, even with our controls, there is a possibility that pertinent unobserved differences in residents could be biasing our results, leaving the possibility open that these estimates are lower-bound estimates of the true relationship. This analysis highlights the need for researchers to explore not only the relationship between ratings and outcomes of interest, but also how ratings are assigned to principals and the potential sorting of principals to schools.

Sensitivity of Student Outcomes Alternate Specifications

As stated previously, one threat to identification is the possibility that principals sort to districts and schools based on the trends in district performance. To interrogate this possibility, we estimate two alternate versions of the model. The first check takes our main specification and adds a district specific linear time trend. The second check replaces the district and year fixed

TABLE 6

Relationship Between New Leaders Measures and Student Outcomes, Sensitivity to New Leaders Covariates

New Leaders Measure	(1)	(2)	(3)	(4)	(5)	(6)
	Original estimates			Estimates without resident controls		
	ELA	Math	Attendance	ELA	Math	Attendance
Panel A: Constructs of competencies (Single regression)						
Factor 1	0.033*	0.044*	0.006**	0.013	0.024 [†]	0.005*
(Human capital)	(0.011)	(0.015)	(0.002)	(0.009)	(0.012)	(0.002)
Factor 2	-0.007	-0.023	0.003	-0.003	-0.002	0.002
(Cultural capital)	(0.011)	(0.016)	(0.002)	(0.015)	(0.026)	(0.002)
Factor 3	-0.004	-0.016*	0.002	-0.013	-0.018	0.001
(Personal leadership)	(0.013)	(0.006)	(0.002)	(0.011)	(0.013)	(0.001)
Panel B: Standards (Separate regressions)						
Standard 1	0.001	-0.011	0.005	-0.006	-0.006	0.004
(Personal leadership)	(0.013)	(0.008)	(0.003)	(0.019)	(0.025)	(0.002)
Standard 2	0.022	0.027	0.006*	0.004	0.009	0.005 [†]
(Instructional leadership)	(0.013)	(0.015)	(0.002)	(0.011)	(0.015)	(0.002)
Standard 3	0.008	0.003	0.004	-0.004	0.001	0.003
(Cultural leadership)	(0.010)	(0.011)	(0.003)	(0.010)	(0.017)	(0.002)
Standard 4 (Adult & team leadership)	0.041*	0.029 [†]	0.007*	0.020	0.018	0.004 [†]
	(0.016)	(0.013)	(0.003)	(0.017)	(0.018)	(0.002)
Standard 5	0.022	0.011	0.005	0.001	0.004	0.003
(Operational leadership)	(0.016)	(0.012)	(0.003)	(0.017)	(0.024)	(0.002)
Baseline ELA and Math score	✓	✓		✓	✓	
Baseline attendance			✓			✓
Pre-APP score	✓	✓	✓			
New Leaders principal covariates	✓	✓	✓			
Observations	28,489	28,489	51,803	28,489	28,489	51,803

Note. Standard errors clustered at school level. All models include student covariates, school covariates, cohort fixed effects, year fixed effects, and district fixed effects. Student covariates include fixed effects for grade, an indicator for having repeated a grade, classification as an English Language Learner, student race/ethnicity, gender, and an indicator for being old for the grade. School covariates include school enrollment, school level, and school-level averages of race/ethnicity, gender, English Language Classification, students repeating a grade, and students old for their grade. Constructs were made from underlying competency data. ELA = English language arts.

[†] $p < .1$. * $p < .05$. ** $p < .01$.

effects with district-by-year fixed effects to non-parametrically account for district time trends. Online Supplemental Table A5 presents the original estimates in bold, with results from these two alternate specifications for each outcome. Estimates remain stable across specifications, with small differences in point estimates and very few changes in significance. Overall, district time trends do not seem to be biasing our results. These results, in combination with our other robustness checks, indicate that our

models likely take into account most sources of bias that would overturn our inferences.

Conclusion

This study investigates relationships between ratings of principal competencies in a residency-based preservice preparation program and student and principal outcomes. We find that certain ratings are predictive of principal VAMs of student achievement in ELA and math, principal

VAMs of attendance, the probability of being endorsed by the New Leaders program, the probability of being placed as a principal in partnering districts, and retention in the second year. These results indicate that the ratings have a useful screening and signaling function in practice (Goldhaber, 2007; Goldhaber & Hansen, 2010; Grissom et al., 2018). The relationship between higher ratings and a greater probability of obtaining a principalship in partnering districts suggests that the ratings are capturing competencies that districts and/or schools prioritize in the hiring process. The relationship between higher ratings and improved student outcomes and principal retention, conditional on placement, suggests that the ratings measure principal competencies that are valuable for the job.

In particular, the Human Capital construct and its constituent standards of Instructional Leadership and Adult and Team Leadership are most predictive of improvements in student outcomes. Student outcomes seem to be more sensitive to a principal's ability to set high expectations for students and teachers, to implement systems such as data-driven instruction and observation and supervision of instruction to ensure those goals are being met, and to develop the school staff through professional and leadership development activities. These results may not be surprising as these domains get to the mechanics of the core mission of teaching and learning. Similar domains have been identified as consequential for student outcomes in prior meta-analyses such as Waters et al. (2003) and which Grissom et al. (2021) posit are the most productive instructional-related activities. These results also suggest principal skills in other domains such as the building operations, the school and community culture, or the general setting of the mission and vision of the school are less relevant to student outcomes if they are further away from the mechanics of teaching and learning. The results are different for principal outcomes. Most standards and the Cultural Capital construct are related to principal retention and the Personal Leadership construct is associated with principal placement. These results suggest that districts take a holistic view a principal's skill-set in making hiring decisions and principals look at a variety of dynamics in making career decisions.

An evaluation of this program found that schools led by Aspiring Principal Program graduates outperformed other schools in the same district led by other new principals by 0.089 *SD* in math and 0.057 *SD* in ELA (Gates, Baird, Doss, et al., 2019; Gates et al., 2014) in 3 years. The associations with measures are about half the size, indicating that differences in these measured skills within the New Leaders program participants can account for a substantial portion of programmatic benefits. Other principal skills or characteristics emphasized by the program but not captured by these performance measures may have also contributed to the program's effects.

This study also highlights some methodological issues that may be especially pertinent to analyzing preservice measures of principal skills. First, we find that measures are related to principal background characteristics, and even in a value-added framework, student outcomes are related to those characteristics. Thus, fully accounting for that bias will be a challenge. These relationships have been seen in the literature before, most pertinently in Grissom et al.'s (2017) study of Tennessee principal licensure scores. There they found a correlation between scores and candidate demographic characteristics. These correlations could reflect true differences in ability, or they could reflect bias. If bias contributes to these correlations, which may be more likely in the third-party ratings in our context, our findings have implications for fairness and equity in the principal evaluations. More research needs to be done to understand the prevalence of this issue and how to fully account for this bias. Second, analyses of preservice measures may intuitively be analyzed with principal performance immediately upon placement in schools. However, consistent with research that shows it may take time for principals to affect student outcomes, our study shows that relationships between measures and student outcomes can grow within 3 years. Thus, more research needs to be done on how relationships evolve over time, particularly as principals gain experience and have time to impact schools.

This study enters a growing literature that has looked at the relationships between principal measures of skills and student and principal outcomes across the stages of the principal pipeline. Direct comparison of results across studies is

hampered by the lack of a standard set of practices. For example, Grissom and Loeb's (2011) study of principal self- and assistant principal ratings of principal competencies found that the organization management domain predicts teacher, parent, and student outcomes. However, the organizational management domain in that study involves instructional improvement practices, a major part of the Human Capital construct in this study. There is similar overlap, when looking at results on domains investigated in Tennessee (Grissom et al., 2018) and Pennsylvania (McCullough et al., 2016) and when comparing results to previous conceptions of school leadership such as those in Robinson et al. (2008).

Overlapping, but not aligned, results suggest more research needs to be done to understand which specific practices are predictive of policy-relevant outcomes, with an eye toward reconciling how different stakeholders define practices and domains and how relationships may vary based on when in the principal pipeline and when in a principal's career they are measured. To aid in this more systematic analysis of measures across contexts and systems, we compile the underlying competencies analyzed in this study into Online Supplemental Table A6. The competencies are linked to the five standards and the three constructs that exploratory factor analysis indicates underlies the data.

Two additional features of this study provide some evidence on the feasibility and effectiveness of a principal pipeline that aligns how principals are trained, selected, evaluated, and developed with common set of practices. First, these relationships are for measures that are used in both a formative and summative capacity. Fall versions of the measures were either not included in the final ratings calculation (Cohorts 14 and 15) or weighted less (Cohort 13). Thus, this study is a proof of concept that these measures can predict important outcomes when both the development of the candidate and the evaluation of the candidate are on the same measures. Of course, the stakes that are attached with providing an endorsement from New Leaders training programs are different from the stakes that are attached in evaluating in-service principals. Furthermore, too much reliance on a defined set of measures can lead to perverse incentives. However, in principle, measures—including job

evaluation measures—can take on this dual formative/summative role as well.

Limitations

The unique context of the study poses some limitations. New Leaders is relatively unique in that the organization invested the time and resources to purposively create these research-informed measures of competencies, integrate them into their preservice program such that it became a core feature of the residency experience, and then used the measures as a means of improving candidates' skills and as a final judgment of the candidate quality. Thus, how these measures would perform if adopted by other preservice programs is an open question. Furthermore, the districts in this study are large, urban districts that are focused on building a pipeline of effective principals. They intentionally partnered with New Leaders, in part because they felt the program objectives were aligned with the needs of their districts. New Leaders recruit both nationally and within participating districts. Applicants are then screened and selected before being invited into the program. Thus, these districts have a deeper labor market to draw from, allowing them a choice among candidates. This flexibility in hiring may not be available to schools and districts in smaller labor markets. Thus, this sample is not likely representative of all principal candidates in the nation. Rather, these candidates are more attuned to the urban environment and are predisposed to display the leadership qualities on which the New Leaders program is predicated. Furthermore, we only observe those graduates that are hired into partnering districts. Thus, the relationships between these measures and student outcomes could differ for principals who chose to work in other settings.

Although, in many ways, the above limitations represent a more "ideal" context, it is this more ideal context that facilitated a proof of concept that competencies can be identified, cultivated, and measured at scale at the earliest part of the principal pipeline as the first investments into principal candidates are made. Furthermore, as they predict policy-relevant outcomes, they can be useful in building and aligning a pipeline system for developing principal competencies.

Having established this proof of concept, further research can understand how such systems can be scaled across principal preparation programs, and how relationships vary in districts and schools in different contexts and with various relationships with preservice training providers.

Acknowledgments

We thank the anonymous reviewers for their insightful feedback that helped improve this manuscript. We also thank the staff from the nine districts who provided the data and Brenda Neuman-Sheldon, Marianna Valdez, Jodut Hashmi, and Jackie Gran from New Leaders who provided both the data and contextual knowledge needed to field the study. This project benefited from the management and data support provided by Mirka Vuollo, Isaac Opper, Diana Lavery, Mark Harris, Alyssa Ramos, Emilio Chavez-Herrerias, Juliana Chen-Peraza, Amanda Edelman, Crystal Huang, Ashley Muchow, Claudia Rodriguez, and Joshua Russell-Fritch.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The authors prepared the work as employees of RAND Corporation.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by a grant from the U.S. Department of Education (Grant No. U411B12002).

Notes

1. As part of the agreement between New Leaders and districts, those who secure positions in partnering districts are expected to serve for at least 5 years. However, principals themselves do not sign a formal commitment with the district, districts are sometimes not aware of this expectation, and rarely is it enforced.

2. The process of coordinating placements of residents into partnering schools varies across partner districts, but usually involves one New Leaders representative that works with the districts to find placements. New Leaders and the district have to navigate a variety of factors, including finding schools with an appropriate position available for the resident to hold during the residency, finding schools where principals were willing to host a resident and had sufficient autonomy to guide the practices of the school,

and finding placements for preservice principals from other programs that partner with the districts.

3. A precursor to estimating the above relationships is ensuring that the measures have sufficient variation in ratings and acceptable internal consistency. If the measures have limited variation, then from a policy perspective they would have limited use because they would be unable to distinguish differences in principal skills. If the measures do not have acceptable internal consistency, they may not reliably measure their intended constructs, making the relationships between these measures and policy-relevant constructs difficult to interpret. Please see Appendix B in the online version of the journal for an investigation into the variation and internal consistency of the measures. Results show sound internal consistency as measured by Cronbach's alpha and item-rest correlations.

4. Some program participants drop out before the end of their residency, meaning they were not measured on the preservice factors and standards. This is observationally rare (less than 5% of participants over the cohorts studied). We investigated the relationship between preresidency standards and dropout and found no significant correlation. We also found no correlation between preresidency standards and endorsement. These findings are not sensitive to model specification. Thus, we focus our discussion on preservice measures at the end of the residency.

5. A limitation of our study is that linking programmatic data to school district data will result in the censoring of candidates who were not placed into partnering districts. As part of our analysis, we look at how measures are related to the probability of New Leaders endorsement and the probability of leaving the program early. We also compare the characteristics of the sample matched to district data with the characteristics of all candidates.

6. Logistic regression models are used as a robustness check; coefficient estimates (marginal effects) and significance are similar to those estimated via linear probability model (LPM). LPM coefficients included for ease of interpretation.

7. While there are 216 program participants among the three study cohorts, 15 have incomplete gender or race/ethnicity data, both of which are key controls for the placement and retention analyses.

References

- Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Measuring test measurement error: A general approach. *Journal of Educational and Behavioral Statistics, 38*, 629–663.
- Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public*

- sector productivity: The case of school principals (NBER Working Paper 17803). www.nber.org/papers/w17803
- Brewer, D. J. (1993). Principals and student outcomes: Evidence from US high schools. *Economics of Education Review*, 12(4), 281–292.
- Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy*, 11(3), 283–309.
- Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review*, 3, 92–109.
- Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2009). *The New York City aspiring principal program: A school-level evaluation*. Institute for Education and Social Policy, New York University. <http://steinhardt.nyu.edu/scmsAdmin/uploads/003/852/APP.pdf>
- Darling-Hammond, L., LaPointe, M., Meyerson, D., Orr, M. T., & Cohen, C. (2007). *Preparing school Leaders for a changing world: Lessons from exemplary leadership development Programs*. Stanford Educational Leadership Institute, Stanford University.
- Davis, S., & Darling-Hammond, L. (2012). Innovative principal preparation programs: What works and how we know. *Planning and Changing*, 43(1/2), 25–45.
- Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics/Revue Canadienne d'économique*, 47(2), 634–663.
- Donaldson, M., Movrogordato, M., Dougherty, S., Al Ghanem, R., & Youngs, P. (2021). Principal evaluation under Element and Secondary Every Student Succeeded Act: A comprehensive policy review. *Education, Finance, and Policy*. <https://direct.mit.edu/edfp/article/16/2/347/97157/Principal-Evaluation-under-the-Elementary-and>
- Eberts, R. W., & Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, 7(3), 291–299.
- Fuller, E. J., & Hollingworth, L. (2014). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*, 50(3), 466–499.
- Fuller, E. J., Hollingworth, L., & Liu, J. (2015). Evaluating state principal evaluation plans across the United States. *Journal on Leadership Education*, 10(3), 164–192.
- Gates, S. M., Baird, M. D., Doss, C. J., Hamilton, L. S., Opper, I. M., Master, B. K., Tuma, A. P., Vuollo, M., & Zaber, M. A. (2019). *Preparing school leaders for success: Evaluation of New Leaders' aspiring principals program, 2012-2017*. RAND Corporation.
- Gates, S. M., Baird, M. D., Master, B. K., & Chavez-Herrerias, E. R. (2019). *Principal pipelines: A feasible, affordable, and effective way for districts to improve schools*. RAND Corporation.
- Gates, S. M., Hamilton, L. S., Martorell, P., Burkauer, S., Heaton, P., Pierson, A., Baird, M., Vuollo, M., Li, J. J., Lavery, D. C., Harvey, M., & Gu, K. (2014). *Preparing principals to raise student achievement: Implementation and effects of the New Leaders program in ten districts*. RAND Corporation.
- George W. Bush Institute. (2016). *Developing leaders: The importance—And the challenges—Of evaluating principal preparation programs*. <https://files.eric.ed.gov/fulltext/ED570672.pdf>
- Goldhaber, D. (2007). Everyone's doing it but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42, 765–794.
- Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47, 218–251.
- Grissom, J. A., Bartanen, B., & Mitani, H. (2019). Principal sorting and the distribution of principal quality. *AERA Open*, 5(2), 1–21.
- Grissom, J. A., Blissett, R. S. L., & Mitani, H. (2018). Evaluating school principals: Supervisor ratings of principal practice and principal job performance. *Educational Evaluation and Policy Analysis*, 40(3), 446–472.
- Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research*. The Wallace Foundation. <https://www.wallacefoundation.org/knowledge-center/Documents/How-Principals-Affect-Students-and-Schools.pdf>
- Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 37(1), 3–28.
- Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, 48(5), 1091–1123.
- Grissom, J. A., Mitani, H., & Blissett, R. S. L. (2017). Principal licensure exams and future job performance: Evidence from the School Leaders Licensure Assessment. *Educational Evaluation and Policy Analysis*, 39(2), 248–280.
- Grissom, J. A., Mitani, H., & Woo, D. S. (2019). Principal preparation programs and principal

- outcomes. *Educational Administration Quarterly*, 55(1), 73–115.
- Haller, A., Hunt, E., Pacha, J., & Fazekas, A. (2016). *Lessons for States: The Every Student Succeeds Act (ESSA) Increases Focus on and Investment in Supporting Principal Preparation and Development*. <https://eric.ed.gov/?id=ED572784>
- Hallinger, P. (2015). Instructional leadership and the school principal: A passing fancy that refuses to stay away. *Leadership and Policy in Schools*, 4(3), 221–239.
- Hallinger, P., & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement*, 9, 157–191.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal*, 86(2), 217–248.
- Herrmann, M., & Ross, C. (2016). *Measuring principals' effectiveness: Results from New Jersey's first year of statewide principal evaluation* (REL 2016–156). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid Atlantic. <http://ies.ed.gov/ncee/edlabs>
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, 4(3), 177–199.
- Leithwood, K., Louis, S. K., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. The Wallace Foundation. <https://www.wallacefoundation.org/knowledge-center/documents/how-leadership-influences-student-learning.pdf>
- Levin, S., & Bradley, K. (2019). *Understanding and addressing principal turnover: A review of the research*. Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/NASSP_LPI_Principal_Turnover_Research_Review_REPORT.pdf
- Louis, K. S., Leithwood, K., Wahlstron, K. L., Anderson, S. E., Michlin, M., Mascal, B., Gordon, M., Strauss, T., Thomas, E., & Moore, S. (2010). *Learning for leadership: Investigating the links to improved student learning*. The Wallace Foundation.
- McCullough, M., Lipscomb, S., Chiang, H., Gill, B., & Cheban, I. (2016). *Measuring school leaders' effectiveness: Final report from a multiyear pilot of Pennsylvania's Framework for Leadership* (REL 2016–106). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. <https://files.eric.ed.gov/fulltext/ED563446.pdf>
- Murphy, J., Goldring, E., Cravens, X., Elliott, S. N., & Porter, A. C. (2007). The Vanderbilt assessment of leadership in education: Measuring learning-centered leadership. *East China Normal University Journal*, 29(1), 1–10.
- National Policy Board for Educational Administration. (2015). *Professional standards for educational leaders*. http://www.npbea.org/wp-content/uploads/2017/06/Professional-Standards-for-Educational-Leaders_2015.pdf
- Rangel, V. S. (2018). A review of the literature on principal turnover. *Review of Educational Research*, 88(1), 87–124.
- Robinson, V. M., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44(5), 635–674.
- School Leaders Network. (2014). *CHURN: The high cost of principal turnover*. <https://www.issueclub.org/resources/20544/20544.pdf>
- Scott, D. (2018). *50-state comparison: School leader certification and preparation program*. Education Commission of the States. <https://www.ecs.org/50-state-comparison-school-leader-certification-and-preparation-programs/>
- Superville, D. R. (2014). States forge ahead on principal evaluation—Education Week. *Education Week*. https://www.edweek.org/ew/articles/2014/05/21/32principals_ep.h33.html
- University Council for Educational Administration. (2018). *National Educational Leadership Preparation (NELP) standards*. <http://www.ucea.org/initiatives/1523/>
- Waters, J. T., Marzano, R. J., & McNulty, B. A. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Mid-continent Research for Education and Learning.

Authors

CHRISTOPHER DOSS is an associate policy researcher at RAND. He is an economist of education that specializes in fielding causal and descriptive studies of educational policies and interventions. His research interests include principal and teacher preparation, early childhood education, behavioral economics, educational technologies, and STEM (science, technology, engineering, and mathematics) education. He is also interested in social and emotional learning

and other student outcomes beyond traditional measures of student achievement.

MELANIE A. ZABER is an associate economist at RAND and co-director of the RAND Lowy Family Middle-Class Pathways Center. Her research has examined household transitions (coresidence, marriage, divorce, bankruptcy), analyzed workforce pipelines (principals, military linguists, building tradespeople), and explored postsecondary finance (market power, state grant aid, student debt). Her research has been funded by the National Science Foundation, the Institute for Civil Justice, and the Social Security Administration. Current projects include an exploration of the persistence of women in STEM careers and an analysis of the long-term education and career outcomes of participants in a high school youth development program.

BENJAMIN K. MASTER is a policy researcher at RAND. His research interests lie in school leadership, human capital development, organizational studies, school and teacher effectiveness, and implementation research.

SUSAN M. GATES is a senior economist at RAND and professor in the Pardee RAND Graduate School. She applies economic theory and methods to help policymakers identify effective practices and make better decisions on a wide range of topics. Her recent education policy research examines different facets of the work that principal preparation programs, school districts, and states are doing to improve school leadership.

LAURA S. HAMILTON is associate vice president of Research Centers at ETS, overseeing a portfolio of work on assessment and learning in K–12, higher education, and career development. Her research explores innovative, equitable approaches to measuring student outcomes and learning opportunities in social, emotional, academic, and civic-learning domains.

Manuscript received September 14, 2019

First revision received August 31, 2020

Second revision received April 1, 2021

Accepted April 13, 2021