

**PRINCIPAL EVALUATION UNDER THE
ELEMENTARY AND SECONDARY EVERY
STUDENT SUCCEEDS ACT: A
COMPREHENSIVE POLICY REVIEW**

Morgaen Donaldson

(corresponding author)
Neag School of Education
University of Connecticut
Storrs, CT 06269
morgaen.donaldson@uconn
.edu

Madeline Mavrogordato

College of Education
Michigan State University
East Lansing, MI 48824-1034
mavro@msu.edu

Shaun M. Dougherty

Peabody College
Vanderbilt University
Nashville, TN 37203
shaun.dougherty@vanderbilt
.edu

Reem Al Ghanem

Department of Educational
Leadership
University of Connecticut
Storrs, CT 06269
ghanemra@bu.edu

Peter Youngs

Curry School of Education
University of Virginia
Charlottesville, VA 22904
pay2n@virginia.edu

Abstract

A growing body of research recognizes the critical role of the school principal, demonstrating that school principals' effects on student outcomes are second only to those of teachers. Yet policy makers have often paid little attention to principals, choosing instead to focus policy reform on teachers. In the last decade, this pattern has shifted somewhat. Federal policies such as Race to the Top (RTTT) and Elementary and Secondary Education Act waivers emphasized principal quality and prompted many states to overhaul principal evaluation as a means to develop principals' leadership practices and hold them accountable for the performance of their schools. The development and dissemination of principal evaluation policies has proceeded rapidly, however, it is unclear whether focusing on principal evaluation has targeted the most impactful policy lever. In this policy brief, we describe where policy makers have placed their bets in post-RTTT principal evaluation systems and comment on the wisdom of these wagers. We describe the degree to which principal evaluation components, processes, and consequences vary across the fifty states and the District of Columbia, and review evidence on which aspects of principal evaluation policies are most likely to improve principals' practice and hold them accountable.

https://doi.org/10.1162/edfp_a_00332

© 2020 Association for Education Finance and Policy

INTRODUCTION

A growing body of research recognizes the critical role of the school principal, demonstrating that school principals' effects on student outcomes are second only to those of teachers (Hallinger and Heck 1998; Waters, Marzano, and McNulty 2003; Robinson, Lloyd, and Rowe 2008; Branch, Hanushek, and Rivkin 2012; Coelli and Green 2012; Dhuey and Smith 2014; Grissom, Kalogrides, and Loeb 2015). Yet federal and state policy makers have often paid little attention to principals, choosing to focus on teachers as the prime target of their policies. In the last decade, this pattern has shifted somewhat. Federal policies, such as Race to the Top (RTTT), the Elementary and Secondary Education Act (ESEA) waivers, and the recent reauthorization of ESEA as the Every Student Succeeds Act, emphasized principal quality and prompted many states to overhaul principal evaluation as a means to develop principals' leadership practices and hold them accountable for the performance of their schools (Jacques, Clifford, and Hornung 2012; CEP 2014). The development and dissemination of principal evaluation policies has proceeded rapidly and, we argue, haphazardly over the past decade. In focusing on principal evaluation, have policy makers placed their bets on the right lever? Moreover, do the policies emphasize the right things?

In this policy brief, we describe where policy makers have placed their bets in post-RTTT principal evaluation systems and comment on the wisdom of these wagers. We describe the degree to which principal evaluation components, processes, and consequences vary across the fifty states and the District of Columbia, and review evidence on which aspects of principal evaluation policies are most likely to improve principals' practice and hold them accountable. This brief is the first to comprehensively catalog principal evaluation policies put in place as a result of federal policies that encouraged principal evaluation reform over the past decade. We show that principal evaluation policies include some promising evaluation practices, such as goal setting and, to a lesser degree, stakeholder surveys (Locke and Latham 2002). Thus, contemporary principal evaluation policies appear more robust than their predecessors.

THE POLICY CONTEXT FOR PRINCIPAL EVALUATIONS IN THE U.S.

In the United States, longstanding approaches to principal evaluation often had little relationship to instructional leadership or student achievement, thus making them a poor bet for improving principal quality (Goldring et al. 2009). In 2009, Goldring and colleagues examined principal evaluation instruments in sixty-five urban school districts in over forty states. They found these instruments primarily focused on school leaders' efforts to establish rigorous learning goals, promote teacher professional community, and hold school staff accountable for student learning. However, these instruments paid relatively less attention to principals' efforts to implement ambitious curricula or monitor instructional quality. In addition, most principal evaluation systems were used for formative purposes and few used evaluation criteria based on standards or evidence (Goldring et al. 2009). Most principals were unclear about the purpose of principal evaluation, did not feel their evaluations were useful, and perceived them to have little impact on their motivation or performance (Thomas, Holdaway, and Ward 2000; Reeves 2005; Davis et al. 2011). As of 2009, principals' evaluation track records suggest it is a poor choice for policy makers seeking a high-leverage policy to improve the quality of principals.

PRINCIPAL EVALUATION IN FEDERAL POLICY

Recent federal reforms have tried to improve principal evaluation to address many of the weaknesses outlined above. Marking a shift from historical approaches, RTTT and ESEA waivers encouraged principal evaluation systems to assess principal performance based on student achievement and school leadership behaviors. One objective of these new systems is to more closely connect the work of principals to improvement of student learning. To meet this aim, RTTT and ESEA waivers required states to develop, adopt, and implement principal evaluation and support systems that differentiate between principal performance based on their effectiveness level (USDOE 2009, 2011). Under ESEA waivers, for example, a minimum of three performance levels is required. Summative annual evaluation is mandated under RTTT but only recommended under ESEA. ESEA recommends principals be evaluated annually in the first three years of service and at least once every three years after that.

Both RTTT and ESEA waivers provide only minimal guidelines to states for the design of principal evaluation systems. Although both policies mandate that principals be evaluated using multiple measures (USDOE 2009, 2011), neither policy specifies what percentage of the final summative score should be tied to student achievement and growth. Furthermore, both policies require states to develop and implement student achievement and growth assessment measures, but they do not specify the measures that must be used. RTTT, however, requires that the chosen assessment tools have high technical quality (i.e., be fair, valid, reliable, and aligned to standards) and that, for tested grades, at least one of the used measures be the state's standardized assessment (USDOE 2009).

Both RTTT and ESEA waivers recommend, but do not require, that states include measures of principal leadership skills and practices. These measures can be supplemented by other measures, such as high school graduation or teacher retention rate (USDOE 2009, 2011). Relatedly, neither policy specifies the number of observations required when observations are used in the evaluation of principals.

The two policies also include the provision of feedback but differ in what they require. Both RTTT and ESEA waivers require that the principal evaluator provide timely and constructive feedback, and that principal evaluation results are used to inform decisions such as professional development, compensation, promotion, and tenure status (USDOE 2009, 2011). ESEA waivers differ from RTTT in that they require principals be involved in the evaluation process (not just recipients of summative feedback), but neither policy specifies whether principals need to complete self-assessments.

We now turn to state policies on principal evaluation. Based on our policy scan, we summarize these policies across the fifty states and Washington, DC, and draw on research to comment on the features of these policies. We suggest ways on how policy makers could improve upon them.

States' Principal Evaluation Policies in the Wake of Reforms

Almost all states have enacted new principal evaluation policies since 2009.¹ Specifically, between 2009 and 2018, fifty out of fifty-one states (98 percent) including

1. We analyzed the principal evaluation systems in all fifty states. We started by gathering publicly available documents pertaining to principal evaluation policies. When no documents were publicly available, we reached out to the Department of Education in the state and requested the documents. Documents included handbooks or manuals, state statutes, state standards, and evaluation system Web sites. We then followed an iterative process,

Washington, DC, enacted new policies related to principal evaluation and/or revised such policies. The trend of passing new or revised legislation accelerated across this period. For instance, in 2010, just two states passed or revised policies (Maryland and Missouri), whereas ten states passed/revised policies in 2017 and eight did so in 2018 (see table A.1 in the online appendix, which is available on *Education Finance and Policy's* Web site at https://doi.org/10.1162/edfp_a_00332). Thus, the majority of mandates for new and revised principal evaluation policies have only become relevant to practice in the past two or three years.

Broad patterns suggest that most states devolved authority for principal evaluation policy design to districts. This is perhaps the most common similarity across states but not the only one worth noting. Forty-three states (84 percent) permit school districts to develop their own principal evaluation systems as long as their systems are consistent with state policy requirements. In thirty-one states (61 percent), all principals, including both probationary and non-probationary school leaders, are evaluated each year.² Another fourteen states (27 percent) are simply more specific in how they differentiate the frequency of evaluations; probationary principals evaluated annually, while non-probationary evaluated every two to four years. In twenty-seven states (53 percent), principals must be evaluated by the district superintendent (or designee), the assistant superintendent, or other district administrators (i.e., those who supervise principals). The remaining states do not specify who is responsible for evaluating principals. Finally, principal evaluator training is explicitly required in twenty-six states (51 percent), recommended in four states (8 percent), but not mentioned at all in principal evaluation policies in twenty-one states (45). In online appendix table A.1, we provide an overview of the context of principal evaluation systems across states.

Components of Evaluation

Much of the innovation and variation in state policies for principal evaluation arises from whether and how they use specific components to evaluate principals. We report these different components in the first panel of table 1 and online table A.2, and briefly summarize them here.

Many states include measures of leadership practice and student achievement in their principal evaluation policies, though few states assign a weight for these components. All states require that districts assign principals summative ratings based on

similar to that used by Steinberg and Donaldson (2016), to gather information about the key aspects of the components, processes, and consequences of principal evaluation systems across all states. We started the document analysis by designing a preliminary matrix that reflected potential components and processes of principal evaluation based on prior research. We used this matrix to gather information about the key aspects of the principal evaluation system in a pilot state—Connecticut—which is part of a larger study on principal evaluation that also includes Tennessee and Michigan. We then refined, eliminated, and added categories. We tested the revised matrix by using it to analyze the evaluation systems in Michigan, Tennessee, and Washington, DC, contexts that differ with regard to collective bargaining context and district enrollment. The pilot process resulted in more revisions and further refining of the matrix. To ensure accurate analysis of the documents, two team members used the final matrix to code each state's principal evaluation policy documents. Initial agreement levels were 0.8, which is often considered an acceptable degree of agreement (Landis and Koch 1977). The two team members met to discuss and resolve any discrepancies between their codes and consulted with the remaining team members as necessary. In the following section, we summarize the findings of this analysis. Note that though Washington, DC, is not a state, we follow prior policy work and include it in the analysis and findings.

2. Probationary leaders are those who do not yet have tenure (in states that tenure school leaders) or have not yet been formally reconsidered for a contract extension (mostly in states that do not grant tenure for leaders).

Table 1. Key Components, Processes, and Consequences of Principal Evaluation Systems across States ($N = 51$)

	Required		Recommended		No Information	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Components of Evaluation						
Student outcome	46	90	3	6	2	4
Teacher effectiveness	4	8	2	4	45	88
Leadership skills and practices	50	98	1	2	0	0
Stakeholders surveys ^a	14	27	28	55	11	22
Other components	7	14	2	4	42	82
Processes of Evaluation						
Goal setting	44	86	3	6	4	8
Mid-year evaluation	28	55	9	18	14	27
Self-assessment	30	59	10	20	11	22
End-of-year evaluation	47	92	0	0	4	8
Observations	34	67	9	18	8	16
In-person meeting(s)	37	73	11	22	10	20
Consequences of Evaluation						
Exemplary rating	6	12	6	12	39	76
Effective rating	6	12	1	2	44	86
Developing rating	21	41	2	4	28	55
Ineffective rating	30	59	2	4	19	37

Notes: Some percentages add up to more than 100 due to rounding errors or due to variables being partly required and partly recommended in some states.

^aSurveys are prohibited in one state, New York, 2 percent.

multiple measures of their performance. Virtually all states ($N = 50$; 98 percent) require and one state recommends a leadership and practice measure in their principal evaluation system. However, only twenty-nine states (57 percent) specify the weight of this component in the principal’s final summative rating. The leadership and practice weight ranges substantially from 15 percent to 100 percent, with the median leadership skills and practices component accounting for about 55 percent to 57 percent of the final summative rating and 50 percent being the modal weight.

In a meaningful break from past principal evaluation systems, forty-six states (90 percent) require and three states (6 percent) recommend that districts include a student outcomes component. Yet, only twenty-nine states (58 percent) specify the weight of the student outcomes component in the principal’s final summative rating, with ranges between 20 and 50 percent and a mode of 50 percent.

Stakeholder surveys (principal performance surveys completed by teachers, parents, students, etc.) are required by fourteen states (27 percent) and twenty-eight additional states (55 percent) recommend them. When used, such surveys are commonly utilized to collect data on principal performance even when districts are not required to use such evidence in summative principal evaluation. In table 2, we present a more detailed look at how stakeholder surveys are used. Stakeholder surveys carry independent weight in the final summative evaluation of principals in only four states (8 percent),

Table 2. Summary of Stakeholder Surveys in Principal Evaluation Systems across States ($N = 51$)

Status	Teacher Survey		Student Survey		Parent Survey		Community Members Survey		Other/Unspecified Survey	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Required	14	27	6	12	6	12	1	2	1	2
Allowed	22	43	21	41	22	43	11	22	6	12
Prohibited	—	—	1	2	1	2	—	—	—	—

Note: The one state that requires other surveys, Mississippi, requires surveys in the form of self-evaluation and supervisor evaluation.

and account for 10 to 30 percent of the summative rating. Only four states (8 percent) require and two states (4 percent) recommend that districts include a teacher effectiveness component in their principal evaluation system, accounting for 5 percent to 15 percent of the principal's final summative rating (see online table A.3 for further detail).

These components of new principal evaluation systems diverge from earlier ones in several ways that make them more promising than their predecessors. First, new systems are much more likely than their predecessors to be based on standards for effective leadership that link principals' actions with desired outcomes (Clifford and Ross 2012). Both RTTT and ESEA waivers encouraged states to develop new principal evaluation systems based on the Professional Standards for Educational Leaders, which address several areas of leadership including: (1) curriculum, instruction, and assessment; (2) equity and cultural responsiveness; (3) professional capacity of school personnel; and (4) ethics and professional norms (NPBEA 2015). The movement to ground principal evaluation in professional standards is a promising one (Clifford, Hansen, and Wraight 2014). Standards-based evaluation systems can reduce subjectivity and provide a more valid assessment of principal effectiveness (Kimball and Milanowski 2009). Also, these systems usually require multiple sources of evidence that attest to educators' performance and considerable evaluator training. Although research examining standards-based principal evaluation is scarce, multiple studies on teacher evaluation have shown that standards-based evaluation scores have moderate positive correlations with student achievement (e.g., Kimball et al. 2004; Milanowski 2004; Milanowski, Kimball, and White 2004).

Second, some systems incorporate feedback on school leader performance from key stakeholders, including teachers, peers, and supervisors. Stakeholder feedback is a promising element in principal evaluation in that it can provide useful, formative feedback (Goldring, Mavrogordato, and Haynes 2015). However, caution is warranted, as very few instruments for stakeholder feedback have been validated (Clifford, Hansen, and Wraight 2014).

Third, starting in 2009 with RTTT and later with ESEA waivers, states have been expected to use student achievement data in evaluating principals (Doherty and Jacobs 2013). The evidence on including student performance in principals' evaluations is mixed. Research indicates that principals affect student test score performance in their school (Dhuey and Smith 2014). However, some have articulated challenges to incorporating student performance in principals' evaluations. These include systematic sorting

of students to schools, measurement error, and constraints on principals' control over the quality of teaching and learning students receive (Grissom, Kalogrides, and Loeb 2015). And yet, over 90 percent of states include measures of principal practice and student outcomes in principals' ratings. Both measures are weighed heavily in summative scores, marking a major shift from historical approaches to principal evaluation.

Fourth, some states encourage school districts to combine summative principal evaluation with efforts to closely supervise school leaders and support their professional growth (Clifford and Ross 2012). For example, some districts are using principal goal setting to engage principals in continuous improvement cycles.³ Often, this is accompanied by a reduced emphasis on school leaders' compliance with district priorities and an increased focus on mentoring and coaching of principals (Anderson and Turnbull 2016; Kimball et al. 2015). In fact, some large districts have shifted the role of principal supervisor from ensuring principals' compliance with district mandates to supporting principals' growth as instructional leaders (Rogers et al. 2019).

Process of Evaluation

As with the components of evaluation, states require some longstanding processes and introduce other, new procedures to principal evaluation (second panel of table 1). Forty-four states (86 percent) require and three states (6 percent) recommend that school leaders engage in goal setting/plan development as part of the principal evaluation process and forty-seven states (92 percent) require them to participate in an end-of-school-year summative written evaluation.⁴ More than half of states require principals to engage in self-assessment ($N = 30$; 59 percent) while ten states (20 percent) recommend it.⁵ As for mid-year evaluation, twenty-eight states (55 percent) require it and nine states (18 percent) recommend it. Finally, the majority of states ($N = 34$; 67 percent) require school leaders to be observed as part of the principal evaluation process, while nine states (18 percent) only recommend it (see table 3, as well as online tables A.4 and A.5 for further details).

Unlike the old single-source evaluation systems, new systems incorporate many different types of evidence on principal leadership behavior and student performance (Henry and Guthrie 2015; Grissom, Blissett, and Mitani 2018). School districts collect data on principals' goals for themselves and their schools as well as their efforts to meet these goals. When structured to reinforce principals' sense of autonomy and competence, goal setting can support principals' intrinsic motivation (Locke and Latham 2002). Moreover, principals appear to value evaluation systems that include goal setting, reflection, and constructive feedback (Sanders 2008; Chacon-Robles 2018). Principals report that the goal setting in evaluation is useful and worthwhile, especially when

3. As described in Anderson and Turnbull (2016), goal settings is a process where principals are asked to identify areas for improvement, collaborate with their supervisors to choose and define targets for improvement that align with standards, and include a plan to achieve them.
4. An end-of-school-year summative written evaluation is generally a summative rating and evaluation report that is delivered at or after the close of the school year. This contrasts with mid-year evaluations that are meant to be formative, with the potential to redirect activities or focus in the latter half of the year.
5. Self-assessment is meant to be a reflective process whereby individuals review their goals and use their own data and analysis to gauge their progress toward achieving those goals, and to arrive at an overall conclusion about their performance.

Table 3. Summary of the Types of In-person Meetings Required or Recommended in Principal Evaluation Systems across States ($N = 51$)

Status	Goal Setting Meeting		Pre-observation Meeting		Post-observation Meeting		Mid-year Meeting		End-of-year Meeting		Other Meeting	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Required	25	49	6	12	13	25	23	45	33	65	6	12
Allowed	2	4	3	6	6	12	1	2	0	0	3	6
No information	24	47	42	82	32	63	27	53	18	35	42	82

Notes: A pre-evaluation meeting is required in Hawaii and Illinois. A Circle Survey meeting is required in Mississippi. Quarterly meetings are required for novice principals in New Hampshire. Unspecified evaluation meetings are required in Utah. Formative evaluation meetings are allowed in Maine. Frequent meetings are recommended in Washington. Frequent meetings throughout the cycle are recommended in Wyoming.

aligned with district-implemented leadership standards, discussed with their evaluator, monitored, and assessed at the end of the year (Sanders 2008). Moreover, they indicate that goal setting helps them maintain a focus on their areas of improvement notwithstanding their workload (Chacon-Robles 2018).

Goal setting appears to be particularly beneficial when accompanied by other reflective practices often encouraged in current principal evaluation systems, self-assessment, and feedback. When principals reflect on their own work, they become more aware of their strengths and weaknesses, which can inform their growth plan and increase their focus on activities related to leadership skills and practices (Alimo-Metcalfe 1998; White, Crooks, and Melton 2002). For example, Sanders (2008) reported that all principals in that sample valued self-reflection and considered it an important factor in the development of their leadership skills and practices.

Another salient practice in new principal evaluation systems is frequent feedback provided by the principal's supervisor. Under traditional approaches to principal evaluation, experienced school leaders were rarely assessed and seldom received feedback that could enhance their leadership (Reeves 2005; Kimball, Milanowski, and McKinney 2009). The infrequent nature of principal evaluation also meant that it did little to facilitate leadership improvement or hold school leaders accountable for their performance, including decisions about promotion or dismissal. In contrast, new systems feature annual ratings of principal performance based on evidence about leadership behavior and student learning (Henry and Guthrie 2015; Grissom, Blissett, and Mitani 2018). These attributes of current approaches increase the likelihood that districts can use evaluation data to support principal growth and improvement and to inform promotion and dismissal decisions. Supporting growth can happen by providing principals with more frequent formative feedback throughout the year instead of exclusively providing summative feedback at the end of the year (Burkhauser et al. 2013). Feedback has the potential to improve principal performance, especially when accompanied with training and professional development (Locke and Latham 2002; Burkhauser et al. 2013).

Consequences of Evaluation

In perhaps the clearest departure from prior principal evaluation systems, most states require districts to assign one of four performance ratings to principals based on their

evaluation results and lay out clear consequences for performing below standard (third panel of table 1). Thirty states (59 percent) require and two states (4 percent) recommend that districts attach consequences to “ineffective” ratings, while twenty-one states (41 percent) mandate and two states (4 percent) recommend consequences for “developing” ratings. The consequences for these two levels of performance are often the same and typically include a remediation plan, more frequent observations and evaluations, intensive intervention, and dismissal if poor performance persists. Six states (12 percent) require and one state (2 percent) recommends that districts attach positive consequences to effective ratings; six states (12 percent) mandate and six states (12 percent) recommend positive consequences for highly effective ratings. Positive consequences for these two levels of performance often include fewer observations, longer evaluation cycles, additional leadership roles, promotions, additional compensation, and public commendation or other acknowledgement (see online table A.6).

In previous decades, the results of principal evaluation had few consequences for school leaders (Reeves 2005). Districts generally did not make connections between evaluation data and efforts to identify professional development opportunities or design improvement plans for principals. In addition, formal evaluation rarely led to school leaders losing their positions. Current evaluation systems have sought to increase the consequences of evaluation. States and districts are beginning to use the results of the principal summative evaluation to hold principals accountable for meeting their objectives, make professional development decisions, determine principal termination, and make salary decisions (Goldring et al. 2009; Kimball, Heneman, and Milanowski 2007; White et al. 2012).

There is more research on some of these consequences than others. Studies of efforts to align principal pay with performance have not produced evidence that pay for performance is associated with improved principal leadership skills and practices or increased student achievement (Hamilton et al. 2012; Marsh et al. 2011; Matarazzo 2014). For example, Hamilton et al. (2012) studied the effects of the Pittsburgh Principal Incentive Program on Pittsburgh’s public-school principals and found that the opportunity to earn an annual permanent salary increase of up to \$2,000 and a bonus of up to \$10,000 caused no change in the average principal performance, as measured by the principal evaluation rubric in the district. Most participating principals reported that the opportunity to earn merit pay did not motivate them to change their leadership practices (Hamilton et al. 2012).

IMPLICATIONS FOR POLICY MAKERS

When policy makers seized on principal evaluation as a policy lever to improve the quality of the nation’s principals, they were making a risky bet. It may have simply been that they doubled down on the importance of educator effectiveness and chose to reform principal evaluation at the same time as they recast teacher evaluation. However, prior to RTTT’s 2009 enactment, there was little evidence that principal evaluation provided principals or districts with useful information, spurred improvements to principals’ practice, or enhanced the quality of schools. Policy makers responded to this information by making several key changes to principal evaluation. New systems based on evidence of leadership behavior and student achievement have the potential

to strengthen leadership practice and school performance. In particular, districts that incorporate multiple measures into their evaluations of school leaders seem likely to promote a greater focus on student learning while holding leaders accountable for the performance of their schools. To date, research and the popular press have described general trends toward promising components and processes in principal evaluation. However, there is little research regarding the efficacy of content, process, and consequences of new principal evaluation systems at the state level. As such, this brief provides the first evidence on the actual prevalence of these new components and processes.

This policy brief surfaces several key implications for policy makers. First, whereas more than 90 percent of all states have enacted principal evaluation policies that include measures of student academic outcomes and more than two thirds require observations of principals, researchers have noted the need for further development of such measures. In particular, scholars have identified challenges to incorporating student achievement data into principal evaluations (Grissom, Blissett, and Mitani 2012) and there has been little research on the technical properties of principal observation instruments (Clifford, Hansen, and Wraight 2014). This suggests that policy makers may want to consider placing less emphasis on both of these components of principal evaluation.

Second, principal goal setting is required in 86 percent of states and research finds that principals value this process, especially when it is linked with district leadership standards and combined with opportunities to reflect on their practice (Sanders 2008; Chacon-Robles 2018). In addition, districts can combine goal setting and other summative approaches to principal evaluation with supervision of school leaders in ways that support their professional growth (Clifford and Ross 2012; Anderson and Turnbull 2016). This suggests that policy makers may want to continue to emphasize goal setting as a key process in principal evaluation and consider ways to encourage principals to select professional development opportunities that align with their goals.

Third, 92 percent of states require districts to provide principals with an end-of-year summative written evaluation, and 55 percent require them to provide school leaders with a mid-year evaluation. Research indicates that principals value constructive feedback (Sanders 2008; Chacon-Robles 2018) and that such feedback can potentially lead to improvements in principal performance, especially when combined with opportunities for professional development (Locke and Latham 2002; Burkhauser et al. 2013). This suggests that policy makers may want to look for ways to increase the likelihood that school leaders will receive constructive and timely feedback as part of the principal evaluation process.

Finally, 59 percent of states require that districts impose consequences when principals receive developing or ineffective ratings, but there is little research on whether such consequences lead to improvements in principal performance. In addition, 12 percent of states require positive consequences when principals receive effective or highly effective ratings, but research has shown no relationship between principals receiving monetary bonuses and either (1) changes in their knowledge and skills or (2) increases in student achievement (Marsh et al. 2011; Hamilton et al. 2012; Matarazzo 2014). This suggests that policy makers may want to consider placing less emphasis on the consequences of principal evaluation.

CONCLUDING THOUGHTS

Empirical evidence demonstrating the importance of principals in promoting favorable educational outcomes is growing and almost all states in the nation have instituted policies changing the nature of principal evaluation. Has this flurry of activity been worth it? Compared with its counterpart for teachers, principal evaluation may be an easier policy lever for districts and states to influence because collective bargaining is less common among principals, and, when principals unionize, they are less powerful than are teachers' unions. In this policy brief, we document the components, processes, and consequences of principal evaluation across the entire United States, highlighting variation in the presence or weight of different elements across states. This is one of the first comprehensive summaries of state policies on principal evaluation and provides next steps for policy makers and researchers.

We find that principal evaluation has been the subject of intense policy change in the past decade. Almost every state has instituted a new principal evaluation policy since 2009, with many new policies coming online in 2017 and 2018. In this new wave of policies, principals continue to be evaluated on measures of leadership, and this measure is often the most heavily weighted of all components of principal evaluation. In a clear improvement over earlier policies, these measures of leadership are more likely to be standards-based. The presence of standards-based leadership measures in principal evaluation policies bodes well for their likelihood of improving principals' practice. Moreover, consistent with the requirements of RTTT and ESEA waivers, we also find that most principals in the United States are also now evaluated on the basis of student performance, which aligns closely to the parallel process of teacher evaluation reform over this same period. This is a substantial change from conventional principal evaluation and may increase principals' focus on student outcomes.

The presence of stakeholder surveys is also a clear deviation from earlier iterations of principal evaluation policies. However, stakeholder surveys play a smaller role than do measures of leadership or student performance, with only 27 percent of states requiring them compared to the 90 percent that require measures of student performance and measures of leadership. This may be a missed opportunity for schools to become more engaged with and responsive to stakeholders.

Policy makers have placed a clear bet on principal evaluation in recent years. Given their investment in this policy lever, policy makers and researchers alike should examine how school districts have implemented these policies, whether these modifications have altered principal behavior, and if they have influenced student outcomes. Other, contemporaneous developments also warrant examination. For instance, what are the implications of these changes to principal evaluation policy in the context of expanded collective bargaining for principals? Are new principal evaluation policies affecting the supply of principals entering or remaining in this role? Policy makers would be wise to be cognizant of these potential ripple effects of recent principal evaluation policies.

Our findings also raise more specific questions about the composition of new principal evaluation policies. First, the role of observations in principal evaluation deserves careful attention. Only two thirds of states require that principals be observed, raising the possibility that evaluators are gathering other forms of data on which to make judgments regarding principals' practice. This may be understandable because leaders' work is broader than that of teachers. Drawing a valid and reliable inference about

principals' leadership skills based on observed practice may be even more difficult than doing so for teachers. This is especially true if, in the absence of observations, it is unclear what indicators are being used by evaluators when assessing principal quality.

A second salient consideration relates to the role of feedback in principal evaluation. Only 73 percent of states require evaluators to meet with principals to provide feedback. This raises questions about the quality of information principals receive on their performance and whether or not they receive recommendations and guidance regarding how to improve their skills.

Our investigation of the dimensions and emphasis of recent reforms in principal evaluation policy reveals that there is substantial variation across states, the implications of which are not yet fully understood. Further investigating the elements of principal evaluation should help to shed light on whether the policy activity of the last decade has led to improvement in principal practice, retention of effective principals, and increases in student performance.

ACKNOWLEDGMENTS

Funding for this paper was provided by a grant from the U.S. Department of Education's Institute of Education Sciences (R305A160100). All opinions expressed in this paper represent those of the authors and not necessarily the institutions with which they are affiliated or the U.S. Department of Education. All errors are solely the responsibility of the authors.

REFERENCES

- Alimo-Metcalf, Beverly. 1998. 360 degree feedback and leadership development. *International Journal of Selection and Assessment* 6(1): 35–44.
- Anderson, Leslie M., and Brenda J. Turnbull. 2016. *Building a stronger principalship: Volume 4. Evaluating and supporting principals*. Available <https://files.eric.ed.gov/fulltext/ED570471.pdf>. Accessed 19 November 2020.
- Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2012. Estimating the effect of leaders on public sector productivity: The case of school principals. NBER Working Paper No. 17803.
- Burkhauser, Susan, Susan M. Gates, Laura S. Hamilton, Jennifer J. Li, and Ashley Pierson. 2013. *Laying the foundation for successful school leadership*. Santa Monica, CA: RAND Corporation.
- Center on Education Policy (CEP). 2014. *Federal education programs: NCLB/ESEA waivers*. Available www.cep-dc.org/index.cfm?DocumentSubTopicID=48. Accessed 3 November 2020.
- Chacon-Robles, Brenda. 2018. Improving instructional leadership: A multi-case study of principal perspectives on formal evaluations. PhD dissertation, University of Texas at El Paso.
- Clifford, Matthew, Ulcca Joshni Hansen, and Sara Wraight. 2014. *A practical guide to designing comprehensive principal evaluation systems: A tool to assist in the development of principal evaluation systems*. Available <https://gtlcenter.org/sites/default/files/PracticalGuidePrincipalEval.pdf>. Accessed 9 November 2020.
- Clifford, Matthew, and Steven Ross. 2012. *Rethinking principal evaluation: A new paradigm informed by research and practice*. Available www.naesp.org/sites/default/files/PrincipalEvaluationReport.pdf. Accessed 9 November 2020.
- Coelli, Michael, and David A. Green. 2012. Leadership effects: School principals and student outcomes. *Economics of Education Review* 31(1): 92–109.

Davis, Steven, Karen Kearney, Nancy Sanders, C. Thomas, and R. Leon. 2011. *The policies and practices of principal evaluation: A review of the literature*. Available https://www.wested.org/online_pubs/resource1104.pdf. Accessed 9 November 2020.

Dhuey, Elizabeth, and Justin D. Smith. 2014. How important are school principals in the production of student achievement? *Canadian Journal of Economics* 47(2): 634–663.

Doherty, Kathryn M., and Sandi Jacobs. 2013. *State of the states 2013 connect the dots: Using evaluations of teacher effectiveness to inform policy and practice*. Washington, DC: National Council on Teacher Quality.

Goldring, Ellen B., Xiu Chen Cravens, Joseph Murphy, Andrew C. Porter, Stephen N. Elliott, and Becca Carson. 2009. The evaluation of principals: What and how do states and urban districts assess leadership? *Elementary School Journal* 110(1): 19–39.

Goldring, Ellen B., Madeline Mavrogordato, and Katherine Taylor Haynes. 2015. Multisource principal evaluation data: Principals' orientations and reactions to teacher feedback regarding their leadership effectiveness. *Educational Administration Quarterly* 51(4): 572–599.

Grissom, Jason A., Richard S. L. Blissett, and Hajime Mitani. 2018. Evaluating school principals: Supervisor ratings of principal practice and principal job performance. *Educational Evaluation and Policy Analysis* 40(3): 446–472.

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. 2015. Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis* 37(1): 3–28.

Hallinger, Philip, and Ronald H. Heck. 1998. Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement* 9(2): 157–191.

Hamilton, Laura S., John Engberg, Elizabeth D. Steiner, Catherine Awsumb Neslon, and Kun Yuan. 2012. *Improving school leadership through support, evaluation, and incentives: The Pittsburgh Principal Incentive Program*. Santa Monica, CA: RAND Corporation.

Henry, Gary T., and J. Edward Guthrie. 2015. *An evaluation of the North Carolina educator evaluation system and the student achievement growth standard 2010-11 through 2013-14*. Available <https://cerenc.org/wp-content/uploads/2015/09/0-FINAL-Evaluation-of-NC-Teacher-Evaluation-9-3-15.pdf>. Accessed 19 November 2020.

Jacques, C., M. Clifford, and K. Hornung. 2012. *State policies on principal evaluation: Trends in a changing landscape*. Available <https://gtlcenter.org/sites/default/files/docs/StatePoliciesOnPrincipalEval.pdf>. Accessed 9 November 2020.

Kimball, Steven M., Jessica Arrigoni, Matthew Clifford, Maureen Yoder, and Anthony Milanowski. 2015. *District leadership for effective principal evaluation and support*. Washington, DC: Teacher Incentive Fund.

Kimball, Steven M., Herbert G. Heneman III, and Anthony Milanowski. 2007. Performance evaluation and compensation for public school principals: Results from a national survey. *ERS Spectrum* 25(4): 11–21.

Kimball, Steven M., and Anthony Milanowski. 2009. Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly* 45(1): 34–70.

Kimball, Steven M., Anthony Milanowski, and Sarah A. McKinney. 2009. Assessing the promise of standards-based performance evaluation for principals: Results from a randomized trial. *Leadership and Policy in Schools* 8(3): 233–263.

Kimball, Steven M., Brad White, Anthony T. Milanowski, and Geoffrey Borman. 2004. Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education* 79(4): 54–78.

Landis, J. Richard, and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1): 159–174.

Locke, Edwin A., and Gary P. Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57(9): 705–717.

Marsh, Julie A., Matthew G. Springer, Daniel F. McCaffrey, Kun Yuan, Scott Epstein, Julia Koppich, Nidhi Kalra, Catherine DiMartino, and Art Peng. 2011. *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report*. Available <https://www.rand.org/pubs/monographs/MG114.html>. Accessed 9 November 2020.

Matarazzo, Melissa F. 2014. Exploring accountability through performance evaluation: How do school and district leaders in three us school districts experience results-based evaluations? Doctoral thesis, Harvard University, Cambridge, MA.

Milanowski, Anthony. 2004. The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education* 79(4): 33–53.

Milanowski, Anthony, Steven Kimball, and Brad White. 2004. The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites. University of Wisconsin CPRE-UW Working Paper No. TC-04-01.

National Policy Board for Educational Administration (NPBEA). 2015. *Professional standards for educational leaders*. Available https://www.npbea.org/wp-content/uploads/2017/06/Professional-Standards-for-Educational-Leaders_2015.pdf. Accessed 9 November 2020.

Reeves, Douglas. 2005. *Assessing educational leaders: Evaluating performance for improved individual and organizational results*. 1st edition. Thousand Oaks, CA: Corwin Press.

Robinson, Viviane M. J., Claire A. Lloyd, and Kenneth J. Rowe. 2008. The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly* 44(5): 635–674.

Rogers, Laura K., Ellen Goldring, Mollie Rubin, and Jason A. Grissom. 2019. Principal supervisors and the challenge of principal support and development. In *The Wiley handbook of educational supervision*, edited by Sally J. Zepeda and Judith A. Ponticell, pp. 433–457. Hoboken, NJ: John Wiley & Sons.

Sanders, Kellie. 2008. The purpose and practices of leadership assessment as perceived by select public middle and elementary school principals in the Midwest. Doctoral thesis, Aurora University, Aurora, IL.

Steinberg, Matthew P., and Morgaen L. Donaldson. 2016. The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy* 11(3): 340–359.

Thomas, David W., Edward A. Holdaway, and Kenneth L. Ward. 2000. Policies and practices involved in the evaluation of school principals. *Journal of Personnel Evaluation in Education* 14(3): 215–240.

U.S. Department of Education (USDOE). 2009. *Race to the Top Program: Executive summary*. Available <https://www2.ed.gov/programs/racetothetop/executive-summary.pdf>. Accessed 9 November 2020.

U.S. Department of Education (USDOE). 2011. *ESEA flexibility: Frequently asked questions*. Washington, DC: U.S. Department of Education.

Waters, Tim, Robert J. Marzano, and Brian McNulty. 2003. What 30 years of research tells us about the effect of leadership on student achievement: A working paper. Eugene, OR: Mid-Continent Regional Educational Lab.

White, David R., Steven M. Crooks, and Jerry K. Melton. 2002. Design dynamics of a leadership assessment academy: Principal self-assessment using research and technology. *Journal of Personnel Evaluation in Education* 16(1): 45–61.

White, Melissa Eiler, Reino Makkonen, Scott Vince, and Jerry Bailey. 2012. *How California's local education agencies evaluate teachers and principals*. REL Technical Brief No. 023. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.