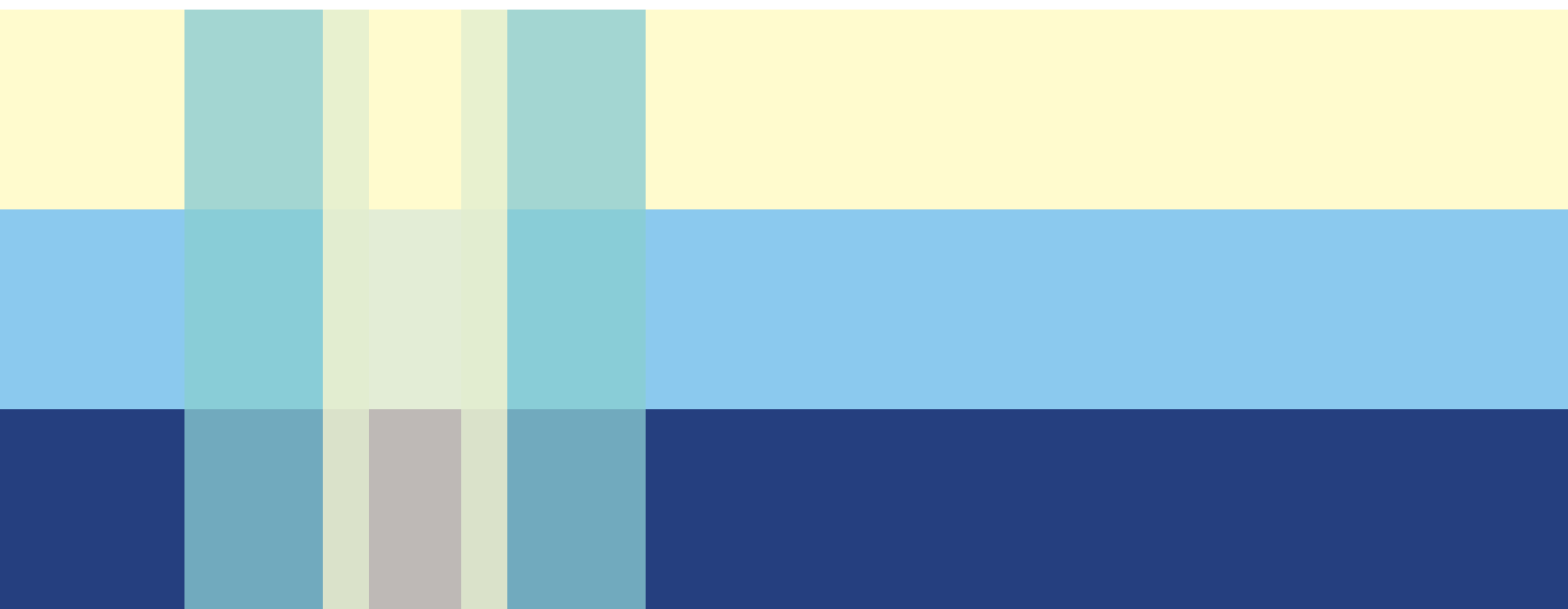


Evaluating and Supporting Principals

POLICY STUDIES ASSOCIATES, INC.

Leslie M. Anderson | Brenda J. Turnbull



Commissioned by:



Evaluating and Supporting Principals

POLICY STUDIES ASSOCIATES, INC.

Leslie M. Anderson | Brenda J. Turnbull



RAND EDUCATION

Commissioned by:



The Wallace Foundation®

PR-2105
January 2016

Preface

This is the fourth report from an evaluation of the Principal Pipeline Initiative. The Wallace Foundation is sponsor of the initiative and of the evaluation, which is conducted by Policy Studies Associates, Inc., and RAND. This report addresses the implementation of principal evaluation and related support as of 2015, viewing implementation in the context of districts' aims, constraints, and capacity. Principal evaluation is part of the Wallace Principal Pipeline Initiative design, intended both as a means of assessing novice principals' performance against clear standards and as a roadmap for tailoring professional development and one-on-one support to identified areas of need. Although educator evaluation has been a prominent topic on the policy scene for many years, there is still much to learn about how educator evaluation systems are designed and implemented. This report provides insights into the strategies and capacities that districts developed to evaluate and support their school leaders.

Contents

	Page
SUMMARY	i
1. INTRODUCTION	1
Background: Principal Evaluation and Support.....	2
This Report and the Overall Initiative Evaluation.....	4
Future Evaluation Reporting.....	7
2. SYSTEM DESIGN: WHY AND HOW THE DISTRICTS RESHAPED PRINCIPAL EVALUATION	9
Background and Purpose	9
Evaluation System Structure	12
Evaluation Procedures.....	24
Principal Perspectives on System Design	34
Summary.....	37
3. PRINCIPAL SUPPORT	39
Novice Principals' Experience with Support	39
Redefining the Supervisor Position to Focus on Supporting Principals	39
Mentor and Coach Support	44
Professional Development.....	47
Summary.....	50
4. SUMMARY AND CONCLUSIONS	51
REFERENCES.....	53
APPENDIX A: SURVEY ADMINISTRATION, WEIGHTING, AND ANALYSES	A-1

Exhibits

PAGE

Exhibit 1: District and partner interview respondents, by district, 2012-15	5
Exhibit 2: Principal survey respondents and response rates, by district	6
Exhibit 3: Summary of initial status and actions taken in principal evaluation, by district, 2011-15.....	12
Exhibit 4: Basic structure of principal evaluation systems in 2014-15, by district	13
Exhibit 5: Detailed structure of principal evaluation systems in 2014-15, by district	14
Exhibit 6: Number of standards and elements of professional practice included in principal evaluation rubrics, by district.....	15
Exhibit 7: Principals who generally agreed that their 2013-14 performance evaluation captured the breadth and complexity of their leadership role and adequately/accurately reflected their performance	16
Exhibit 8: Excerpt on organizational leadership from Denver Public Schools school leadership framework and principal evidence guide	19
Exhibit 9: Student growth measures used to evaluate principal performance, by district	21
Exhibit 10: Terminology and benchmarks used for principal summary ratings, by district	23
Exhibit 11: Percent of principals reporting that they received feedback on the results of their evaluation in the previous year, by cohort	28
Exhibit 12: Principal ratings of the quality of the feedback they received on the results of their evaluation in the previous year, by cohort	29
Exhibit 13: Principals told they needed to improve in at least one area of professional practice based on evaluation results for 2013-14, by district	30
Exhibit 14: Areas of professional practice that principals were told they needed to improve based on evaluation results in 2013-14	31
Exhibit 15: Principals evaluated in 2013-14 who agreed that their district’s evaluation system was accurate and fair, provided clear expectations for performance, and was useful and worthwhile.....	35
Exhibit 16: Difference in principal perceptions of the support they received from their supervisor/evaluator and mentor/coach in 2013-14 versus 2014-15	43
Exhibit 17: Principals who received support from their supervisor/evaluator, mentor/coach, or through professional development who STRONGLY AGREE that the support helped them in various ways	48
Exhibit A-1: Principal survey weights by district and year	A-1



Summary

Background on the Principal Pipeline Initiative and This Study

Since 2011, six large urban and suburban districts have been changing how they evaluate and support novice principals—those in their first three years on the job—as part of a multi-pronged strategy to make these principals more effective. Under the terms of their participation in the Principal Pipeline Initiative sponsored by The Wallace Foundation, these districts have agreed to implement four key components designed to develop a cadre of highly capable novice principals:

- Adopting standards of practice and performance for school leaders that would govern principal preparation, hiring, evaluation, and support
- Improving the quality of preservice preparation for principals
- Using selective hiring and placement practices to match principal candidates with schools
- Implementing on-the-job evaluations and support for novice principals

The initiative also brought the expectation that leadership policies and practices would build the district's capacity to advance its educational priorities. All components would achieve district-wide scale and long-term sustainability.

This report analyzes the progress of these districts in implementing the fourth key component, evaluation and support systems aligned with the district-adopted standards for leaders. Consistent with the initiative's philosophy that evaluations can be a positive source of guidance for improving practice, districts have agreed to provide novice principals with support tailored to their needs, as identified by evaluations. The ultimate goal of this support—which includes support from supervisors, coaching or mentoring, and professional development—is to strengthen principals' capacity to improve teaching and learning.

The six participating districts receive grants from The Wallace Foundation ranging from \$8.5 million to \$13.25 million over five years, along with foundation-funded technical assistance. Districts use these resources to expand on and improve their existing practices and to build sustainable policies for school leadership.

Principal Pipeline Initiative Districts

- Charlotte-Mecklenburg Schools, North Carolina
- Denver Public Schools, Colorado
- Gwinnett County Public Schools, Georgia
- Hillsborough County Public Schools, Florida
- New York City Department of Education, New York
- Prince George's County Public Schools, Maryland

This is the fourth in a series of reports from an evaluation of the Principal Pipeline Initiative by Policy Studies Associates and RAND. (Other components of the initiative are discussed in other reports.) Data for this report were collected from late 2011 through spring 2015 by conducting interviews with district administrators, focus groups of novice principals, surveys of novice principals, and reviews of district documents. More details about methods can be found in the body of the report.

Summarized below are the main findings about principal evaluation and support across the six districts.

Designing Principal Evaluation Systems

In revising their principal evaluation systems, districts put the main emphasis on using evaluation as a tool to help principals improve.

All six districts had a history of evaluating principals' performance, but after joining the Principal Pipeline Initiative, they made many changes in their evaluation systems—and are still refining them. All of these districts have developed rigorous systems with multiple measures to evaluate principal performance.

District leaders believed their redesigned principal evaluation systems would produce sound data and feedback that could be used to help novice principals succeed, even when they arrived with good preparation and demonstrated skills. In no district did leaders express a desire to use evaluations to dismiss large numbers of novice principals.

In all six districts, the revised principal evaluation systems emphasized measures of student achievement growth and measures of professional practice, as required by their states. Within this framework, districts made different choices about what and how to measure, although they had more latitude to set measures of practice than of growth.

States mandated the basic framework of using student growth and professional practice as the main measures in principal evaluations. Five of the six states with Principal Pipeline Initiative districts also assigned particular weights that had to be used to calculate an overall principal rating.

To determine student growth, all six districts used results on the state test as the main measure, consistent with state requirements. Four districts supplemented state test scores with data from other assessments and student performance indicators. In the five districts with state-specified weighting formulas, student growth counted for at least 40 percent of the overall rating, and as much as 70 percent.

To evaluate professional practice, the six districts developed standards and competencies for principal leadership, which they used as a basis for their principal evaluation systems. Every district designed or adapted an evaluation rubric for supervisors to use in rating principals on the district's leadership standards and competencies. The areas commonly addressed by these rubrics included instructional leadership, human resource management, school climate and culture, planning, and school management and operations.

Most principals agreed that the standards used to evaluate their performance were comprehensive and appropriate. Principal supervisors and other district staff continue to tweak the rubrics as they gain more experience in applying them.

Across the six districts, 75 percent or more of novice principals who were evaluated in their first or second year as a principal in 2013-14 agreed that the evaluation system captured the breadth and complexity of their leadership role and adequately/accurately reflected their performance.^a

Recognizing the need for further improvement, district staff members were in the process of refining their evaluation rubrics, as well as the guidance provided to evaluators to help them gather evidence. For example, some supervisors have suggested changes in the wording of some leadership standards or rubrics or have noted competencies that should be added or better defined.

The steps used to evaluate principals' professional practice were designed to encourage two-way discussions between principals and their supervisors.

In contrast to systems of teacher evaluation, which often rely on scheduled, discrete formal observations, principal evaluation in these districts became a year-long conversation between principal and supervisor, intended to produce steady improvement in principal practice.

In all six districts, supervisors spent considerable time in schools. During these school visits, they talked with principals, teachers, and other school staff; observed principals' performance, including their interactions with school staff; and reviewed school improvement plans, teacher professional development plans, and other documents. Determining the type and amount of evidence required to substantiate a particular principal rating was still a work in progress in most districts, however.

A majority of novice principals received directions for improvement based on their evaluations. Instructional leadership was the most common area cited for improvement.

More than half of novice principals were told after their 2013-14 evaluations that they needed to improve in at least one area of practice, according to surveys administered in 2015. About half of the novice principals who received directions for improvement said they were told to improve their instructional leadership.

Providing Principals with Support

A large majority of principals who were told they needed to improve received support in the area flagged for improvement.

Specifically, 86 percent of principals who were advised to improve their instructional leadership skills reported receiving support in this area. The percentages of principals who reported receiving support were similar for human resource management (82 percent) and other management skills (86 percent).

^a The survey asked about the previous year's evaluation to capture the full cycle of feedback and support offered in the following year. The survey results across districts were weighted to reflect each district equally.

Districts are redefining the job of the principal supervisor by shifting focus from overseeing compliance and operations to helping principals succeed as instructional leaders. Districts recognize they need to build supervisors' capacity to deliver this kind of support.

With seed money from The Wallace Foundation, the Principal Pipeline Initiative districts had added more supervisors so that each could work with fewer principals. Districts were also reorienting supervisor's responsibilities to place much greater emphasis on effectively supporting principals' efforts to improve.

At the time of this study, districts were still working out ways to strengthen supervisors' capacity to fill this redefined role. Districts were provided training to help supervisors deliver effective feedback, structure their school visits, and provide coaching and support to principals. In some districts, supervisors were meeting or visiting schools together to discuss and bring more consistency to their coaching and feedback.

Large majorities of principals agreed that their supervisors provided knowledgeable, tailored support. But principals gave slightly more favorable ratings to their mentors or coaches.

Over the period studied, novice principals' perceptions of the support they received from supervisors have been generally positive and have risen. For example, in 2014-15, 80 percent of the principals surveyed in Pipeline Initiative districts reported that their supervisor helped them set effective goals and develop an action plan to meet those goals (up from 75 percent in 2013-14). Focus groups with principals revealed that some principal-supervisor relationships were not living up to district expectations. Comments suggested that some supervisors lacked the capacity to help principals, did not visit schools often enough, or did not instill sufficient trust for principals to candidly discuss their weaknesses.

In all six districts, novice principals also received individualized, on-call support from mentors or coaches. The specific types of assistance varied, but often these coaches or mentors served as trouble-shooters, helping principals address day-to-day problems, or as a bridge between principals and supervisors. In the survey, principals' ratings of support provided by mentors/coaches were slightly higher than those for supervisors.

Principals expressed limited satisfaction with the professional development they received from the district.

All of the Principal Pipeline Initiative districts offered professional development for principals, ranging from traditional courses to professional learning communities in which principals met regularly to discuss problems and receive support from district staff. Although districts sought to tailor their professional development to address needs identified by principal evaluations, principals participating in focus groups often described professional development as compliance-oriented or overloaded with information.

Principals surveyed in spring 2015 did not rate professional development as a significant source of support. Only 23 percent strongly agreed that professional development addressed the pressing issues in their school, for example, and just 32 percent strongly agreed that it led them to make changes in their work. These were well below the comparable percentages for support from supervisors or mentors or coaches.

1. Introduction

The Wallace Foundation is supporting six urban school districts in efforts to make novice principals more effective. The districts have participated since 2011 in the Principal Pipeline Initiative, which embodies a multi-part strategy for strengthening school leadership through district policy and practice:

1. Leader standards to which sites align job descriptions, preparation, selection, evaluation, and support.
2. Preservice preparation that includes selective admissions to high-quality programs.
3. Selective hiring, and placement based on a match between the candidate and the school.
4. On-the-job evaluation and support addressing the capacity to improve teaching and learning, with support focused on needs identified by evaluation.

The initiative also brought the expectation that district policies and practices related to school leaders would work together to build the district's capacity to advance its educational priorities. All components would closely reflect the district's strategic priorities and would achieve district-wide scale and long-term sustainability.

The foundation made a multi-year commitment to the six sites. The districts receive grants of \$8.5 million to \$13.25 million over five years, along with technical assistance supported by the foundation, to build their principal pipelines and the capacity of principal supervisors. They use the funding and assistance to expand and improve on their existing practices and to build sustainable policies and investments in school leadership. The districts are:

- Charlotte-Mecklenburg Schools, North Carolina
- Denver Public Schools, Colorado
- Gwinnett County Public Schools, Georgia
- Hillsborough County Public Schools, Florida
- New York City Department of Education, New York
- Prince George's County Public Schools, Maryland

Background: Principal Evaluation and Support

This report focuses on one of the four key components of the Principal Pipeline Initiative, systems of evaluation and support for novice principals. Like the other components, evaluation and support has been a focus for each of the districts in each year of the initiative. Educator evaluation has been a prominent topic on the policy scene during this time, gaining popularity as a lever for reforming urban districts and becoming a condition for the receipt of substantial federal funding to states and to many districts within them, as we briefly discuss here.

Principal Evaluation: The Policy and Research Background

The Principal Pipeline Initiative has coincided with a time of intense attention to educator evaluation in policy and politics.

EVALUATION AS A KEY INGREDIENT IN REFORMS THAT ADDRESS HUMAN CAPITAL. *The Widget Effect*, a report of The New Teacher Project (later renamed TNTP), argued that prevailing practices in teacher evaluation did a disservice to students by failing to identify those teachers whose performance was inadequate (Weisberg et al., 2009). The report gained attention with its claims that too many teachers were receiving high ratings in a perfunctory process. Teacher evaluation with consequences became a cornerstone of reforms in Washington DC, Chicago, and many other cities, leading to opposition from teacher unions along with uncertainty and confusion among teachers (Sartain, Stoelinga, & Brown, 2011). Subsequent research has explored issues around teacher evaluation in depth. It has assessed the validity of various evaluation methods (see, e.g., Kane et al., 2013). Researchers have also raised cautions about hasty implementation or a punitive stance (National Research Council, 2009) and have explored the intended and unintended consequences of using value-added methods to evaluate individual teachers' performance (Harris & Herrington, 2015).

Principal evaluation has drawn less attention. Nationally, we know from principals' responses to the Schools and Staffing Survey in 2011-12 that 91 percent of all public school principals had a performance evaluation during that year, that evaluation was an annual event for 72 percent, and that student test scores were part of the evaluation for 52 percent (data retrieved from National Center for Education Statistics). Some earlier research had pointed to the shortcomings of principal evaluation systems that offered little or no guidance for practice improvement (Goldring et al., 2008; Portin, Feldman, & Knapp, 2006).

Advocates and researchers have developed recommendations for effective principal evaluation in this decade. They include considering the principal's context, incorporating standards of leadership practice, using evaluation to build capacity, and focusing on multiple measures of performance (Clifford & Ross, 2012). A similar list of recommendations, based on a review of the admittedly limited research on principal evaluation, also spoke to the desirability of aligning evaluation with outcomes (e.g., student achievement and effective instruction), reliable and valid evaluation tools and procedures, a focus on a manageable number of high-impact evaluation criteria, and engagement of principals in establishing goals and assessing their own performance (Davis et al., 2011).

STANDARDS-BASED ACCOUNTABILITY AS A POLICY OF THE OBAMA ADMINISTRATION. With funding available to disburse at the height of the Great Recession, the U.S. Department of Education was in a strong position to direct attention to an education policy agenda. In order to compete for large Race to the Top grants from the U.S. Department of Education, states had to enact policies of several kinds, among them the evaluation of teachers and school leaders using criteria that addressed both student growth and professional practice.¹ Five of the states housing Principal Pipeline Initiative districts (Florida, Georgia, Maryland, New York, and North Carolina) won large Race to the Top grants in 2010, and Colorado won a smaller grant in the following year. Thus all the districts were subject to this program's requirements. In implementing Race to the Top, 9 of the 12 states that had won grants in 2010 found that principal evaluation brought fewer concerns than teacher evaluation for several reasons, such as: principal evaluation affected fewer individuals; principals were often involved in helping design the evaluation system; and principals were accustomed to accountability for student performance (Government Accountability Office, 2013).

Principal Evaluation and Support in the Principal Pipeline Initiative

The Principal Pipeline Initiative aimed to develop intentionally aligned systems of preparation, hiring, evaluation, and support for novice principals, governed by standards that specified the capabilities and performance that the district wanted to see in principals. Districts would systematically evaluate novice principals' instructional leadership capabilities, give them feedback, and, on the basis of the evaluation results, provide professional development and skilled mentoring to support them in further developing their needed capabilities. The initiative defined novices as those in their first, second, and third years in the position.

An important dimension of the initiative's design was active partnership between grantees and The Wallace Foundation. This included not only each district's ongoing dialogue with senior foundation officials about detailed work plans and budgets, but also extensive opportunities to receive technical assistance from providers vetted by the foundation and to participate in the initiative's Professional Learning Communities. The technical assistance included skilled help for each district in setting up its new data system, called a Leader Tracking System, as a repository for individual-level, longitudinal information about aspiring and novice leaders. The system would include data on the individual's preparation and certification history, positions held and the schools in which he or she had held them, assessed skills and performance in each position, and participation in professional development and coaching.

The initiative design provided for the ongoing improvement of pipeline components through data. For example, by using the Leader Tracking System to review aggregate data on the performance of the graduates of each preservice program, districts could identify the programs with which they wanted to continue close partnerships and those that needed improvement.

¹ Other policy actions that were prerequisites for Race to the Top eligibility were development or adoption of standards for students, data systems, and a structured process for turning around low-performing schools.

This Report and the Overall Initiative Evaluation

This report addresses the implementation of principal evaluation and related support as of 2015, viewing implementation in the context of districts' aims, constraints, and capacity. Previous reports in this series described the districts' plans and first-year activities (Turnbull, Riley, Arcaira, Anderson, & MacFarlane, 2013), analyzed the preparation and support for school leaders offered by districts and their partners (Turnbull, Riley, & MacFarlane, 2013), and analyzed implementation of all components as of 2014 (Turnbull, Riley, & MacFarlane, 2015). These reports and a future final report on implementation are part of a multi-year evaluation of the Principal Pipeline Initiative that has a dual purpose: to analyze the processes of implementing the initiative's components in the participating districts from 2011 to 2015; and then to assess the results achieved in schools led by principals whose experiences in standards-based preparation, hiring, evaluation, and support have been consistent with the initiative's design.

The evaluation team is issuing reports while districts are introducing changes in their policies and practices under the initiative, with the result that the findings in this and other implementation reports are preliminary. District data on individual principals, such as the numbers of principals hired under particular procedures, are not yet complete but will become available for evaluators' use through the Leader Tracking Systems. A separate study is beginning to analyze costs.

A key purpose for the implementation reports is to inform other districts about the activities, accomplishments, and challenges of the six participating districts. Data collection for this report took place during the four years after the initiative's August 2011 launch, while implementation remained a work in progress. In each year, district leaders described policies and practices that they saw as consistent with the initiative's intent, but in general not enough time had passed for them to offer more than preliminary impressions of the results of those policies and practices. Most district leaders spoke in interviews about key policies and practices they still intended to change in order to meet new or continuing challenges.

Data Sources

This report is based on an analysis of data collected by the evaluation team from the following sources: (1) semi-structured interviews with administrators in district central offices and in outside programs that partnered with the districts (e.g., in universities), conducted during site visits in spring 2012, 2013, 2014, and 2015 (Exhibit 1); (2) focus groups with novice principals (i.e., those in their first, second, and third years in their position) in spring 2013 (N=36), 2014 (N=32), and 2015 (N=31); (3) surveys of novice principals in spring 2013, 2014, and 2015 (Exhibit 2); (4) documents including the districts' evaluation-related materials (guides, manuals, and rubrics) and their proposals, work plans, and progress reports for the foundation; and (5) observation of and participation in cross-site meetings from 2011 through 2015, including observation of presentations and panel discussions by district leaders.

Exhibit 1: District and partner interview respondents, by district, 2012-15

District	2012	2013	2014	2015
Charlotte-Mecklenburg	14	22	11	13
Denver	19	21	13	22
Gwinnett County	14	15	16	19
Hillsborough County	17	18	30	18
New York City	8	14	18	15
Prince George's County	19	23	16	16
Total	91	113	104	103

Exhibit reads: In Charlotte-Mecklenburg, 14 district or partner staff members were interviewed in 2012.

Site-visit interviews were arranged by the project director in each district, responding to specifications from the evaluation team. In each year we requested interviews with the project director, the superintendent (or, in New York City, another high-level official in the central office); other members of the executive team such as the directors of human capital, curriculum and instruction, and data systems; and central-office staff and partner-program leaders who were, collectively, knowledgeable about standards, job descriptions, preparation programs, hiring and placement, supervision, evaluation, and support for principals. Where two or three people worked closely together on a particular function, the team typically conducted a joint interview with them. In some cases project directors arranged one or more larger group interviews (e.g., the eight principal supervisors in Hillsborough County were interviewed in two groups in 2014, as were the eight principal coaches in that district).

The semi-structured interview protocols included not only factual questions but also probes for the respondents' perceptions of what district practices appeared to be serving the intended purposes, what practices appeared to need improvement, and what changes (if any) were under consideration. During the visits that preceded one of our special-focus reports (i.e., the 2015 visit before this report on evaluation and the 2013 visit before a report on principal preparation), additional time was spent in gathering more information about that topic. The comments made by district staff in the initiative's twice-yearly gatherings also informed the evaluation team's understanding.

Focus groups with novice principals were conducted in each district in 2013, 2014, and 2015. Project directors identified the participants. As with the interview protocols, there was some tailoring of the focus-group protocols to ensure attention to the topic of the next upcoming report. The focus groups' comments provided useful insights that informed the team's understanding of the more systematically collected data, but because the participants were not necessarily representative of their peers, caveats accompany the focus-group comments referenced in this report.

More complete data were gathered from novice principals in web-based surveys addressing their perceptions and experiences. The surveys were administered in spring 2013, spring 2014, and spring 2015 to all first-, second-, and third-year principals in the six districts. In the 2013 wave of surveys, 353 principals responded out of the total of 617 novice principals surveyed across the six districts, for an overall response rate of 57 percent (Exhibit 2). A total of 541 of the 821 novice principals responded in 2014, for an overall response rate of 66 percent, and 514 of the 791 novice principals responded in 2015, for an overall response rate of 65 percent.

Exhibit 2: Principal survey respondents and response rates, by district, 2013-15

District	2013		2014		2015	
	Number of respondents	Response rate, in percent	Number of respondents	Response rate, in percent	Number of respondents	Response rate, in percent
Charlotte-Mecklenburg	43	96%	52	96%	56	93%
Denver	56	78	68	83	55	83
Gwinnett County	39	93	47	100	41	98
Hillsborough County	55	90	42	89	46	81
New York City	102	31	269	51	248	50
Prince George's County	58	89	63	93	68	96
Total	353	57	541	66	514	65

Exhibit reads: In 2013, 43 principals completed the survey in Charlotte-Mecklenburg, comprising 96 percent of all first-, second-, and third-year principals in the district.

Because we wanted to place principal evaluation in the context of a full cycle of evaluation (which typically ended in late spring or summer) and also to report on the feedback and follow-up support that followed an evaluation cycle, each year's surveys asked principals about the evaluation conducted during the *previous* year and about the feedback and support that had followed. Thus, for example, survey responses gathered in spring 2015 reflect principals' reported experience with their evaluations that were completed in 2014 and the feedback and support that followed during the 2014-15 school year. They do not reflect perceptions of the evaluation cycles that were in progress as of spring 2015. The 2015 survey gathered more detailed responses on evaluation and support than earlier surveys, but all rounds of surveys addressed these topics with a core set of questions. Findings are compared across years in this report, for those questions that were repeated across years.

Data Analysis

For this study of initiative implementation across six districts, the district is the primary unit of analysis. Therefore, where this report presents overall survey findings from all districts, the analysis gives equal weight to each district. The raw survey responses, if compiled across districts, would over-represent New York City—which had the largest numbers of novice principals—and under-represent the other districts. Because of the significant cross-district differences in numbers of novice principals, we applied post-stratification weights so that each district would be equally represented in overall analyses. Accordingly, overall cross-district findings are presented as weighted data. For a complete explanation of the survey weights, see Appendix A.

In the analysis of survey data, due to the relatively low response rate from principals in New York City (50 percent in 2015), we tested for non-response bias to determine whether respondents were representative of all principals in the district. We used New York City’s progress reports submitted to The Wallace Foundation to determine whether there were significant differences between the percentages of survey respondents who reported having graduated from particular preparation programs versus the percentages of graduates of those programs reported in the more complete data from the district. This was the only data point in our survey for which frequencies could be obtained from a different and more complete source for comparison. The comparison showed that a few programs were overrepresented among survey respondents (e.g., three programs were overrepresented in 2015). Therefore additional analyses tested a sample of survey items for differences in responses from graduates of each of those programs versus graduates of all other programs. We found only a few significant differences among group means, and these few significant differences did not follow a clear pattern or cluster within a particular set of items. We therefore determined that within-district survey weights were not necessary for New York City.

Qualitative analysis was iterative. Interview transcripts and notes were coded by the evaluation team, using software for qualitative data analysis, according to key features and expectations of the Principal Pipeline Initiative. Multiple iterations of analysis identified and refined the specific themes, descriptions, and analyses presented in this report. Drafts were reviewed by the particular team members who visited the site for factual accuracy and revised as necessary. The districts’ progress reports to The Wallace Foundation were a supplementary source for detailed factual descriptions of policies enacted; similarly, their initial proposals were a source for facts about policies and practices in place before the initiative. Finally, project directors in the districts conducted a further fact-check prior to publication.

Future Evaluation Reporting

As the initiative and the evaluation continue, final reports on implementation and on effects will assess the actions taken by districts and their partner institutions in this initiative. A 2016 report will assess overall implementation, identifying the structures and policies put in place, the results observed, and factors that have supported or impeded the sites’ progress in carrying out their plans. After the end of the sites’ grant period, in 2018, the final report from the evaluation will analyze the effects of “pipeline principals”—those prepared, selected, and supported in ways consistent with the initiative’s design—on important school outcomes, including growth in student achievement.

2. System Design: Why and How the Districts Reshaped Principal Evaluation

All six districts had a history of evaluating principals' performance, but they made many changes under the initiative. Here we discuss the concerns and priorities that they brought to evaluation systems and the systems that they crafted to serve local purposes while meeting state requirements. Most made gradual yearly changes in evaluation systems. All were continuing to tinker with the structure and content of their evaluation systems, redefining the number, type, and value of the measures used to evaluate principal performance; considering ways to further strengthen the rubrics used to measure professional practice; and deliberating over the consequences of the ratings.

Background and Purpose

Districts embarked on the Principal Pipeline Initiative in 2011 wanting to equip their novice principals to succeed, and they believed that evaluation systems could help serve this purpose.

Leaders in these districts believed that improved evaluation systems could help their novice principals. They agreed with the premise of the initiative: not only should future cohorts of principals arrive after strong preparation and rigorous hiring processes, but on-the-job evaluation should reveal their strengths and remaining weaknesses and point the way for tailored support. Evaluation rubrics² would be aligned with the same standards of principal professional practice that drove preparation and hiring, so that aspiring principals, novices, and their supervisors would receive consistent messages about what the district expected.

From the start of the initiative, district leaders wanted evaluation systems to provide better information that could support the improvement of sitting principals. In no district did leaders show an appetite for weeding out larger numbers of novice principals through evaluation. Three districts (Denver, New York, and Prince George's County), in fact, were concerned about the high rate of voluntary or involuntary principal departures they had seen before the start of the initiative. They wanted evaluation to be informative, as illustrated by the following statements from district leaders in Denver and Prince George's County, respectively, in 2012:

Takeaways

1. These districts saw evaluation as a way to develop principals, not a way to dismiss more of them.
2. Most novice principals reported that their district evaluation systems were appropriate.
3. States specified the big building blocks for principal evaluation—student growth and professional practice—but most of these districts were able to set their own standards for professional practice.
4. There were gaps in the guidance given to principal supervisors on collecting evidence and assigning principal ratings.
5. Evaluation often unfolded as what one principal called “an all-the-time conversation” with a supervisor.

² A rubric is a tool used in evaluation that spells out the criteria on which practice is to be rated and describes different levels of quality in relation to these criteria.

The old evaluation system didn't clearly define what the instructional leadership components looked like. We're taking big steps forward, putting a stake in the ground around those components, with more clarity about overall performance. The old structure didn't have final rating categories. You'd just be renewed or non-renewed.

We have had the same principal evaluation for years and years. It was perfunctory and didn't get at the leadership standards we wanted.

The design of the Principal Pipeline Initiative emphasized standards-based principal evaluation with a focus on instructional leadership. More specific system requirements came from new federal incentives and state policies. The federal Race to the Top competition brought requirements for systems of teacher and leader evaluation that met federal criteria, which included giving "significant" weight to measures of student growth and also incorporating measures of professional practice. Beginning in 2012, states could also receive federal waivers of accountability requirements of the Elementary and Secondary Education Act, with systems of teacher and leader evaluation based on student growth and professional practice among the conditions they had to meet to receive those waivers. Five of the six districts in the Principal Pipeline Initiative were in states that had won Race to the Top grants in 2010, and the districts had pledged to cooperate with implementation. Colorado won a smaller grant in 2011. Race to the Top put these states on a somewhat faster timetable than others toward compliance with the federal requirements for principal evaluation, but by 2015 there were 43 states subject to those requirements because they had received waivers under the Elementary and Secondary Education Act.³

State policy in place by the beginning of the Principal Pipeline Initiative dictated that the state would have to approve the districts' systems of principal evaluation, and that these systems would include measurement of student growth and professional practice. But the implementation process over the following years included partial or pilot implementation while evaluators were trained and the system refined, and it often included negotiation with the state over new or modified standards and measures. Implementation stages were somewhat different in each district. For example, Hillsborough County moved quickly to meet Florida's aggressive timeline, putting a new system of principal evaluation in place for the 2011-12 school year. Later, the district sought and received state approval each time it adjusted its system to reflect revisions in its principal standards and competencies, with the result that a newly revised local system was in place in 2014-15. Prince George's County began small-scale piloting of a new evaluation system to generate formative feedback for principals in 2011-12, while Maryland was moving slowly with its new requirements for principal evaluation. Prince George's County's standards took shape over the following year; the district submitted its evaluation system for state approval in 2013; and it rolled out a state-approved system districtwide in 2013-14.

³ The Every Student Succeeds Act, enacted in December 2015, changed the accountability provisions and ended the waiver authority of the Elementary and Secondary Education Act. By summer 2016, states will no longer be subject to the waiver authority's specifications for principal or teacher evaluation, and it is likely that many state policies and requirements will change.

The balance between state and district initiative in shaping an evaluation system also varied, depending on state policy and capacity. North Carolina moved forward quickly and offered districts little flexibility, and Charlotte-Mecklenburg has followed the North Carolina system from the beginning. That system had seven standards for professional practice in place by 2010, then added student achievement as an eighth standard in 2012-13. Georgia, on the other hand, made slow progress toward compliance with Race to the Top requirements, and the existing Gwinnett County evaluation system eventually furnished a model for the state-mandated system that was rolled out in 2014. New York State developed an evaluation system with several required features but left other features open to negotiation between districts and unions. New York City and its principals' union agreed in 2013 on the system to be used for 2013-14 through 2015-16, and the state allowed use of that system.

Most of the six districts made gradual changes to their evaluation tools and procedures before and during implementation of the Principal Pipeline Initiative. Charlotte-Mecklenburg was a partial exception because the district moved quickly to implement the required state standards, but discussion and training around the use of these standards was ongoing in that district during the initiative. The typical practice among these districts was an initial year of pilot or partial implementation of a system, followed by (1) continuing fine-tuning of the principal standards in consultation with the state and (2) years of working with the principal supervisors who evaluated principals, aiming both to familiarize them with the system and also to use their feedback to improve the rubrics and procedures.

By 2015, district leaders had made changes in their evaluation systems (Exhibit 3). District leaders and principal supervisors saw these systems as part of the overall support structure intended to help principals effectively lead their schools. The evaluation rubric, they believed, could be a navigation tool to help guide principals' professional development and to structure the delivery of targeted support. In Prince George's County, for example, a district administrator described the evaluation system as a way to help school leaders understand the skills and abilities they needed as leaders in order to help a school improve. A supervisor described the way she used the evaluation data: "...my goal is to be able to use that evaluation to have a conversation about areas of strength and weakness. And to push [the principal] as much as I can to leverage the strengths, as well as really being intentional about how to address the weaknesses." A Hillsborough County administrator, similarly, said the evaluation rubric could foster an ongoing district conversation about leadership that would develop increasingly nuanced definitions and interpretations of effective leadership.

Exhibit 3: Summary of initial status and actions taken in principal evaluation, by district, 2011-15

District	Status of principal evaluation system as of 2011-12	Key actions taken through spring 2015 in principal evaluation systems
Charlotte-Mecklenburg	State required the use of specific standards.	Continued to use state standards. Hired more principal supervisors and trained them in using the evaluation system.
Denver	School performance and conditions largely dictated principal retention decisions. Standards for principal practice were not identified, and principals received little feedback.	Developed standards and a new evaluation system specifying measures of student growth and professional practice, as required by the state. Hired more principal supervisors and trained them in using the evaluation system.
Gwinnett County	Used a longstanding system of measures of student growth and principal practice.	Fine-tuned the existing measures. Hired more principal supervisors and trained them in using the evaluation system.
Hillsborough County	Had developed a new evaluation system, as required by the state, and hired more principal supervisors, who were beginning to learn that system.	Developed new principal standards and changed the evaluation system to align with them. Trained the principal supervisors in using the evaluation system and in providing feedback.
New York City	School performance and conditions largely dictated principal retention decisions. Standards for principal practice lacked specificity, and principals received little feedback.	Developed and received approval for a principal practice rubric aligned with the Quality Review rubric used for rating schools. Trained principal supervisors in using the evaluation system.
Prince George's County	School performance and conditions largely dictated principal retention decisions. Standards for principal practice were not identified, and principals received little feedback.	Developed standards and a new evaluation system specifying measures of student growth and professional practice, as required by the state. Hired more principal supervisors and trained them in using the evaluation system.

Evaluation System Structure

Each district's principal evaluation system had two major parts: measures of student growth and measures of professional practice (Exhibit 4). Gwinnett County and Hillsborough County each included an additional component measuring other aspects of principal performance, such as management of human resources and operations.

State mandates called for measures of student growth and measures of professional practice, and these mandates largely determined the weight assigned to each part of the system in calculating an overall rating. Growth in student performance counted for at least 40 percent of the overall rating in the five districts that had formulas, reflecting state requirements. (Charlotte-Mecklenburg did not have a percentage system for compiling ratings, and this was also consistent with state policy, since North Carolina did not require use of percentages.) Measures of student growth counted for 70 percent in Gwinnett County and 40 or 50 percent

in the other four districts that used percentages. All systems also included supervisor ratings, with weights ranging from 60 percent of the total rating in New York City and 50 percent in Prince George’s County to 12 percent in Gwinnett County. Some districts had additional measures using parent surveys, teacher surveys, or data on human capital development or building management. Some used measures of school conditions and practices, such as culture, curricula, and instruction, to rate individual principal performance.

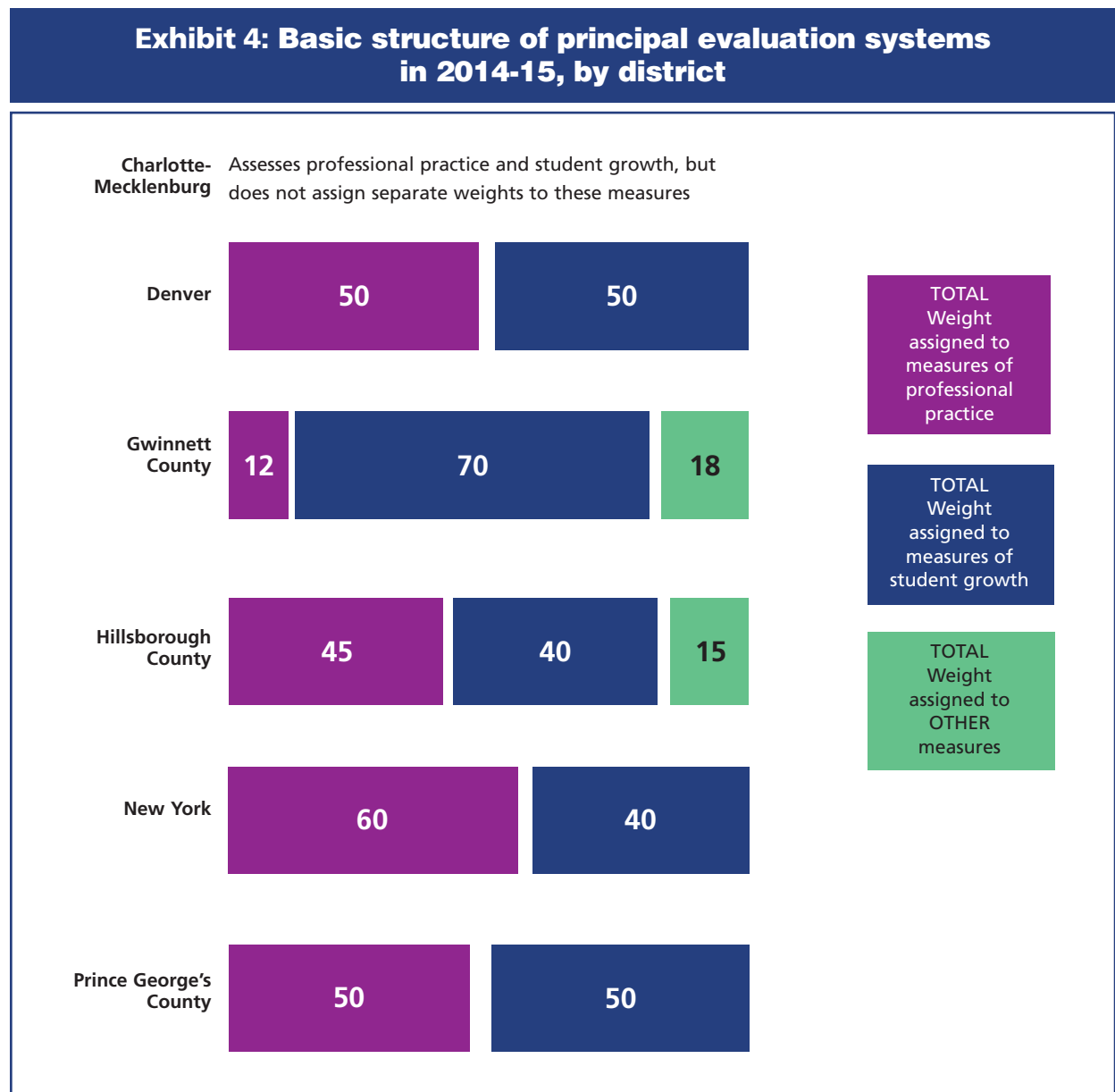


Exhibit reads: Charlotte-Mecklenburg’s principal evaluation system does not assign separate weights to measures of professional practice and student growth. Denver, however, assigns a weight of 50 percent to measures of professional practice and 50 percent to measures of student growth.

Within a framework that included measures of student growth and professional practice, all the districts made different choices. They used different types of data to measure professional practice and student growth, or used similar data in different ways (Exhibit 5). The following sections identify important choices that they made, and the rationales for those choices.

Exhibit 5: Detailed structure of principal evaluation systems in 2014-15, by district

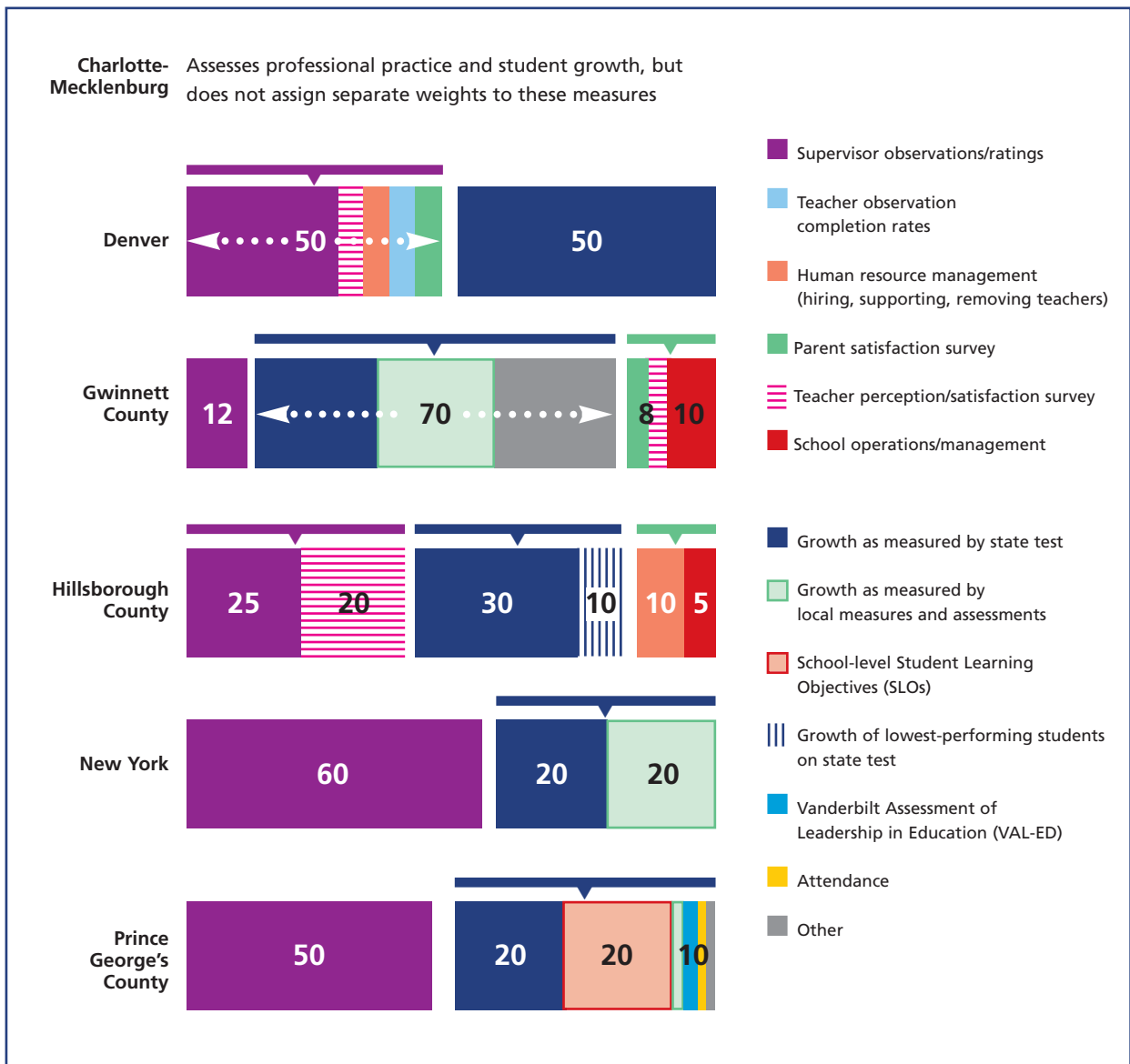


Exhibit reads: The structure of Denver’s evaluation system includes a measure of professional practice, which is weighted at 50 percent of a principal’s final evaluation score. When assigning the final score, principal supervisors can take into account a variety of measures in addition to the supervisor’s school-based observations, including results from the teacher perception/satisfaction survey, the parent satisfaction survey, human resource management data, and/or teacher observation completion rates. The other 50 percent of a principal’s final evaluation score is based on student growth as measured by the state test.

Measures of Professional Practice

Developing standards and competencies for principal leadership was one of the districts' first tasks under the Principal Pipeline Initiative. Their work on standards and competencies began in 2011-12 and continued over the years. The districts incorporated their principal standards and competencies into the rubrics for principal professional practice that they developed to meet state evaluation mandates. Evaluation rubrics were therefore based on the same standards and competencies that the districts used to shape principal preparation and hiring. Most districts revised their standards documents repeatedly for clarity and completeness. In fact the use of standards in principal evaluation systems played an important part in bringing to light issues of clarity or completeness in the standards, which districts then addressed (Turnbull, Riley, & MacFarlane, 2015).

Across the six districts, the number of standards on which principals were evaluated ranged from three to eight, with a total of 10 to 26 elements of practice within those standards (Exhibit 6). Some districts (Denver, Hillsborough County, and Prince George's County) had developed longer lists of standards early in the Principal Pipeline Initiative but streamlined them for the sake of more powerful communication of their top priorities (Turnbull et al., 2015). The five common areas of focus in district standards were instructional leadership, human resource management, school climate and culture, planning, and school management and operations.

Exhibit 6: Number of standards and elements of professional practice included in principal evaluation rubrics, by district

	Number of standards	Number of measured elements of practice
Charlotte-Mecklenburg	7	20
Denver	6	11
Gwinnett County	8	—
Hillsborough County	5	26
New York City	3	10
Prince George's County	8	25

Exhibit reads: Charlotte-Mecklenburg's principal evaluation rubric included seven standards and measured 20 elements of professional practice.

Principal supervisors and principals agreed, for the most part, that the standards used to evaluate principal performance were comprehensive and appropriate. Across the six districts, 75 percent or more of novice principals who were evaluated in their first or second year of the principalship in the 2013-14 school year agreed that the evaluation system captured the breadth and complexity of their leadership role and adequately/accurately reflected their performance (Exhibit 7).⁴ Agreement is defined here as a response of

⁴ The surveys asked about the previous year's evaluation in order to provide a picture of the full evaluation cycle, including feedback and support offered in the following year. Readers should also note that we weight the survey findings across districts so that the results reflect each district equally.

“somewhat,” “to a considerable extent,” or “to a great extent” to the question “to what extent do you agree,” where the two other possible responses were “minimally” and “not at all.” The percentages of principals reporting they agreed at least somewhat did not differ significantly across districts. They also did not change significantly over the years of this evaluation: principals’ positive views held steady, on average, from the evaluation cycle of 2011-12 through the 2013-14 cycle. This positive response stands in notable contrast to prior research suggesting that principal evaluation systems five to ten years earlier often lacked clear performance expectations or standards (Goldring et al., 2008), and failed to focus on the appropriate leadership competencies (Goldring et al., 2008; Seashore Louis et al., 2010).

Exhibit 7: Principals who generally agreed that their 2013-14 performance evaluation captured the breadth and complexity of their leadership role and adequately/accurately reflected their performance

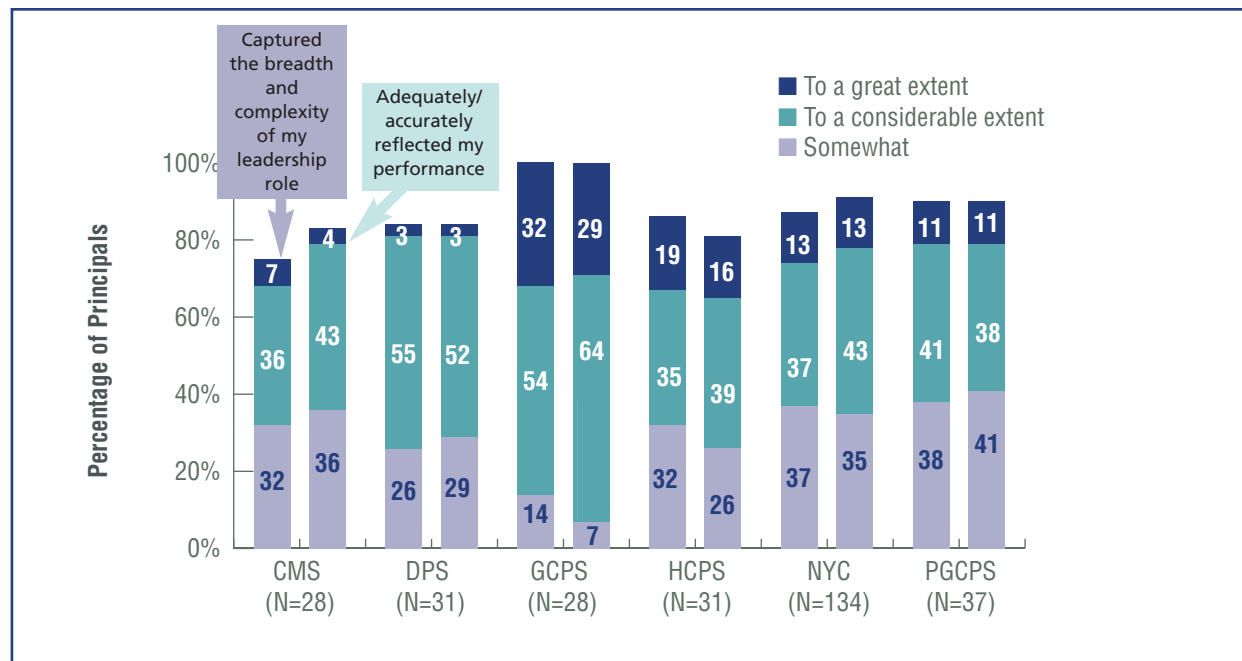


Exhibit reads: Seventy-five percent of principals in Charlotte-Mecklenburg who were evaluated in their first or second year of the principalship in 2013-14 agreed that the principal evaluation system captured the breadth and complexity of their leadership role and 83 percent agreed that it adequately/accurately reflected their performance. The other responses, not shown in the exhibit, were “not at all” and “minimally.”

Central office staff saw the principal evaluation rubrics as sources of needed insight for principals and for principal supervisors. In every district, principal supervisors considered their principal evaluation rubrics to be useful for tracking and measuring leadership skills. A coach in Denver made a typical comment when describing the rubric as exceptionally well organized such that anyone could look at it and know the lead-

ership competencies that the district valued: “So it really paints the picture of what’s really a highly effective instructional leader. ‘What skills do I have to be exhibiting? What should be happening in my school?’” Principal supervisors, too, could do their work better thanks to a rubric that addressed important aspects of the school’s functioning, according to a Hillsborough County administrator:

[The principal evaluation system] allows [supervisors] to truly have a clear understanding about what’s happening within their schools. I think that they will be much more educated and have much more clarity when they go to do the evaluations for their principals. The successes that have occurred at those schools, the things that they might be still working on, still struggling with. I think the supervisor will be much more equipped to be an effective evaluator.

Nevertheless, the rubrics were not perfect—and may never be. As principal supervisors began to use a rubric to evaluate principal performance, some identified gaps and ambiguities in their leadership standards and evaluation rubric. Some asked whether the standards captured all the competencies that define an effective leader. A Denver district leader, for example, said that because principals were expected to develop assistant principals’ leadership capacity, this should be “explicitly called out in the rubric.” District leaders were also trying to provide better wording and descriptors to convey a rubric’s meaning. One in Hillsborough County said: “Now, ... we’re going to do some word changes because as [supervisors] have used [the rubric], they’re saying, ‘Can we change this word? Because it’s a little bit too ambiguous.’”

THE DILEMMA OF DIFFERENTIATION. Use of the new evaluation systems was bringing up fundamental questions about fairness. In particular, there were tensions between the value of accommodating differences among principals and schools and the value of holding uniformly high expectations for students’ learning conditions. Supervisors in several districts believed that their district’s evaluation rubric should differentiate expectations by principal experience or school context, although none of the rubrics in use in 2015 did so. As one supervisor explained, applying a single rubric to all principals seemed unfair to novices:

It is too much. I mean, you almost have to be superhuman to be able to be highly effective, or distinguished, in all of those categories. And how do we make it a little more bite-sized? Even though we know that leaders need all of that, how do we chunk it? Maybe what’s a priority for year one; what’s a priority year two? I don’t know. But just trying to see how we can make it a little more manageable.

Others commented that schools present principals with different levels of challenge based on their needs and contexts:

With principals, how do you keep the same standards and high expectations for everyone recognizing that certain schools are easier to lead than others? For example, a high poverty school [is harder to lead] than an affluent school and a leader who is placed in a high-functioning school does not have to work as hard to have those systems operate well because they already are. Is that the result of what the leader has done or what she or he has inherited?

District leaders wrestled with the issue of differentiation, acknowledging that these concerns were legitimate and reasonable. Still, they chose to hold constant the goal of effective leadership in every school. “I don’t know that we are ever going to fully solve for that. I don’t know that it’s a solvable thing given that we’re not going to ... water the system down for a first-year principal,” as a Denver leader put it. Similarly, a district leader in Hillsborough County, although aware of the problems, argued that the expectations for all principals had to be the same because “the outcomes for students that we’re expecting are the same.”

INCLUDING MEASURES OF SCHOOL PRACTICE. Another debate was emerging about whether rubrics should include measures of school practice in addition to individual principal practice. With a rubric that had been intentionally developed to emphasize measures of school practice—including measures of school culture, curricula, and instruction—New York City administrators were divided in their opinions about this rubric. Some felt that evaluating principals on the basis of school conditions signaled to principals that they had a role in improving teacher and student performance. One administrator explained that including measures of school practice had become a way to “unearth” problems in the school—such as struggling teachers or a negative school culture—that should be a principal’s responsibility to solve. “You do not hear people try to skirt responsibility for the [school] piece [anymore]. They might struggle with it, or they might disagree [with what is or is not quality practice], but [the rubric] has eliminated that layer of, like, ‘I’m not responsible for this.’”

On the other hand, a supervisor in the district explained why she thought this was a flawed strategy because it left out measures of some principal skills: “It is a tool to evaluate the performance of a school, it wasn’t necessarily created to evaluate the abilities and the performance of a principal. ... it doesn’t really get to the dispositions, the ability to communicate, the ability to strategize.”

Denver’s rubric straddled this divide between emphasizing principal behaviors and emphasizing school practices. It defined the principal *and* school behaviors that should be demonstrated for each leadership standard, as illustrated in the evidence guide that the district used in the most recent evaluation cycle (Exhibit 8). For example, principal behaviors related to organizational leadership would include organizing school staffing and scheduling to maximize both instructional time for students and professional learning time for teachers. School-level practices showing organizational leadership would include few interruptions to instructional time and teachers having ample time to collaborate with one another.

Exhibit 8: Excerpt on organizational leadership from Denver Public Schools school leadership framework and principal evidence guide

	<i>Effective</i>	<i>Distinguished</i>
Principal Behaviors	<p>Utilizes innovative staffing and scheduling to ensure the school day and school year maximize instructional time to benefit all students.</p> <p>Applies a schedule that maximizes time for teachers to learn, innovate, and plan together.</p> <p>Balances multiple and competing priorities in a manner that aligns with the values, vision and goals of the school.</p> <p>Provides clear rationale for resource decisions based on the school's mission, strategies, and learning goals.</p> <p>Makes strategic and sound, legal, and budgetary decisions.</p> <p>Focuses on both short and long-term fiscal management decisions that are grounded in the strategic goals of the UIP.</p> <p>Ensures that budget planning and implementation represent a focus on equity for all student populations.</p> <p>Develops external resources that align with the school budget in alignment with the school's UIP.</p> <p>Ensures that the school building is a safe, clean, and aesthetically pleasing school environment.</p> <p>Develops and ensures effective implementation of safety and risk-management plans (e.g., lockdown drills, fire drills, tornado drills).</p>	<p><i>In addition to "Effective:"</i></p> <p>Staffing and scheduling are uniquely designed to ensure that the school day and year fully maximize the use of time to improve student achievement and staff collaboration.</p> <p>Supports others throughout the school community to ensure that everyone organizes and manages time to advance student learning priorities.</p> <p>Collaborates with the school community to creatively maximize funds.</p> <p>Proactively communicates difficult budget decisions, and secures additional resources to achieve goals.</p> <p>Maintains the confidence of stakeholders during times of significant financial stress.</p>
School Behaviors	<p>Teachers have ample time to collaborate with one another.</p> <p>Students receiving specialized instruction and interventions also receive grade-level, core instruction.</p> <p>There are seldom interruptions to instructional time, and teachers and staff are able to focus their planning time and committee work on driving student achievement.</p> <p>Yearly budget decisions are anchored to current needs and student data and put the needs of students first.</p> <p>Decision-making is transparent and all stakeholders understand the reason behind decisions related to the use of resources.</p> <p>The school environment is viewed as safe, clean, aesthetically pleasing, and representative of the school's values, vision, and mission.</p> <p>The school community is well prepared for crisis situations and is practiced in the protocols required to effectively respond to crises.</p>	

Measures of Student Growth

Every district used the state test as a measure of principal effectiveness with respect to student performance growth, but beyond that commonality there were many differences (Exhibit 9). Student growth on the state test was the only measure used in Denver and Hillsborough County, but Hillsborough County looked at growth among the lowest-performing students in addition to overall student growth. The other four districts used additional measures to evaluate principal effects on student growth. Charlotte-Mecklenburg's measure of student growth included data from end-of-course and end-of-grade assessments, career and technical education post-assessments, K-2 assessment, and measures of student learning objectives. Gwinnett County used local assessments and also compared school performance to similarly situated schools in 25 national "peer" districts selected specifically for benchmarking purposes. In addition to state assessment data, New York City used growth on both local and state assessments. Prince George's County had seven separate measures of student growth, including state assessments, school-level student learning objectives, district assessments, attendance, grade retention, college- and career-ready participation rates, and VAL-ED.⁵

Denver expected to add two measures of student growth in the future: (1) a status measure that would show the percent of students achieving proficiency; and (2) individualized school performance goals that principals would set in consultation with their supervisors. A district administrator expressed enthusiasm for these additions "because right now, just having one measure on the student growth side is not really [painting] the whole picture." As he explained: "[We want to] make sure that there's lots of different data points for our principal supervisors to consider when they're making a performance decision about their [principals]."

Other Evaluation Measures

TEACHER AND PARENT SURVEYS. Denver, Gwinnett County, and Hillsborough County each included teacher ratings of principal performance in their principal evaluation systems. Both Denver and Gwinnett County gathered teacher ratings by administering a satisfaction survey. In Denver, survey results could be factored into the ratings of professional practice but did not have a specific weight; in Gwinnett County, a survey of both parents and school staff had a weight of 8 percent of the principal's final evaluation score.

Teacher ratings had almost the same weight in a principal's overall score as supervisors' ratings in Hillsborough County, and four years ago teachers' ratings had the same weight as supervisors' ratings. According to a district administrator, principals had long argued that teachers did not have direct knowledge of specific principal competencies that they were expected to rate, especially at the high school level. Among the solutions that had been proposed, and that the district was considering, was to have teachers rate a sub-set of principal competencies, but not all.

⁵ VAL-ED was treated as a measure of achievement on the grounds that it measures principal practices associated with student performance.

Exhibit 9: Student growth measures used to evaluate principal performance, by district^{6, 7}

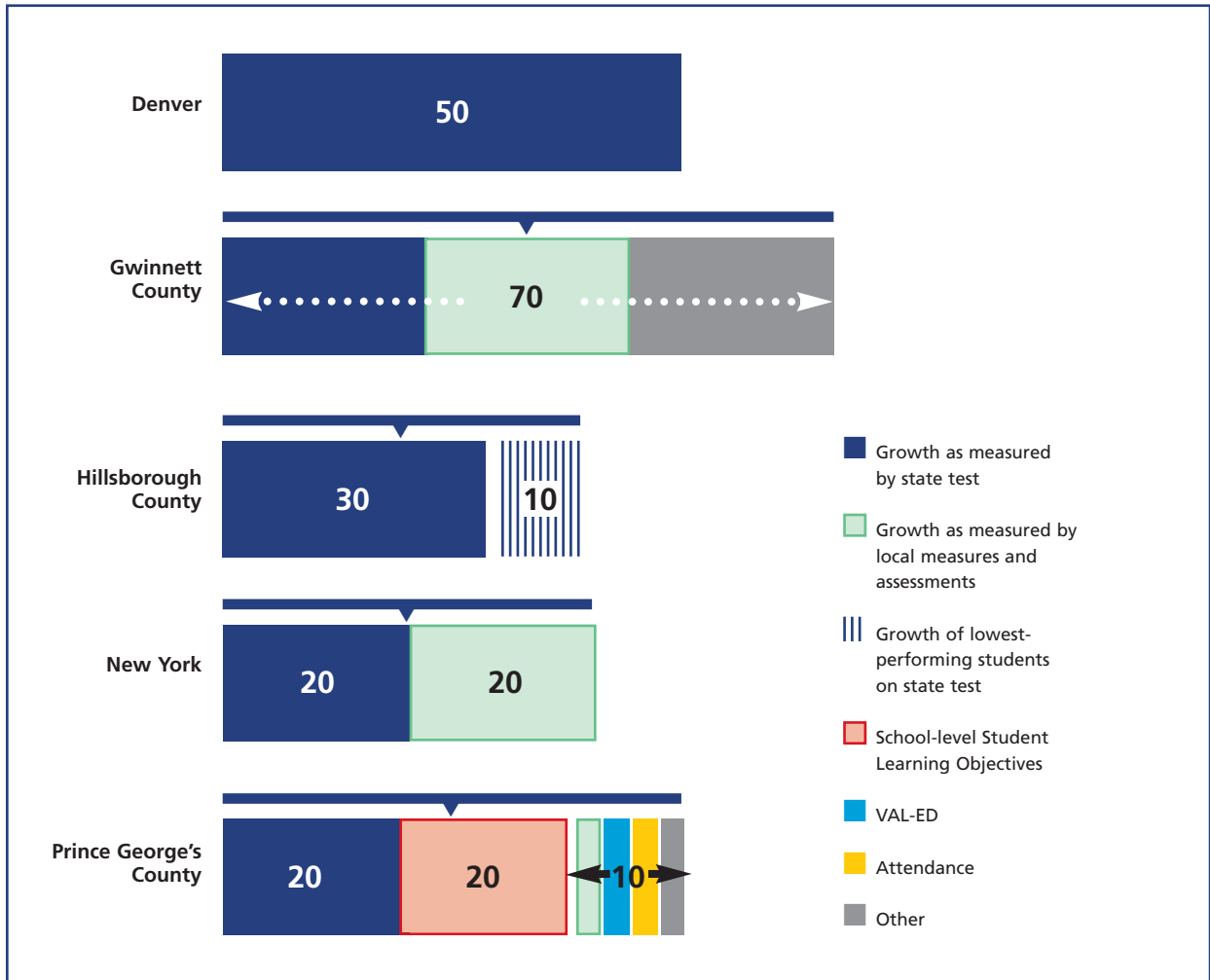


Exhibit reads: Hillsborough County used two measures of student growth to evaluate principal performance, including the overall school-level growth of students as measured by the state test and growth of the lowest-performing students based on the state test. The district weighted the measure of student growth on the state test at 30 percent of a principal's final score and the growth of the lowest-performing students at 10 percent of a principal's final score.

Source: Extant data on system specifications.

⁶ Note that Charlotte-Mecklenburg is not included in the exhibit because it did not use a percentage system in compiling ratings.

⁷ In 2014-15, New York based 20 percent of high school principals' evaluation scores on local measures such as graduation rates and credit accumulation; 40 percent of K-2 school principals' scores on local assessments; and 40 percent of elementary and middle school principals' evaluation scores on student growth as measured by the state test only. Prince George's County included two "Other" measures: one addressing college and career readiness; the other addressing retention in grade. The district did not plan to include state test scores in the first two years (i.e., 2014-15 and 2015-16) of statewide implementation of new assessments. Instead, it placed greater weight for that period on the other measures of student growth. Also, the weights and measures shown are for elementary, K-8, and middle school principals; data for high school principals are weighted slightly differently and include a measure of the year-to-year change in the percent of students retained in 12th grade.

STAFF MANAGEMENT. Two districts incorporated measures of principals' capacity to manage their staff. In Denver, supervisors were encouraged to consider a variety of staff management skills when evaluating principals' professional practice, including: (1) teacher observation completion rates; and (2) handling of human relations decisions, including capacity to hire, develop, and retain effective teachers and to dismiss ineffective ones. Hillsborough County sought to direct principals' attention to human resource management by giving specific weights to their practices in that area. Two percent of a principal's final evaluation score was awarded for practices in each of five areas, including: managing teacher contracts; meeting or exceeding the teacher observation schedule throughout the year and providing ongoing feedback to enhance practice; and conducting observations that were consistently aligned with those of other observers and that correlated with teacher value-added scores. A district administrator explained that before 2013-14, the points awarded in this area were "just a given" in a principal's evaluation. When the central office began taking points away from some principals, "the principals started focusing on it. I think they're really paying attention to those things now, and they should be, because that has a direct impact on student achievement."

SCHOOL OPERATIONS MANAGEMENT. Both Gwinnett and Hillsborough Counties included a measure of school operations in their principal evaluation systems, weighting these skills as distinct parts of a principal's overall evaluation rating. Gwinnett County awarded 10 percent of a principals' final evaluation score for: (1) submitting to the central office high-quality student and staff data in a timely fashion; (2) properly maintaining the instructional materials and physical assets of the school building; and (3) managing the school staff and physical plant. Hillsborough County awarded up to 5 percent of the final score for: (1) maintaining the school resources, such as school equipment; (2) submitting error-free staffing data to the central office; and (3) complying with legal and policy standards.

Final Summary Ratings

The timing for assigning final summary ratings to principals depended on whether and how the district incorporated late-arriving data on student growth. Charlotte-Mecklenburg and Prince George's County assigned ratings at the end of the school year. Gwinnett County, Hillsborough County, and New York waited until student growth data arrived from the state in the fall, some months after the end of the school year for which the principal was being evaluated. Denver changed the entire evaluation cycle from school year to calendar year to allow for use of the data that did not arrive until fall.

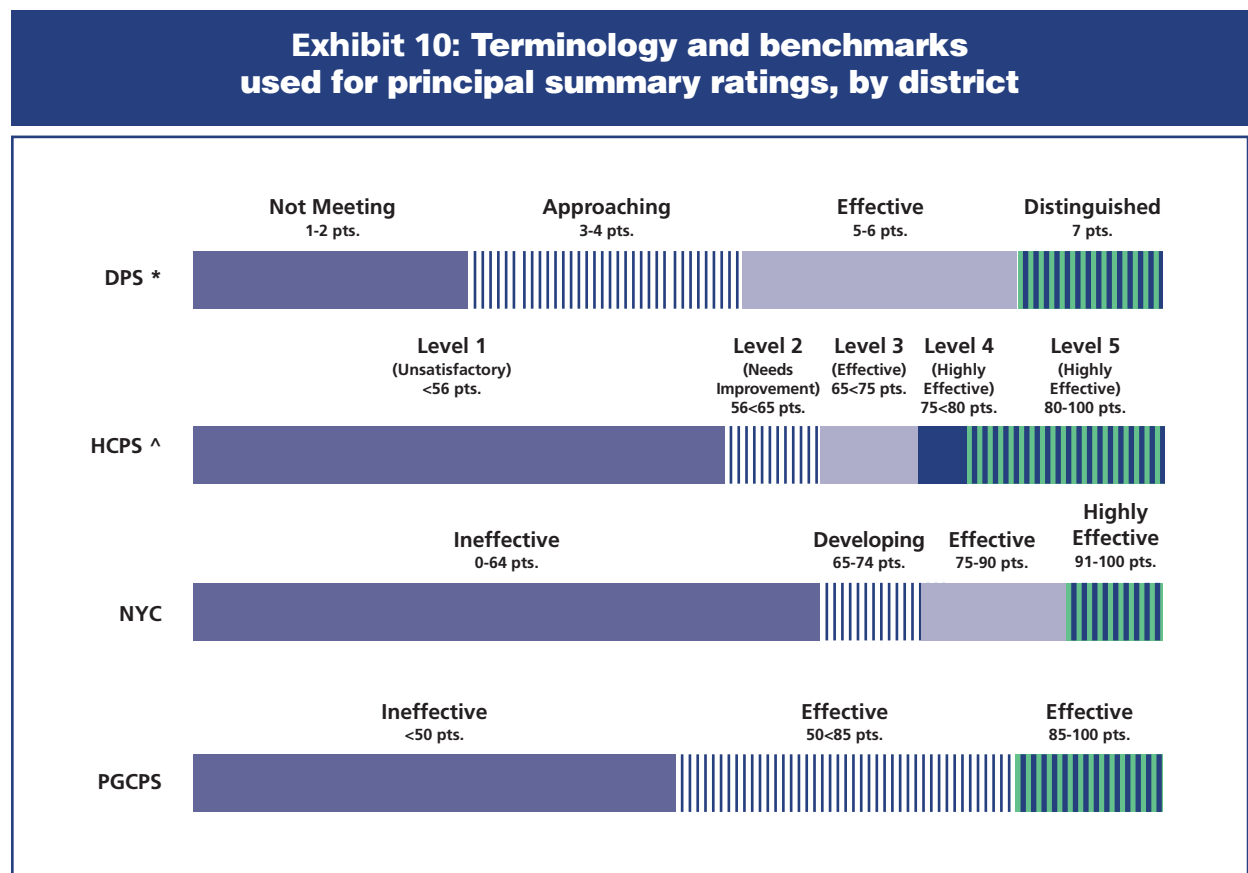
Four of the six districts used a point system to assign principals a final summary rating. As required by all six states, the number of rating categories had evolved beyond a simple "satisfactory/unsatisfactory," with three to five categories used in 2015. Denver, Hillsborough County, New York, and Prince George's County assigned points, usually ranging from zero to 100, to each principal based on that principal's weighted and aggregated scores for professional practice, student growth, and, in Hillsborough County, other leadership measures (Exhibit 10). These districts set benchmarks for each of the rating categories based on a point value.

In Charlotte-Mecklenburg, principals received a rating of "Needs Improvement," "Effective," or "Highly Effective" only after they had been an administrator in the district for three years, at which point the district would have three years of data on student growth under that principal's leadership. Consistent with

state requirements, though, the district set a relatively high bar for achieving an effective rating, requiring that principals achieve the middle rating on a five-point scale on all of the seven professional practice standards as well as meeting expected student growth targets.

Gwinnett County did not compile principals’ scores across all four components of its evaluation system (professional practice, school operations, parent and teacher satisfaction surveys, and student growth). Rather, it assigned a rating for professional practice (Exemplary, Proficient, Needs Development, and Ineffective) and a separate score for student growth (on a 100-point scale) based on its Weighted School Assessment.

Student growth measures could trump other measures in the evaluation systems of three districts, Denver, Gwinnett County, and New York. Gwinnett County demonstrated the value it placed on student growth through its weighting system, which allocated 70 percent of a principal’s score to the measure of student



* Points converted to a 100-point scale

^ Reflects 2013-14 benchmarks. Point distribution and ratings benchmarks recalculated every year. For purposes of state reporting, Hillsborough County converts its 5-point scale to the state's 4-point scale by combining Levels 4 and 5 into the state's "Highly Effective" rating category.

Exhibit reads: Denver Public Schools assigned principal summary ratings based on a four-part scale that ranges from "Not Meeting" (based on a summary score of 1-2 pts.), to "Approaching" (based on a summary score of 3-4 pts.), to "Effective" (based on a summary score of 5-6 pts.), and "Distinguished" (based on a summary score of 7 pts.).

growth. As one district administrator explained, continuous improvement in student performance was a major priority for the district: “Understand that the [Weighted School Assessment] is universally understood—or ought to be universally understood—as an incredibly significant measure of the school and principal’s performance. It just is, and has been for a long time.”

A supervisor’s recommended professional practice rating for a principal could be overridden in Denver if the central office determined that the school was not showing sufficient gains in student achievement to justify an “Exemplary” professional practice rating. This did not often happen, though, according to a district administrator:

I think there were somewhere between four and eight recommendations that were overridden.... It’s a rarity that we would override that decision or that recommendation. Principal supervisors [are] in schools every week so they know more than what’s on a spreadsheet and so they have the discretion and the ability and the empowerment to make the final call. But it’s rare that it deviates from what the [student growth] data’s saying anyway.

Similarly, New York’s principal evaluation system required that if principals were rated as “Ineffective” on both the state and local measures of student learning, then they must receive an overall rating of “Ineffective,” potentially overriding whatever rating they had received on their professional practice.

Evaluation Procedures

The Principal Pipeline Initiative design did not spell out evaluation system designs or procedures, but it called on districts to ensure that principals received constructive feedback from evaluation. From the beginning of the initiative, whether a district was designing a new evaluation system or refining an existing one, district leaders spoke of their efforts to keep principals informed and engaged.

The steps that went into evaluating principals’ professional practice were designed in each district to unfold something like a conversation between principals and their supervisors. In contrast to systems of teacher evaluation, which have tended to rely on a set number of formal observations to assess practice, these principal evaluation systems had procedures for goal setting and evidence gathering that were designed to ensure ample opportunities for two-way discussion. Feedback for principals was infused into the process of data collection in hopes of supporting principals’ professional growth. Final performance ratings then directed principals’ attention more formally to areas of needed improvement and, if necessary, the districts placed principals on an action or improvement plan when their performance was rated as ineffective or poor. Every district also acknowledged that dismissal could ensue from a low performance rating, although this rarely happened.

Goal Setting

Principal self-assessment and goal setting was the first step in four districts' evaluation process. This did not mean that principals were able to pick and choose which of the district standards they would have to meet. Rather, it reflected a common practice in employee performance assessment: because research indicates that “participation in the goal-setting process tends to enhance goal commitment” (Lunenburg, 2011, p. 3), organizations often ask employees to reflect on their areas for improvement, to work with their supervisors in specifying improvement goals, and to commit to a plan. It also reflected a recommendation drawn from the literature on principal evaluation (Davis et al., 2011). Most district administrators, supervisors, and principals said that these individual discussions helped supervisors focus their feedback and support. In Charlotte-Mecklenburg, Denver, Hillsborough County, and Prince George's County, principals met with their supervisor (and also, in Charlotte-Mecklenburg, their coach) to review and discuss their self-assessments and to set professional goals for the year. In Hillsborough, for example, the district provided a template, the School Leader Goal Setting and Professional Learning Plan, for the principal and supervisor to complete together. In that template, they would define school improvement goals, principal improvement goals, action steps, and sources of support (e.g., professional development and supervisors' targeted assistance). A district administrator explained that the goal-setting process—along with some targeted supervisor training—had helped supervisors focus on improving principal practice.

...what do [principals] have to do better [related to the] competencies in order to improve teachers in order to improve reading achievement? What do they have to do differently?

Principals in Gwinnett County and New York set goals for the school year, but they focused on school-level goals rather than individual ones for themselves. In Gwinnett County, the principal worked with teachers, parents, and community members to develop a Local School Plan for Improvement, setting targeted goals based on student achievement data and identifying strategies for achieving them. In New York, principals developed a Comprehensive Education Plan in consultation with the school leadership teams, setting instructional goals for the school year and defining how the school budget aligned with those goals.

Evidence Collection

Supervisors in all districts spent considerable time in schools. They were expected to collect a variety of evidence from a variety of sources to document and rate principals' professional practice. Generally, this meant school visits that involved discussions with principals, teachers, and other school staff; observations of principal performance, including principal interactions with school staff; and reviews of artifacts, including plans for school improvement and teacher professional development.

Unlike teacher evaluation systems that require principals to observe teachers a pre-determined number of times, the evidence-collection process for supervisors had no set number of interactions. And unlike principals, who spend about 10 percent of their time on teacher observations (Horng et al., 2009), supervisors in most districts collected evidence of principal practice all year long. In Denver, supervisors were in schools

as many as 40 times a year, which, a district administrator explained, gave them a very clear view of what was happening in each building. A district leader in Prince George’s County described evidence collection as ongoing and unceasing for principal supervisors: “I can’t even calculate how many hours [supervisors] spend [collecting evidence] because they’re looking at that principal from the beginning of the year through the end of the year and collecting data every time they go into those buildings.”

Determining the type and amount of evidence required to substantiate a particular principal rating was still a work in progress in most districts. A district administrator in Prince George’s County, for example, said supervisors were just beginning to make changes in the evidence they used to evaluate principals. “I’m more confident next year that there will be more uniformity to what all [supervisors] are collecting. And there’ll be a little bit more uniformity to the [feedback] conversations [they have with principals].” A principal supervisor in Denver said that while he recognized that the principal evaluation rubric was creating a common language around definitions of principal leadership, the evidence guide connected to that rubric was still “growing.”

Districts’ principal evaluation rubrics often included examples of principal practice to illustrate evidence of principal effectiveness, but they did not explicitly define the amount and type of data, by standard, that a supervisor would need to assemble in order to evaluate principals accurately. For example, one district defined an “Accomplished” rating on a particular dimension of instructional leadership as a principal who “Helps teachers and staff set, monitor, and achieve challenging goals based on student outcomes; anticipates, adapts, and persists in the face of obstacles and responds in a positive solutions-oriented manner.” The rubric did not, however, describe the type of data a supervisor would use to measure this competency.

Some district interviewees acknowledged that they might need to work on better defining the evidence to support effectiveness ratings. As one district administrator explained:

If our standards are used as checklists, I’m not convinced they actually give us good evidence of effectiveness as a leader. ... We have evidence referenced in our framework, but it’s like “leads for,” “communicates with energy about,” or “calls out inequities.” As a principal supervisor, can you really be there enough? Like, what would be my evidence of that? So I think the standards are really important because they give us a [common] language. I think the next step is figuring out what is truly evidence of effective leadership.

Hillsborough County appeared to be moving in the direction of an evidence guide. An administrator described the district’s recent efforts to develop a “critical attributes” document to help trigger supervisors’ thinking about what principals should be doing if they are “exemplary.” Still, the administrator worried that a catalog of attributes might lead supervisors to approach the evaluation task as a perfunctory exercise.

Prince George’s County and Charlotte-Mecklenburg had gone farther in defining the evidence needed to support a principal rating. Prince George’s County’s evaluation rubric, for example, listed evidence that a supervisor could use to support a rating for each element of practice on which principals were evaluated.

While the evidence lists were not comprehensive and did not substitute for supervisor judgment, they did include several concrete examples of objective evidence on which to rate principal performance, such as:

- Guidelines for the use of core, supplemental, and intervention curricula exist and are understood by all staff members.
- Materials-in-use are regularly reviewed, updated, and replenished to ensure classrooms are adequately supplied.
- When targets are set for student achievement, appropriate training is provided for all staff members to meet these targets.
- E-mail, memoranda, and other communication protocols are developed to streamline and reduce information received by staff members.
- Time is provided for staff members of common subjects and grade levels to share and discuss curricular expectations.
- All staff members possess individual professional development goals and are making progress toward them.
- The Principal frequently provides written feedback to staff members based on observations of instructional practice and common planning time.

Assigning Principal Ratings

There were no clear-cut procedural rules about how supervisors used evaluation evidence to assign principal ratings. In most districts, supervisors were described as having significant control over the decision-making process. Denver and Hillsborough County, for example, encouraged supervisors to take into account a variety of data when rating their principals, but gave them full discretion in assigning values to each of these measures. As a district administrator explained:

We don't have a lot of rules or metrics around how to look at the data. We do provide a lot of guidance and we do a lot of training and calibration but it's not like if a certain percentage of families feel a certain way about a school then that translates into a lower score in community leadership.

Another district leader gave an example of how supervisors were expected to exercise discretion in using data to assess principal performance:

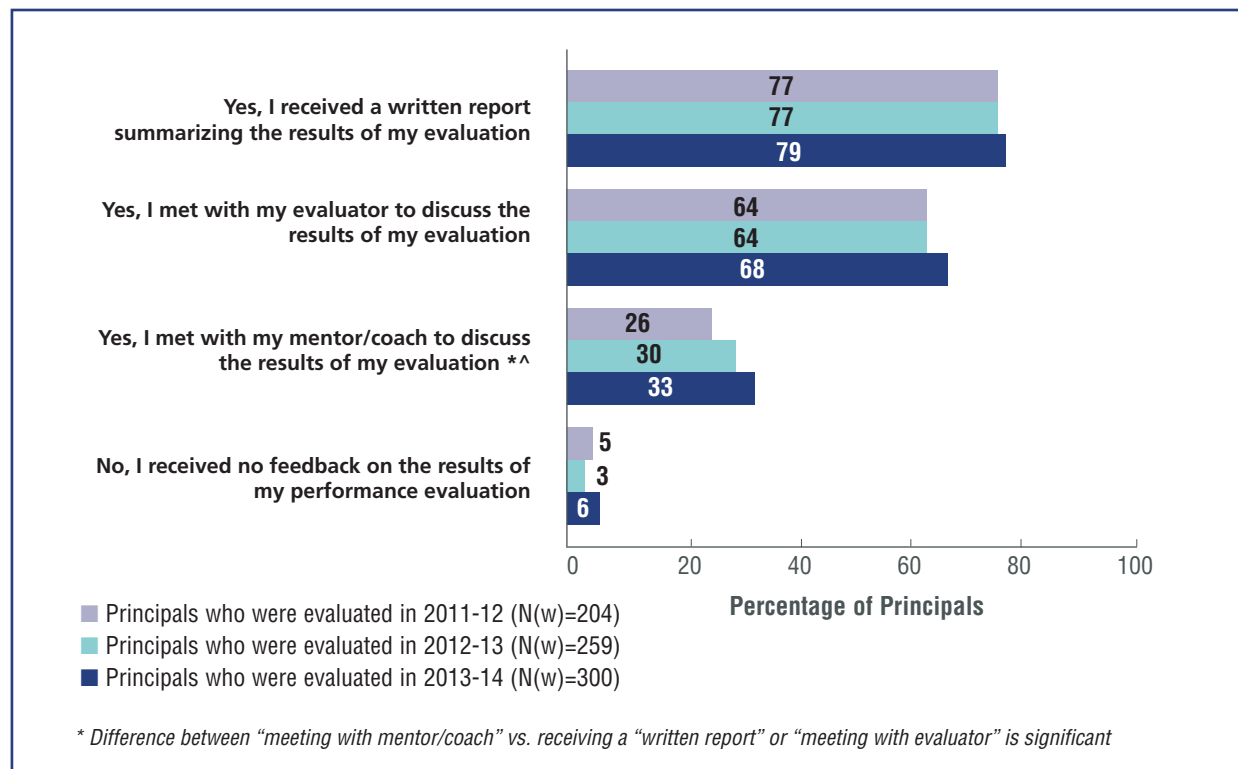
...teacher retention is part of our rubric. The supervisor could use data [on the teacher retention rate for a school], but it has to go beyond data because if a principal comes in the school in a turn-around situation and they lose teachers, they shouldn't be penalized. So on paper, it might look like you're not doing a good job of retaining teachers, but in reality, [dismissing teachers] was needed.

Providing Formal Feedback

No district leader was able to provide the exact number of principals who fell into each rating category in 2013-14. However, all described the majority of principal ratings as falling in the middle—as Developing, Approaching, or Effective—with few principals falling into the very lowest or very highest performance categories. One district leader’s explanation was typical: “There’s four buckets you can fall into and we don’t have a ton of folks at the tails. So it’s really deciding whether a principal is an effective or slightly less than effective [leader].”

Survey data suggest that principals were indeed receiving feedback from their supervisors regarding the results of their evaluation (Exhibit 11). Over 90 percent of principals who were evaluated in 2011-12, 2012-13, and 2013-14 reported having received some form of feedback, either written or verbal, on the results of their evaluation. Three-quarters of principals who were evaluated in 2011-12, 2012-13, or 2013-14 reported receiving written feedback on the results of their evaluation, and approximately two-thirds reported meeting with their supervisor to discuss their evaluation results.

Exhibit 11: Percent of principals reporting that they received feedback on the results of their evaluation in the previous year, by cohort



^ Not every principal who was evaluated in 2011-12, 2012-13, and 2013-14 had a mentor or coach. The weighted numbers who did were 129(w) in 2011-12; 166(w) in 2012-13; and 166(w) in 2013-14.

Exhibit reads: Seventy-seven percent of novice principals who were evaluated in 2011-12 reported receiving a written report summarizing the results of their evaluation compared with 77 percent of novice principals who were evaluated in 2012-13 and 79 percent who were evaluated in 2013-14.

Supervisors also described delivering feedback to their principals on an ongoing, informal basis in addition to the required conferencing. This was consistent with the district leaders’ desire to establish a continuous cycle of principal observation, support, and feedback. As one Denver administrator explained, principal evaluation and support are “inextricably linked,” and supervisors diminish the potential power of evaluation if they only pull out the rubric twice a year to assign a score:

We have a pretty robust framework for our principals and if we’re not using that as a planning tool, a coaching tool, a development tool, support, a way to model, etc., etc., a way to create common language in our building. If we’re not doing all that then we’re missing the broader purpose of why this work is so valuable.

Principal survey data suggest that the quality of the feedback principals received has been consistently good over time. Over three-quarters of the principals who were evaluated in 2011-12, 2012-13, or 2013-14 and who received feedback reported that the quality of the feedback they received was good or excellent (Exhibit 12).

Exhibit 12: Principal ratings of the quality of the feedback they received on the results of their evaluation in the previous year, by cohort

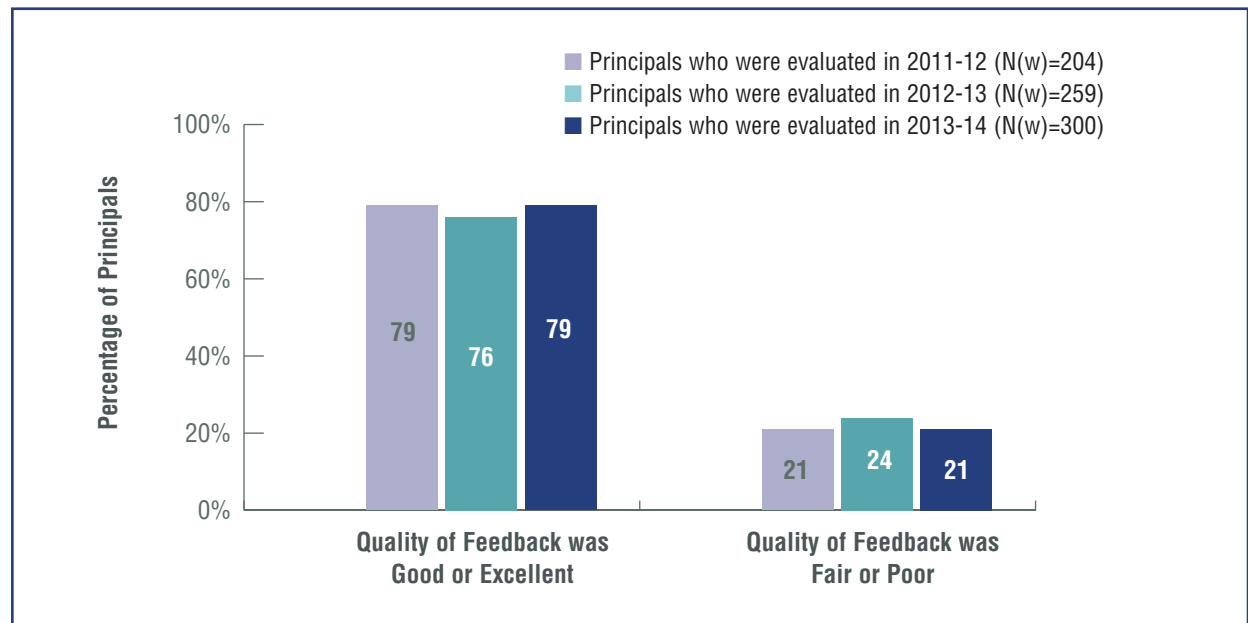
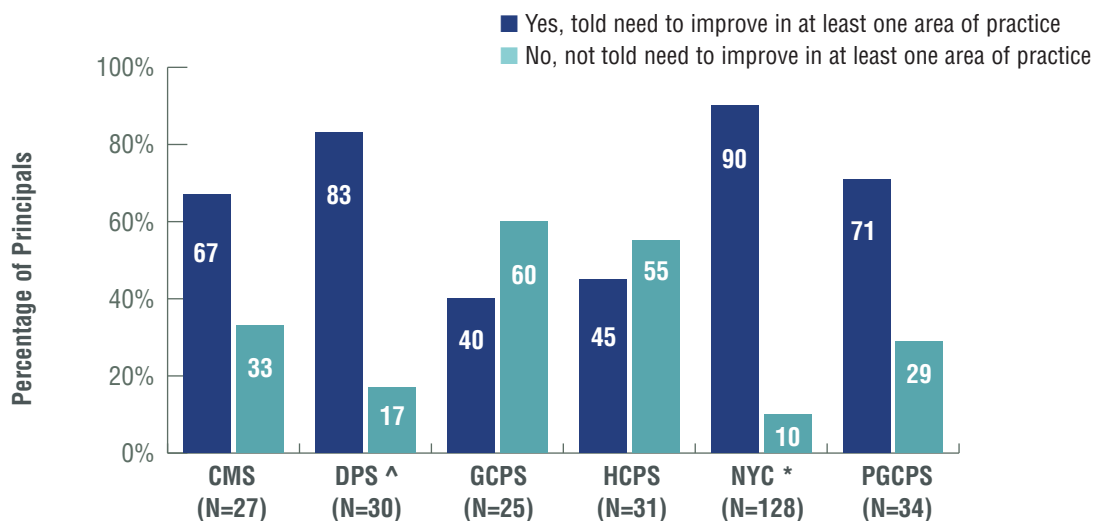


Exhibit reads: Seventy-nine percent of novice principals who were evaluated in 2011-12 reported that the quality of the feedback they received was good or excellent compared with 76 percent of novice principals who were evaluated in 2012-13 and 79 percent who were evaluated in 2013-14.

The new evaluation systems gave some direction for improvement to most novice principals. More than half of novice principals were told after their 2013-14 evaluations that they needed to improve in at least one area of practice, according to surveys administered in 2015. This was the case for 69 percent of principals who had been evaluated in their first year as principals in 2013-14 and 59 percent of those who had been in their second year. Of these principals, approximately half were told they needed to improve in only one or two areas, but approximately 20 percent were told they needed to improve in five or more areas of professional practice.

There were differences across districts in the percentage of novice principals who were told that they had areas for improvement (Exhibit 13). We have no evidence suggesting that this reflected the quality of the principals' work, however. The districts may have had different expectations for the extent to which supervisors would tell novice principals they needed to improve. Ninety percent of novice principals in New York who were evaluated in 2013-14 reported that they were told they needed to improve in at least one area of practice, a statistically significant difference from the 67 percent of novice principals in Charlotte-Mecklenburg, 45 percent of novice principals in Hillsborough County, and 40 percent of novice principals in Gwinnett County. The difference between Denver (83 percent of principals told they needed to improve) and Hillsborough County (45 percent) and Gwinnett County (40 percent) was also statistically significant.⁸

Exhibit 13: Principals told they needed to improve in at least one area of professional practice based on evaluation results for 2013-14, by district



* Difference from CMS, GCPS, and HCPS is significant

^ Difference from GCPS and HCPS is significant

Exhibit reads: Sixty-seven percent of novice principals in Charlotte-Mecklenburg who were evaluated in 2013-14 reported they were told they needed to improve in at least one area of their professional practice based on the results of that evaluation.

⁸ On reviewing this report for factual accuracy, Hillsborough County pointed out that in another sense 100 percent of principals in that district were told that they needed to improve since all were required to write two goals for improvement in their own performance.

Instructional leadership was the area in which the novice principals most commonly reported having been told they needed to improve (Exhibit 14). Among those novice principals who were given any areas of improvement after a 2013-14 evaluation, about half reported on the survey that they had been told they needed to improve in at least one of the following practices, which we classified as instructional leadership:

- Helping teachers improve their professional practice
- Pressing teachers to set high standards for student learning
- Using data to inform instruction
- Tracking student academic progress
- Conducting instructional planning with teams of teachers

Exhibit 14: Areas of professional practice that principals were told they needed to improve based on evaluation results in 2013-14
N(w)=300

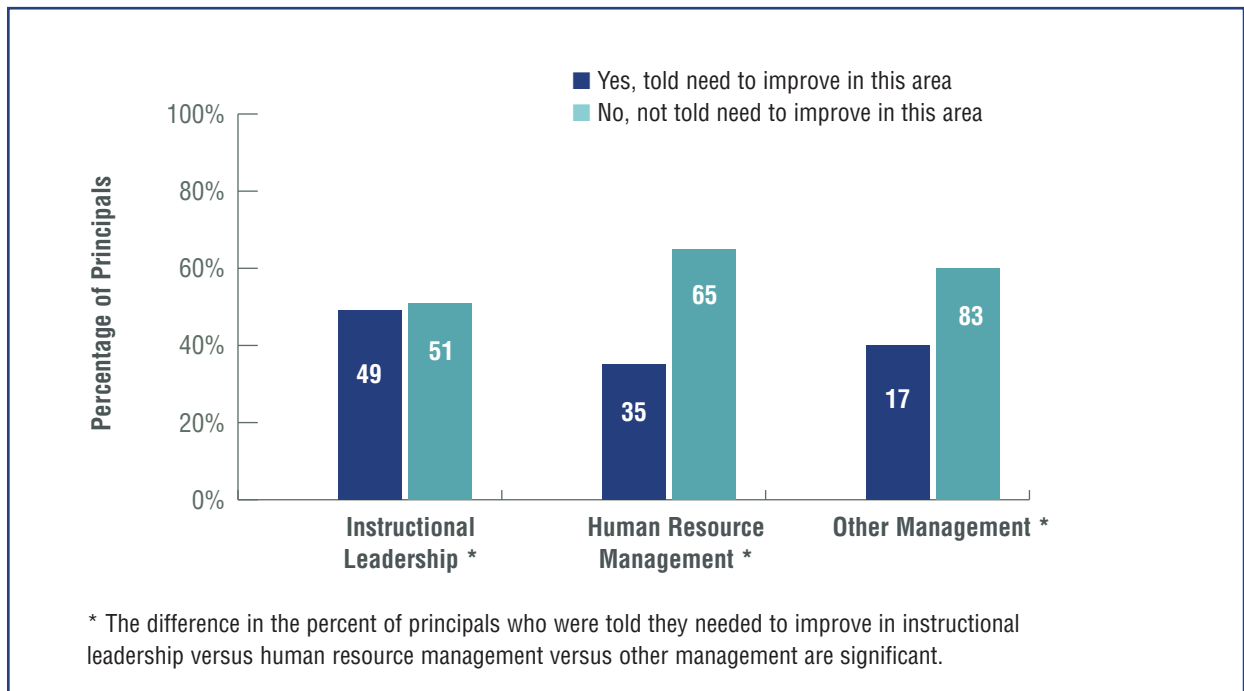


Exhibit reads: Forty-nine percent of novice principals who were told they needed to improve in at least one area of their professional practice based on the results of their evaluation in 2013-14 were told they needed to improve in instructional leadership.

On average across districts, approximately one-quarter of novice principals reported being told they needed to improve in at least one of these practices (that is, of the half who reported they were given any direction for improvement, half reported that it was in at least one of these instructional-leadership practices). The largest percentages reported needing to improve in interaction with teachers, including helping teachers improve their professional practice (30 percent of all novice principals) and pressing teachers to set high standards for student learning (29 percent). There were differences across districts, paralleling the overall differences in the extent to which the districts gave critical feedback to their novice principals. Fewer than a quarter of principals in Gwinnett County and Hillsborough County reported having been told they needed to improve in any area of their instructional practice.⁹ In Charlotte-Mecklenburg, Denver, New York City, and Prince George's County, by contrast, more than half the novice principals evaluated in 2013-14 reported having been told they needed improvement in at least one area of instructional-leadership practice.

Human resource management was an area of improvement called to the attention of about one-sixth of all novice principals after their 2013-14 evaluation. This area included such competencies as hiring personnel, releasing or counseling teachers to resign their positions, retaining highly effective school staff, addressing staff concerns, and developing staff leadership capacity. Fewer were told they needed to improve their other management skills, such as managing their own time on the job, managing budgets and resources, or maintaining campus facilities.

Consequences of Summary Ratings

Every district's principal evaluation system required that principals who received a low rating or score must also receive an action or improvement plan. Furthermore, leaders in every district said that principal dismissal could be a consequence of evaluation results. Nevertheless, district leaders were unsure about or legally prohibited from sharing in interviews the number of novice principals to whom these consequences were applied.

IMPROVEMENT PLANS. Despite the requirements for action plans or improvement plans to follow low ratings, there were no hard and fast rules about the improvement process and where it might lead. Hillsborough County principals, for example, were invited to select someone, such as a coach or administrator, to provide support in a specific area. In Prince George's County, principals in need of improvement had ongoing conversations with their supervisors and were given time to improve.

DISMISSALS. In Charlotte-Mecklenburg, Denver, and Prince George's County where summary evaluation data had been available for at least one full year in spring 2015, it was as yet unclear how many principals were dismissed as a result of their evaluation ratings. Every district stated, though, that dismissals were among the consequences imposed on principals based on their evaluation results. In New York, for example, state legislation said that after a principal receives two ineffective ratings, "termination may be considered" [NY-6, sec, 3012-c (6)]. Novice principals in their first probationary year, however, could be dismissed by their supervisor based on performance evidence. In Denver, where principals

⁹ However, a Hillsborough County district leader notes that most principals in the district set at least one goal for improvement in the area of instructional leadership.

had neither tenure nor collective bargaining rights, a district leader explained that “the vast majority” of principals who received an ineffective rating were “usually reassigned or not invited back to be in that role the following year.”

Leaders in one district said that they knew of no new principal who had been dismissed at the end of their first two-year contract and few who had been dismissed at the end of their next four-year contract. An administrator said that there were not “hard and fast” rules about improvement: “We don’t have a matrix where if you have two ‘developings,’ you’re not renewed or something like that.” Still, another district leader explained that although there were few dismissals as a direct consequence of a final evaluation score, the district did have attrition through the contract cycle that was the result of performance management: “We have very frank conversations with principals and tell them that they are not progressing and that if it continues [they will be dismissed].” As one district leader explained: “A lot of them make the choice [to leave], so our [dismissal] numbers wouldn’t always show that.”

Supervisor Training for the Evaluation Role

All districts were working with their principal supervisors on calibration, that is, an effort to increase consistency and fairness in the principal evaluation system. Hillsborough’s principal supervisors had participated in weekly calibration exercises for a few months as of spring 2015. The exercises focused on one element of one competency area at a time, discussing the type of evidence supervisors should amass to rate the element. A district administrator explained that the purpose was to ensure consistency across supervisors in the amount and type of evidence supervisors collected to support their decisions about ratings. A related purpose, she said, was to help supervisors understand that an “artifact” is not “evidence”—that the existence of a professional development plan, for example, is not sufficient evidence to judge a principal’s efforts to develop staff, without evidence of the plan in action and its results. She also said she hoped that the calibration exercises and discussions would lead to better differentiation between the rating levels, and described another district administrator questioning supervisors’ ratings, asking, “Wait a minute. Why is that ‘accomplished’? I have that in ‘progressing.’ Look at the words.”

To determine whether supervisors were consistent in their assessment of principal practice, New York City set up evaluation simulations. Supervisors reviewed videos and artifacts and assembled evidence to rate each indicator of leadership practice. They shared with each other their most compelling evidence and rationale for assigning a particular rating, and members of the group discussed whether they agreed. The purpose of the exercise was to determine whether the group was consistent in its assessment of principal practice, raising questions such as: “Are we on the same page? Are we developing a consistent lens for looking at evidence through the rubric? Are we providing a coherent and consistent experience for the field when we go out there?” A district administrator had seen in the ongoing, cumulative training that supervisors were developing deeper insights into the rubric and how to distinguish “effective” from “highly effective” principal practice. “Because of the nuances of the rubric, every time we come across the same indicator, we might come up with something a little more specific,” the administrator said.

Prince George’s County and Denver also had supervisors work in small groups. Supervisors reviewed sample evidence, assigned principal ratings based on the standards, and discussed their ratings as a group, each explaining their rationale for the rating they assigned. A district administrator in Prince George’s County said that most supervisors calibrated very closely and were developing a common understanding of the behaviors they expected to see in principals and in schools. A district administrator in Denver, however, acknowledged that four calibration sessions a year was not enough, saying that supervisors needed more training to ensure adequate alignment with each other in evidence collection and interpretation.

Calibration training for supervisors was less intense and structured in Charlotte-Mecklenburg and Gwinnett County. There, informal team meetings focused on a review of the evidence and a discussion of whether a particular principal demonstrated effective practices, and why. A Gwinnett County administrator said supervisors would visit schools together, talk to principals, and observe classrooms. “We’ll talk about how we’re seeing classroom instruction, about the instructional leadership of the principal. ... We’ll bring a couple of case studies to supervisor team meetings and discuss the ratings.” District administrators in Charlotte-Mecklenburg commented that because of high turnover among the principal supervisors, the district needed to continue to work on educating principal supervisors on the evaluation tool.

Principal Perspectives on System Design

Among the districts in the Principal Pipeline Initiative, principals gave consistently positive reviews of the quality of their evaluation systems, suggesting a high rate of principal buy-in. On survey questions that were repeated across two or three years of this study, the reviews did not change significantly over time. Most principals who had been evaluated in 2013-14, in every district, agreed that the evaluation system was accurate and fair, provided a common language for professional practice, set clear and consistent expectations for their performance, and was generally useful for purposes of informing their professional practice and identifying areas in which they needed to improve, according to our survey (Exhibit 15). Principals’ responses were gathered with a five-point scale, and the top three responses are shown in the exhibit. Perhaps the most striking finding is that 85 percent of the principals generally agreed that the evaluation “provided results that were worth the effort.” This finding is significant in light of the fact that the districts’ evaluation systems varied significantly in scope and complexity. It is also striking when compared with earlier research which found that, on the whole, principals saw limited usefulness of performance evaluation with respect to feedback and professional learning (Portin, Feldman, & Knapp, 2006).

**Exhibit 15: Principals evaluated in 2013-14 who agreed that their district's evaluation system was accurate and fair, provided clear expectations for performance, and was useful and worthwhile
N(w)=300**

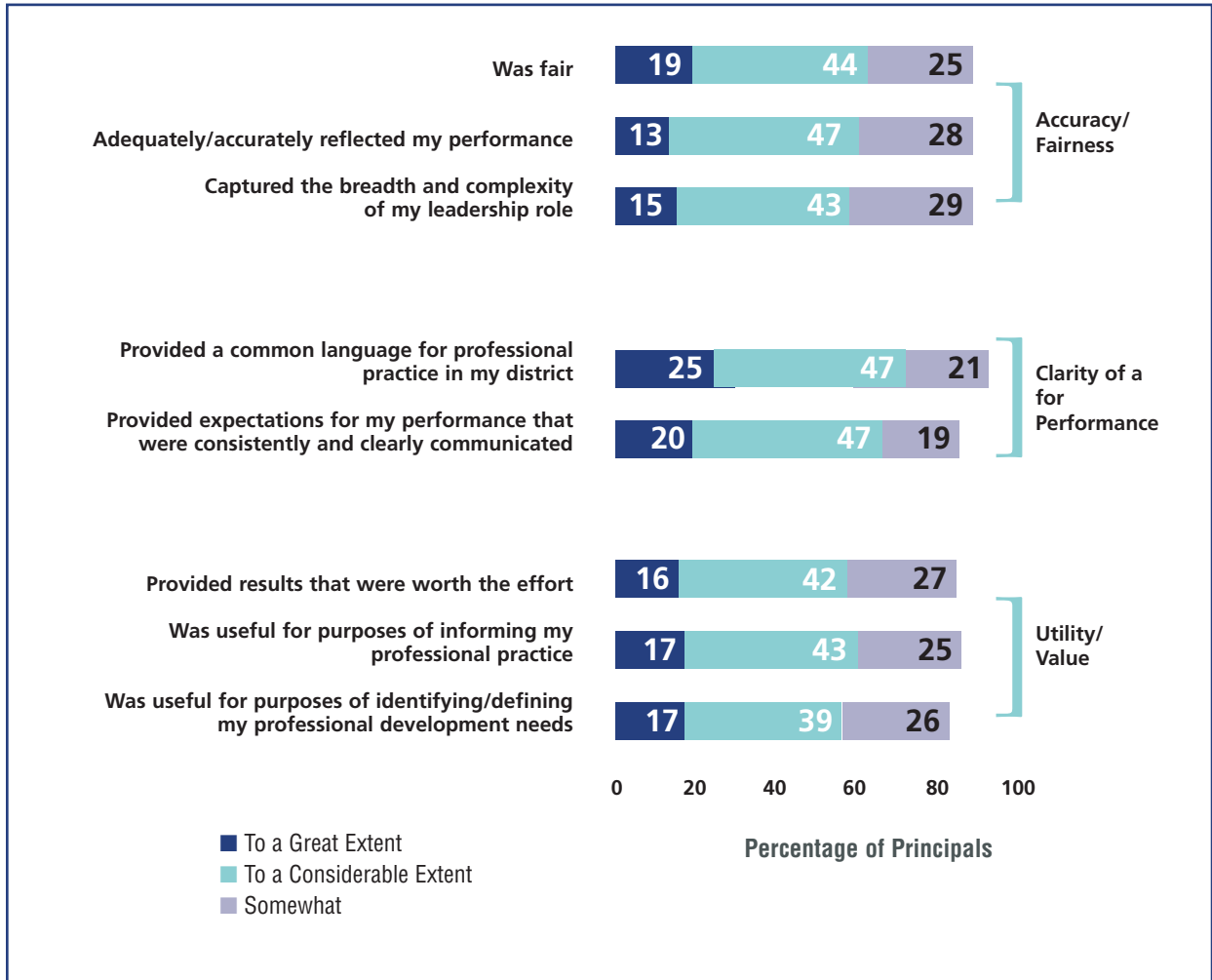


Exhibit reads: Eighty-eight percent of novice principals who were evaluated in 2013-14 agreed that their district's evaluation system was fair, saying they agreed at least "somewhat" with the statement. The other responses, not shown in the exhibit, were "not at all" and "minimally."

Consistent with the survey results, novice principals participating in focus group interviews expressed general approval of their districts' evaluation systems. Some described particular strengths of these systems, including the regular visits from their supervisors and the feedback conferences that were built into the evaluation procedures. In Hillsborough County, a principal described the benefits of the regular feedback: "... it allows us to progress monitor—it's a pulse check." A principal in Gwinnett County, echoing the intentions expressed by many district leaders across the six districts, explained that he did not think of "this big evaluation" but "an ongoing conversation all the time about what are your goals, how are you working toward those goals, and are you making progress or not. So it's not sit down and have one meeting and be evaluated with feedback for next year because it's an all-the-time conversation."

While generally satisfied with the evaluation system, some principals participating in focus group interviews in every district offered suggestions for making more explicit the evaluation rubrics used to measure their professional practice. One principal, for example, said that the rubric helped him know the standards that he needed to meet, but was too generic and unrelated to specific activities to generate useful, authentic feedback: "I feel like [the evaluation is] just a hoop we jump through. Feedback for me is much more authentic when it's on something that's truly authentic to what I'm doing, whether it's a new initiative, or just guiding my thinking around something I'm doing." In another district, a principal explained that the district needed a rubric that better defined what highly effective professional practice looked like. He wanted examples that would be clearer than his supervisor's definition of highly effective as someone who goes "over and beyond" the job requirements. "I'm not sure if [supervisors] are even clear about what 'over and beyond' actually means," the principal said. And a principal in a third district criticized the evaluation rubric that held his work in a turnaround school to the same standards of performance as other principals who were in higher performing schools and had been in their schools longer:

... so coming in as a new principal this year with no [grace period], the walls are really high [but] you're getting this "Go slow to go fast; build relationships" message and you're like, "Great, that doesn't perfectly align to the framework of how I'm evaluated."

Principals in Hillsborough County and Denver expressed frustration about the inclusion of teacher ratings in their evaluations. One principal argued that unlike teachers, his supervisor had been a principal and knew what the job entailed. Principals in both districts explained that they might work under circumstances that necessarily upset teachers, such as a turnaround school, yet the evaluation system did not account for those circumstances. As one principal explained:

...if on my teacher perception survey lots of people are [upset], well, last time I checked, [the district] wanted me to [upset] lots of people. And I don't have another way to get rid of some of them, so that would not be surprising if half my staff leaves at the end of the first year, which is typical in a turnaround [school]. But then that looks bad because half the teachers were unhappy.... To me, I don't understand how all those pieces play together to give you a picture of the work.

Summary

Leaders in all six districts hoped that principal evaluation could be a tool to guide principals' attention and learning, with positive effects on principal effectiveness and school functioning. These hopes were consistent with the aims of the Principal Pipeline Initiative. At the same time as they were implementing that initiative, districts followed their states' specific requirements for evaluation-system design by including measures of student growth and professional practice. Within these broad categories of measures, each district made distinctive choices about what would be measured and how it would be weighted. Most of them continued to revisit these choices from year to year.

The districts grounded their measures and rubrics for professional practice in principal standards and competencies that they had developed under the Principal Pipeline Initiative to guide not only evaluation but also principal preparation and hiring. The districts were in the process of translating the standards and competencies into clear and complete rubrics for use in evaluation, and also continuing to work on guidance for the principal supervisors to use in gathering evaluation evidence.

Among novice principals in each district, at least 75 percent agreed that their evaluation system captured the breadth and complexity of their leadership role and adequately/ accurately reflected their performance. More than half reported that they had received feedback indicating that they needed to improve in at least one area of practice.

Instructional leadership was the area in which the largest number of novice principals were told that they needed to improve. In particular, relatively high proportions received critical feedback on their performance in helping teachers improve their professional practice (30 percent of all novice principals) and pressing teachers to set high standards for student learning (29 percent). This is likely to mean both that novice principals had more to learn about instructional leadership and also that principal supervisors were paying particular attention to this area of leadership practice.

The districts generally treated their evaluation data as a resource for the individual learning and management of principals rather than for districtwide analysis to support policy choices. Some interviewees mentioned the possibility of analyzing the overall evaluation data in order to identify the common weaknesses and give them greater attention in principal preparation, especially in the programs that districts ran themselves. In general, though, this was something they said they hoped to do in the future.

Districts continued to refine procedures and build capacity for evaluation. Principal supervisors were meeting—more often in some districts than others—to work on consistency in their ratings. Districts also saw a need to improve the feedback and support that principals received from supervisors, as we discuss in the next chapter.

3. Principal Support

Districts worked to build support systems that could help principals learn from their evaluation results. Just as they worked to prepare principal supervisors to use new evaluation instruments, they were also striving to prepare supervisors to assist principals directly and, often, to help principals find other assistance from mentors, coaches, or district professional development.

Novice Principals' Experience with Support

Among principals who were told they needed to improve in at least one area of professional practice, most reported on the survey that they had received support in that area. Eighty-six percent of those who were told they needed to improve their instructional leadership skills reported having received support, and the percentages were similar for human resource management (82 percent) and for other management skills (86 percent). There were no statistically significant differences in the reported rates of support between principals who were first-year principals in 2013-14 and those who were second-year principals in that year.

Takeaways

1. Most novice principals reported receiving support in those areas for improvement that evaluation had flagged.
2. Districts were redefining the principal supervisor job around giving feedback and supporting improvement.
3. Most novice principals reported that their supervisors were knowledgeable and addressed their specific needs.
4. Professional development offerings were less helpful than support from supervisors, mentors, or coaches, according to principals.
5. Novice principals highly valued their help from mentors or coaches, and those who had mentors or coaches rated all sources of support more highly than those who did not.

Redefining the Supervisor Position to Focus on Supporting Principals

Districts were changing the role of principal supervisor, focusing it on helping principals develop and improve. This was an aim of the Principal Pipeline Initiative from the beginning, and it gained added momentum when a multi-district study by the Council of the Great City Schools, commissioned by The Wallace Foundation, concluded that principal supervisors in large urban districts often lacked the time and training to provide substantial support to principals (Corcoran et al., 2013). The foundation gave each Principal Pipeline Initiative district additional funding ranging from \$430,000 to \$1 million in 2014, as seed money to redefine and strengthen the principal supervisor position. As of spring 2015, the districts were still addressing fundamental questions about supervisors' capacity to carry out this changed role, particularly in working with novice principals who might not know how to ask questions and get what they need from their supervisors.

Reducing supervisors' span of control so that they could effectively support principals' continuous improvement was one of the district priorities everywhere except in New York City, where the large number of principals made a major reduction in supervisors' caseload impractical. Denver, Gwinnett County, and Prince George's County had added supervisors before 2014 in order to reduce each supervisor's caseload. For 2014-15, seed money from The Wallace Foundation supported the addition of more supervisors in Charlotte-Mecklenburg and Gwinnett County and more assistant supervisors, in an enlarged role, in Hillsborough County. Reflecting on a small caseload, a principal supervisor in one district described what she was able to accomplish with a manageable number of principals:

With my nine schools, I probably talk with every one of my nine principals several times a week and I'm in every one of my schools weekly if not every 10 days. When I sit down to do the summary evaluations, I can go through every one of those standards and I can tell you what every one of my principals is doing and where they're strong and where their growth areas are. I think that's pivotal. ... Everybody needs feedback. Everybody wants to continue to grow and get better as a leader and so I think that we're beginning to better understand that.

District leaders knew they needed to define the principal supervisor role differently than in the past, with a focus on principal success. This was not a new idea in 2015. "We've been shifting the role of principal supervisor for three years now and we're in our fourth year of it," as one administrator put it. Some district leaders, but not all, said they felt they were clear about the new role definition. A focus on improving instruction was the key, according to a New York district leader, but that leader also acknowledged that many sitting supervisors had been hired years ago to focus on compliance-oriented tasks and community engagement: "...and so their original orientation to the work wasn't to be the instructional supervisor of a bunch of people. ... So it's like shifting the mindset." Gwinnett County leaders also appeared confident that they were communicating the right expectations by emphasizing that supervisors should work with principals in three big areas: instructional, operational, and civic engagement. They said that, as a result, the supervisors' work was starting to change: "... the visits that are made to our schools are more structured, they are more focused. They are much deeper than they were three years ago."

A district leader in Denver expressed doubt that the work of redefining the role was easily completed, however. This leader said that unless the district also defined what it meant to be an effective principal supervisor, supervisors' performance would be uneven:

So defining the effectiveness of a principal supervisor and having a framework to say "This is what an effective supervisor does and this is what it looks like" [is essential]. So role definition and then how do you know if you're effective? I mean, imagine doing such hard work and not knowing those two things? Those are huge.

Districts provided supervisors with training to develop their coaching and feedback skills, with Wallace support. District leaders worried that without this training, supervisors would struggle to have the difficult conversations about performance that might help principals improve. Universities helped two districts. In Charlotte-Mecklenburg, Queens University provided some training in coaching to principal supervisors. A

supervisor observed that this was extremely helpful because most supervisors did not have coaching experience and coaching is “a huge skill set” to develop. He also described informally training with other supervisors in coaching, visiting schools with other supervisors and discussing what they saw and what they would have coached around. “That’s been a tough process, but a good one,” he said. Hillsborough County provided feedback training to supervisors with assistance from the University of Washington Center for Educational Leadership. In the training, supervisors did school walkthroughs and then practiced the feedback they would give to a principal based on what they saw. The district also provided in-house training on how to provide ongoing feedback.

Executive coaches working in Denver were training some of their clients who were principal supervisors on delivering feedback to principals. Coaches described role playing and practicing delivering feedback to a principal, getting supervisors to think about what the feedback sounds and feels like and what are the main points they want to deliver. They also described helping supervisors develop their questioning skills rather than their directing skills, focusing the supervisors on asking open-ended, probing questions intended to unlock their principals’ thought processes and shed light on why they were doing what they were doing.

For some supervisors, shifting into the role of principal supporter and coach was a challenge. Leaders in every district described the varied pathways their supervisors were taking toward becoming instructional leaders, and said that the shift away from focusing on operations management made the journey difficult for some supervisors. A leader in one district said that supervisors who had “a great skill set for operations” were working to improve their instructional leadership, too, but so far only some had mastered “the whole overall leadership skill set.” Another district administrator described supervisors’ previous job as “firefighters” in the district: “All you would ask them about was ‘How is the facility? Is the school safe?’ You would not have asked them about instruction or about teacher performance, and now that’s shifted.” A supervisor echoed these observations, explaining that despite the high-quality training he was receiving, the climb to becoming an instructional leader was steep:

I don’t think there’s any way unless the roles change further that I would be able to go in to high-performing principals’ schools and know more about what good instruction looks like than they do, but I think I’m getting to the level where I know what it looks like and would be able to make suggestions.

In another district, administrators were of two minds about the capacity of their supervisors to support principals. Some expressed full confidence, describing the supervisors as “very successful [former] principals” who were strong building managers and instructional leaders and had the capacity to lead principals. Monthly supervisor meetings were cited as an example of the district’s efforts to ensure supervisors were well-prepared to support principals by giving them an opportunity to share ideas: “We have a think tank. We talk about what the common issues are and devise plans of support.” Another administrator, however, wondered whether supervisors were fully aware of how to coach and provide feedback to principals:

Our [supervisors] are still young in their leadership in this capacity, and again, it’s about teaching adults. They are not the principal anymore. They are the supporter of principals and that’s a different lens and we are still learning what that looks like.

Principals in several districts described a sea change in the frequency and quality of supervisor visits to their schools. Having their supervisors in their schools more regularly was a welcome and, in their view, effective policy change. Simply making visits more frequent was one important change. “When I had a supervisor that had 47 schools, I did not see him once in my school building, never, the entire year,” a principal said in a focus group. In another district, a principal said her supervisor in 2014-15 was in her school “a lot.” For this principal, having someone “pop in to see what’s going on” was a source of comfort: “I mean, it’s just good to know that somebody knows you’re still out there, because for years we were on islands by ourselves.” A principal in still another district described the supervisor visits in 2014-15 as having an entirely different shape, rhythm, and intensity than in past years:

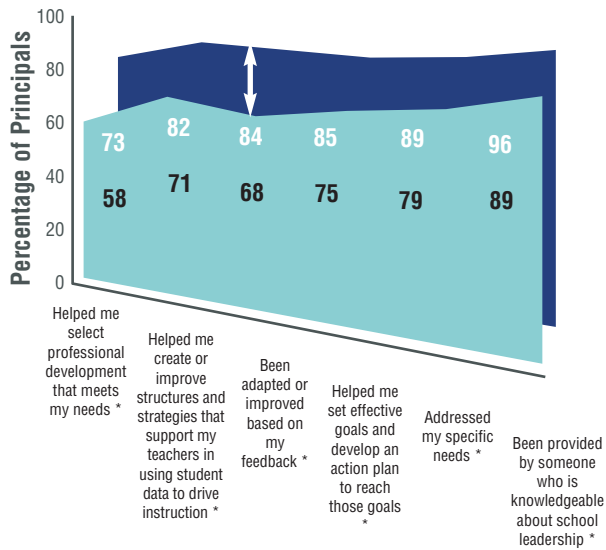
...this year, my [supervisor] is seeing me in the midst of the work where he’s attending our leadership team meetings, he’s at our faculty meetings, he’s in our grade-level planning meetings and sharing ways that I might [do things differently], and just really helped focus on what’s the data, what’s our story, what are we doing well and what are we not doing so well and how can we refine those practices.

Principal survey data for 2015 suggest that, while coaches were still the preferred source of support as they had been in the previous year, supervisors were starting to close the gap (Exhibit 16). In 2013-14, there was a statistically significant difference between principal perceptions of the support they received from their mentor/ coach and perceptions of the support they received from their supervisor/ evaluator. By 2014-15, the differences had narrowed so much that they were no longer statistically significant. For example, in 2013-14, 73 percent of novice principals reported that their mentor/coach had helped them select professional development that met their needs compared with 58 percent who reported that their supervisors supported them in this way; a difference of 15 percentage points. By 2014-15, the difference had dropped to six percentage points. Similarly, in 2013-14, 16 percentage points separated principals’ perceptions of their mentor/coach versus their supervisor/evaluator with respect to each provider’s capacity to adapt their support based on a principal’s feedback; by 2014-15, the difference had decreased by half to eight percentage points.

Exhibit 16: Difference in principal perceptions of the support they received from their supervisor/evaluator and mentor/coach in 2013-14 versus 2014-15

2013-14 Principal Perceptions

Size of difference in principal perceptions of supervisor/evaluator versus mentor/coach support

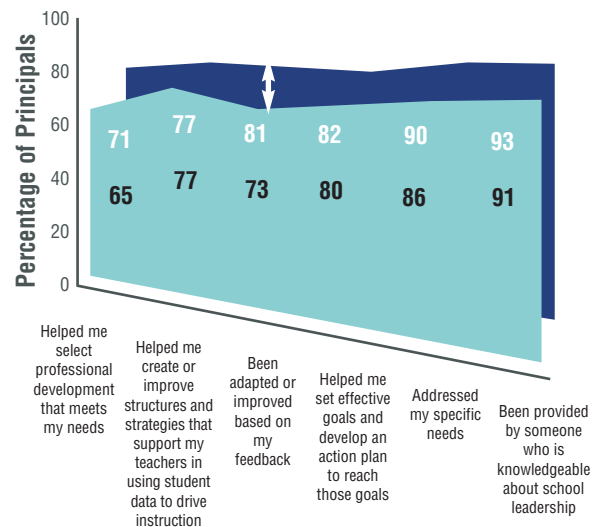


- Support received from Supervisor/Evaluator in 2013-14 [N(w)=523]
- Support received from Mentor/Coach in 2013-14 [N(w)=381]

* The difference between supervisor/evaluator and mentor/coach is statistically significant

2014-15 Principal Perceptions

Size of difference in principal perceptions of supervisor/evaluator versus mentor/coach support



- Support received from Supervisor/Evaluator in 2014-15 [N(w)=504]
- Support received from Mentor/Coach in 2014-15 [N(w)=345]

Exhibit reads: In 2013-14, 73 percent of novice principals who agreed that their mentor/coach had helped them select professional development that meets their needs compared with 58 percent who agreed that their supervisors supported them in this way, which was a difference of 15 percentage points. In 2014-15, the difference between principals who agreed that their mentor/coach versus their supervisor/evaluator helped them select professional development that met their needs had dropped to six percentage points.

Still, principals perceived some issues with their supervisors' support. Some principals described experiences that would suggest that supervisor support was inconsistent, including supervisors who were difficult to work with, were unwilling or lacked the capacity to help a principal address a need, or simply did not show up in the school often enough to help the principal develop professionally. In one focus group, for example, while one principal described seeing her supervisor regularly, another principal said his supervisor "checked out two months ago. I've got nothing." Principals commented that their supervisors were stretched too thin, whether because of a heavy caseload or their central-office obligations. The latter concern came up in each district. As one principal put it: "...if their job is to be instructional coaches to us as school leaders and 35 percent of their time is taken up by meetings in this [district headquarters] building, how are they coaching us and how are they helping us?" And some principals raised concerns about supervisors' capacity as instructional leaders, saying, for example:

...if someone comes to me and they don't know more than me, it's really hard for me to buy into [them]... You want someone to come to the table that's going to stretch you and help you fill in those gaps and give you logical things to think about. But if that person who's brought in to support you doesn't have that capability, then that's really useless.

Combining the role of supervisor with the role of providing support appeared to be working out successfully, according to most principals' reports in the survey. However, issues of trust did come up in focus groups. Some principals were unsure about the extent to which they could trust their supervisors with information about their needs. Principals in two districts described keeping their supervisors at arm's length, feeling unsure of how candid they could be without suffering negative consequences. As one said:

Well, you don't ask certain questions of your boss. You know, it depends on your rapport and the relationship, of course and the interpersonal dynamics there. ... [But] you've got to be very careful. So I don't know that [my supervisor is] the first person I'm going to when I have to ask a sensitive question.

A district leader acknowledged that the supervisor/principal relationship was not yet what it needed to be for all novice principals: "In some cases, [the principals] don't know where to push, where to not, where to ask questions, and how to be an advocate for their school."

Mentor and Coach Support

In each of the six districts, mentors or coaches offered support to novice principals.¹⁰ In 2014-15 across the six districts, 90 percent of first-year principals, 74 percent of second-year principals, and 38 percent of third-year principals had a mentor or coach. The figures had not changed significantly since 2012-13. The overall rate of mentoring or coaching reported by novice principals in these districts, 69 percent in 2014-15,

¹⁰ "Mentoring" and "coaching" did not have consistent definitions across districts. In this report we use "mentors or coaches" and "mentoring or coaching" when presenting cross-district findings but use the local term when analyzing the supports in a particular district.

was somewhat higher than the rate of mentoring or coaching reported by all principals in urban districts in 2011-12 on the federal Schools and Staffing Survey, which was 62 percent (Manna, 2015). Districts varied in the type of support offered, along a continuum from helping principals learn to think for themselves to assisting them in the day-to-day management of their schools.

Mentors and coaches saw themselves as trouble-shooters, helping principals address the day-to-day problems that arose. A coach in a pilot initiative in New York City explained that they were trying to identify the needs—which were extremely varied—and then respond, rather than bring a set agenda. Consequently, they worked through repeated cycles of learning and adaptation:

So it's been—for me the whole experience has been like learning how to navigate the differences and then trying to respond to them in an appropriate way. That's been the challenge, a good challenge that I think we're still working well with.

A mentor in another district described helping principals in whatever way made sense: “they call or email, and if needed, I go over to their buildings and we role play. And if I don't have the resource or the support to give them, I'm like, okay, this is who we need to get.” In Denver, which used a model of executive coaching, a coach explained:

...the whole idea of executive coaching is really what your client needs; and they articulate those needs. We're not here, as coaches, to tell [principals] what they need. They have [principal supervisors] to tell them what they need.

Coaches said they were often the bridge between principal and supervisor, helping principals understand their supervisors' expectations and helping the supervisors understand their principals' needs. Districts had set up procedures for these interactions. In Gwinnett County, for example, mentors took responsibility for ensuring that a supervisor knew about a principal's needs. A supervisor described how the process worked:

I would meet mostly with the mentor and talk about concerns and the principal knew it. When I would visit, [the mentor] would say she knew about it. Our conversations [with the principal] centered around the same topic/ concern if there was a concern. For every new principal, I'd know who the [mentor] was. The [mentor] would schedule an appointment with me and ask if I had any concerns. We'd kind of work like a tag team. It could be informal, but there is at least monthly communication.

And a Gwinnett County mentor explained:

To me at least, it typically has been, if [the supervisor is] seeing something, they'll let me know they're seeing something. That gives me a heads up to have something I can go in and work on. However, if I see a red flag, and I start getting very concerned, I will camp on the [supervisor's] doorstep.

A joint effort by district staff in different roles was also the expectation in Prince George's County. A coach, mentor, and principal supervisor would meet at the beginning of the year to discuss the roles they would play in providing principal support. A district administrator explained that through continuous dialogue with the principal's supervisor and mentor, the coach—having had recent experience in the principalship and “understand[ing] the pressures and nuances of the first year”—could accurately diagnose and effectively communicate a principal's needs to their supervisor and mentor. “So it's just been a nice relationship and communication among the ‘triad.’”

There were limits to the information sharing, however, at least in some circumstances. Mentors and coaches described limits on the amount of information they would share with supervisors. A mentor said:

We want to continue to cultivate the relationships [with the supervisors] so that we can both come to a good understanding of what are the priorities ... and focus areas of the school. ... But we are definitely not spies, and we are definitely not running back [to the supervisors] with information.

Principals' comments in focus groups painted a picture of the coach or mentor as their lifeline, providing day-to-day, hands-on support that principals said was vital to their immediate survival but that their supervisors usually did not have the bandwidth to provide. Principals described their mentors as “cheerleaders,” “someone to bounce ideas off of,” “available at all times,” “an amazing resource,” “the most effective source of support,” and a “lifesaver.” One principal's description reflected the perspectives expressed by many principals across the six districts:

Having a [mentor] has probably been the most effective [support]. And I know for my mentor, it's a 24-hour-a-day job. I mean, I've had to have those middle of the night conversations with my mentor about things that keep me up at night, and she's just been very open and very wise and given great advice. So that has probably been the number one [thing] for me as a new principal is having that [mentor].

Aware of the sheer volume of their needs, however, novice principals were making distinctions between the support they could ask of their mentor or coach—a ready resource—versus that which they could ask of their supervisor: “because you don't want to call your [supervisor] for every little thing, especially as a first- or second-year principal,” as one said.

Having the support of a mentor or coach appeared to raise principals' opinions of the effectiveness of all the support they received from their district. Principals who had a mentor or coach were more likely to rate the support they received from their supervisor and from professional development higher than principals who did not have a mentor or coach in 2015. While the differences in the average ratings of support were not large—means of 4.2 versus 3.8 for supervisor support and 4.0 versus 3.8 for professional development—they were statistically significant. Indeed, the differences suggest that districts' investments in mentors or coaches extended the benefits of all the support principals received.

Most principals did not report discussing the results of their evaluation with their mentors or coaches, however. Only about a third of the principals who had mentors or coaches reported on the survey that they had met with them to discuss the results of their evaluation. It is possible that principals made distinctions between their day-to-day needs—served largely by their mentors or coaches—and the needs documented through a formal evaluation system.

Professional Development

The Principal Pipeline Initiative called on districts to support novice principals through high-quality professional development, but it appears that districts were focusing more attention on strengthening individual support from supervisors, mentors, or coaches than on strengthening group-based professional development. Districts offered professional development for principals, but the connection between identified principal needs and available professional development was a work in progress.

District-wide data on principal needs played a part in planning district programs of professional development. In Charlotte-Mecklenburg, for example, district leaders identified managerial leadership as a common gap in principal practice and therefore offered professional development sessions around helping principals “manage all the different responsibilities and their nuances.” Similarly, leaders in Prince George’s County reported designing professional development offerings based on principal evaluation results: “... that’s how we make our big buckets for what we’ll do throughout the year. We kind of take this collective look with what our principals are struggling with across the board, and think about what maybe new processes that we’re going to put in place.”

Principals did not rate professional development as a significant source of support, however. It lagged behind supervisor and mentor or coach support, according to principals surveyed in spring 2015 (Exhibit 17). The percent of principals who strongly agreed that professional development addressed the pressing issues in their school, addressed their specific needs, or deepened their understanding of school leadership was consistently lower—by as many as 20 percentage points—than the percent strongly agreeing that their supervisor, mentor, or coach supported them on these issues.

In focus groups, principals described professional development as “compliance-oriented” and organized for large groups. They often described the professional development in their district as overloaded with information. No principal referred to district-sponsored professional development as helpful or focused on addressing individual needs.

In addition to traditional course offerings, district leaders in Charlotte-Mecklenburg, Denver, Gwinnett County, and Prince George’s County described novice principals’ professional learning communities (PLCs) as another source of professional development and support. These PLCs met regularly to discuss problems of practice and receive support from district staff—such as instructional facilitators, assessment specialists, or finance specialists—to develop or strengthen a skill. District leaders said that these groups provided real and meaningful support for principals; a Denver official said of principals: “... they’ll actually pick up the phone and call ... somebody that they know is really good at something that they’re struggling with. They’ll do co-visits to schools. I know a couple principals that talk to three or four other principals five or six times a week.” Group walkthroughs were also a feature of novice principal PLCs in Charlotte-Mecklenburg, where a principal said that six principals from the PLC would visit one school and help the principal identify their “blind spots” so that they could think about how they would improve their leadership.

Exhibit 17: Principals who received support from their supervisor/evaluator, mentor/coach, or through professional development who STRONGLY AGREE that the support helped them in various ways

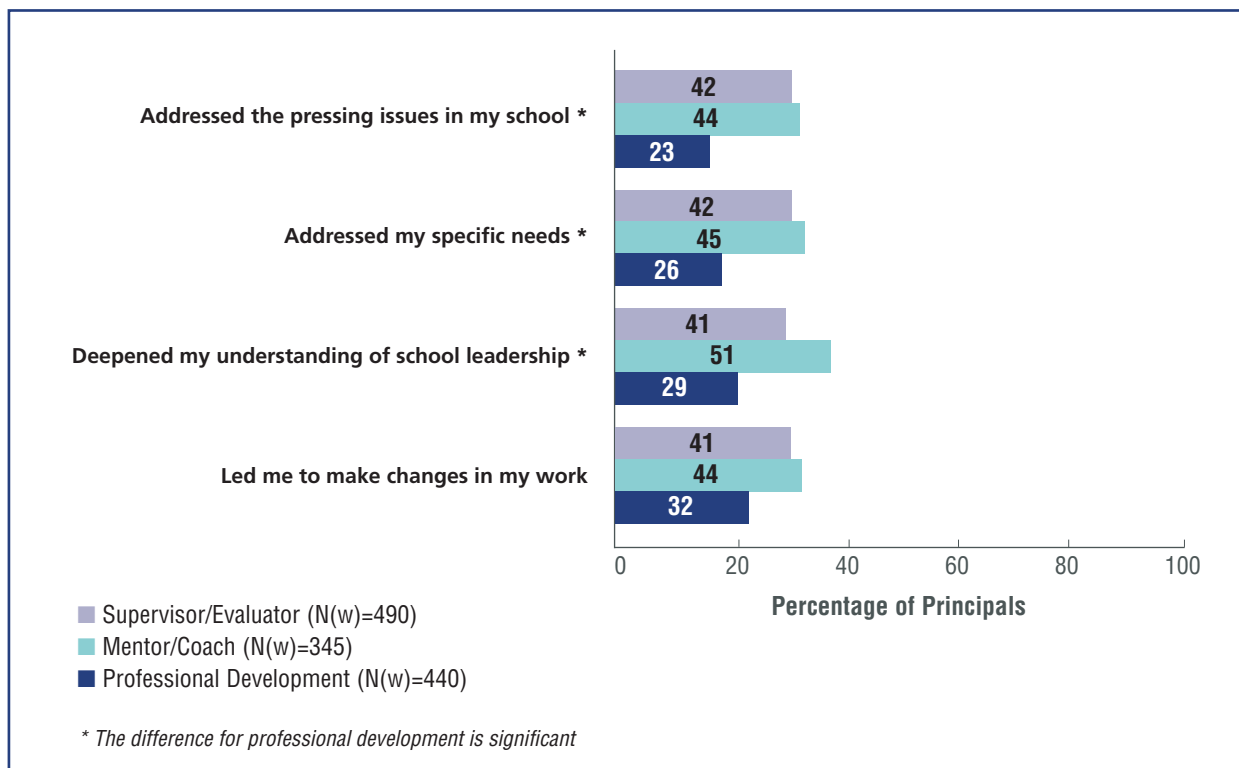


Exhibit reads: Forty-two percent of principals agreed that the support they received from their supervisor/evaluator helped them address the pressing issues in their school to a great extent.

Denver administrators said they recognized that their new principals who had come through strong preservice preparation had already outgrown the “nuts and bolts” induction training that the district had typically provided. The district quickly developed a PLC structure intended to give novice principals the opportunity to discuss their successes and challenges with a network of peers. District administrators said the resulting New Principal Breakfast had become a focus group for principals to exchange ideas and strategies and for central office staff to identify—and serve—the very different real-time needs of their novice staff:

Our intent was to just pull them together and give them the support, the networking support that they gave us feedback that they wanted. And I would say over the course of the year, that has definitely turned into—for me, it’s been a focus group where I get their feedback of how their year has gone, and so I’ve really spent a year with them and I have some really good notes on what they do need moving forward.

Monthly PLC meetings in Prince George’s County, similarly, provided professional development to principals in such areas as curriculum, instruction, teacher evaluation, and data-based decision making.

Districts were still working to strengthen the connection between individual principal needs and available professional development. After several unsuccessful attempts in recent years, for example, a district leader in Denver described the complexities of building a professional development system that would directly respond to principal needs: “That’s just challenging, to take a principal out of their building, when so much is going on, to offer PD that is just-in-time, that’s differentiated, that is really meaningful, and actually moves the growth of that leader.” One problem that administrators identified was that professional development was usually not available on the right timeline. A mid-year evaluation in January might call for support that was only offered in December. The district leader explained that the district was working to develop professional development opportunities aligned with the evaluation rubric and to offer them at multiple times during the year. The district was also trying to make supervisors more aware of professional development opportunities through an online, digital professional development course catalog. However, the district held no illusions about the extent to which principals saw these courses as go-to sources of support:

...we did a lot of work, published this document, and then the second it’s published, it’s no longer accurate. So we lose the confidence of [our principals] who say “I need help and that’s not an option for me because I’m not going to waste my time even trying to do that.” And so I think that we’ll have to do an intentional push to say “We stand by this; this is going to be helpful to you.” But we’ll definitely have to build confidence in our users ... that we can meet their needs.

Traditional notions of professional development through course offerings were not enough to help principals improve, according to some district leaders. A Hillsborough leader explained that course offerings did not provide the job-embedded experiences or practice that principals needed. For example, principals needed to develop their capacity to have hard conversations with teachers based on classroom observations, and this leader then thought:

What I want to give them is, what would that look like? So what could be good examples of job-embedded experiences that this person would get to improve in that area? We haven't done that yet. So all [supervisors] think about is "What PD can I give them?" They're not at the point yet, about what can I do to support them in the field. So I think that's a tool that, that needs to be provided to them, this summer, as we go further.

Summary

The Principal Pipeline Initiative emphasized on-the-job support for novice principals, and the six districts made this a priority. A high proportion of principals reported receiving support in the areas in which they needed to improve, and district leaders were working hard to build capacity for better support. Principal supervisors, some newly hired and others doing a job that had been redefined, were striving to support their principals as instructional leaders. By 2015, principals were generally rating their supervisors as a source of knowledgeable, tailored help almost comparable to their mentors or coaches. This reflected growth in the supervisors' ratings since the previous year. Focus groups with principals revealed, however, that some principal-supervisor relationships were not living up to district expectations, whether because of a lack of supervisor capacity or because of the principal's concerns about candidly discussing his or her weaknesses. The on-call, nonjudgmental role of a mentor or coach continued to offer benefits that principals said they appreciated.

District professional development did not, as yet, offer effectively tailored help to address individual principals' identified needs. Districts were trying to improve their professional development by aligning it to overall data on all principals' learning needs and by instituting professional learning communities. However, principals expressed limited satisfaction with what they received. Thus the individual capabilities of a principal's supervisor and mentor or coach (if one was available) were the chief resources offered by districts for principals' professional learning.

4. Summary and Conclusions

As of 2015, every district had developed rigorous systems to evaluate principal performance, designing complex, multi-measure systems that included measures of professional practice and student growth to evaluate principal performance. These systems reflected state policies (under the federal Race to the Top and waivers of the Elementary and Secondary Education Act accountability provisions). Districts began piloting and rolling out their evaluation systems early in the Principal Pipeline Initiative, if not before, and they continually tinkered with the measures, the procedures, and the training for evaluators.

Across districts, the principal evaluation measures varied in number and type and were weighted differently, but stakeholders, including district leaders, principal supervisors, and novice principals, generally agreed that the evaluation system was appropriate and captured information that would ultimately help principals improve. Unlike most teacher evaluation systems that mandate a predetermined number of observations, these principal evaluation systems fostered an ongoing conversation between principal and supervisor intended to produce steady and continuous improvement in principal practice. Principal evaluation was not a scheduled, discrete activity but a year-long conversation between principal and supervisor.

While districts had made a great deal of progress in developing rigorous principal evaluation systems, they were still working to strengthen them. They knew they needed to continue to spell out the evidence that would support accurate ratings of professional practice and, especially, to train principal supervisors for greater consistency in ratings.

Top leaders in these districts were committed to the Principal Pipeline Initiative idea of providing support to each novice principal that took evaluation results into account. They were redefining the role of the principal supervisor to become the principal's primary source of support and professional development. The districts provided additional staffing and training to build principal supervisors' capacity to support principal growth. In spring 2015, districts had made many strides toward preparing supervisors for their new roles. This was still a work in progress, but most novice principals saw their supervisors as knowledgeable and helpful, according to early evidence.

Although the districts also reported efforts to reorganize their professional development offerings to address principals' individual learning needs, they had been less successful in this regard, based on survey and interview data from novice principals. Group-based professional development was rated lower than principal supervisors, coaches, or mentors as a resource for addressing individual needs or deepening understanding of principal leadership. Mentors and coaches were given high ratings for their help and support by the novice principals with whom they worked.

Districts did not yet have clear plans for using evaluation data to inform improvement of other pipeline components. The results of principal evaluations were not yet informing principal hiring and placement practices or preservice program content and quality. This may be because the evaluation systems were so new. While some district leaders recognized that the evaluation data could and should be factored into dis-

trict decision making with regard to principal training, hiring, and placement, none had used the evaluation data in this way nor had district leaders universally expressed this as an objective. They did, however, hope that systems of evaluation and support could help them reduce the number of principal dismissals by identifying principal needs early and providing them with support that would help them improve.

Overall, these districts saw their principal evaluation system as a way to direct attention to the competencies that they valued in school leaders and to direct individual principals' attention to the specific competencies that they needed to develop further. Still, it takes time to refine evaluation system design, define evidence, ensure consistency in ratings, and build comprehensive support for principal learning, and districts were still working to achieve these goals.

References

- Clifford, M., & Ross, S. (2012). *Rethinking principal evaluation: A new paradigm informed by research and practice*. Alexandria: National Association of Elementary School Principals and National Association of Secondary School Principals.
- Corcoran, A., Casserly, M., Price-Baugh, R., Walston, D., Hall, R., & Simon, C. (2013, October). *Rethinking leadership: The changing role of principal supervisors*. Washington DC: Council of the Great City Schools.
- Davis, S., Kearney, K., Sanders, N., Thomas, C., & Leon, R. (2011). *The policies and practices of principal evaluation: A review of the literature*. San Francisco, CA: WestEd.
- Goldring, E., Cravens, X.C., Murphy, J., Elliott, S.N., Carson, B., & Porter, A.C. *The evaluation of principals: What and how do states and districts assess leadership?* Paper presented at the American Educational Research Association Conference, New York, NY, March 2008.
- Harris, D.N., & Herrington, C.D. (2015, March). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, 44, 2, 71-76.
- Hornig, E., Kalogrides, D., & Loeb, S. (2009). *Principal preferences and the unequal distribution of principals across schools* (CALDER Working Paper No. 36). Washington, DC: Urban Institute.
- Kane, T.J., McCaffrey, D.F., Miller, T., & Staiger, D.O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle: Bill & Melinda Gates Foundation.
- Louis, K.S., Leithwood, K., Wahlstrom, K.L. & Anderson, S.E. (2010). *Learning from leadership: Investigating the links to improved student learning*. St. Paul, MN: University of Minnesota, Center for Applied Research and Educational Improvement.
- Lunenburg, F.C. (2011). Goal-setting theory of motivation. *International Journal of Management, Business, and Administration*, 15, 1-6.
- Manna, P. (2015). *Developing excellent school principals to advance teaching and learning: Considerations for state policy*. New York: The Wallace Foundation
- Mitgang, L., Cummins, H. J., & Gill, J. (2013). *Districts matter: Cultivating the principals urban schools need*. New York, NY: The Wallace Foundation

National Research Council. (2009, October 5). *Letter report to the U.S. Department of Education on the Race to the Top Fund*. Washington DC: Author.

Portin, B.S., Feldman, S., & Knapp, M. S. (2006). *Purposes, Uses, and Practices of Leadership Assessment in Education*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Sartain, L., Stoelinga, S.R., & Brown, E.R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago: Consortium on Chicago School Research.

Turnbull, B.J., Riley, D.L., Arcaira, E.R., Anderson, L.M., & MacFarlane, J.R. (2013). *Building a stronger principalship, Vol. 1: Six districts begin the Principal Pipeline Initiative*. Washington DC: Policy Studies Associates.

Turnbull, B.J., Riley, D.L., & MacFarlane, J.R. (2013). *Building a stronger principalship, Vol. 2: Cultivating talent through a principal pipeline*. Washington DC: Policy Studies Associates.

Turnbull, B.J., Riley, D.L., & MacFarlane, J.R. (2015). *Building a stronger principalship, Vol. 3: Districts taking charge of the principal pipeline*. Washington DC: Policy Studies Associates.

U.S. Government Accountability Office (2013, September). *Race to the Top: States implementing teacher and principal evaluation systems despite challenges*. Washington DC: Author.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York: The New Teacher Project.

Appendix A: Survey Administration, Weighting, and Analyses

Survey Administration

In the winter of 2013, 2014, and 2015, PSA contacted the PPI project directors in each of the six districts to explain the purpose of the survey and to request the names and email addresses of all first-, second-, and third-year principals. Once PSA received the principal rosters, we reviewed the lists for accuracy. Web-based surveys were administered in spring 2013, spring 2014, and spring 2015 to all first-, second-, and third-year principals in the six districts. We sent multiple reminder emails to non-responding principals. In the 2013 wave of surveys, 353 principals responded out of the total of 617 novice principals surveyed across the six districts, for an overall response rate of 57 percent. A total of 541 of the 821 novice principals responded in 2014, for an overall response rate of 66 percent, and a total of 514 of the 791 novice principals responded in 2015, for an overall response rate of 65 percent.

Survey Weighting and Analyses

To determine post-stratification survey weights for each district, the inverse of the number of respondents from the district out of the total number of respondents to the survey was divided by six, the total number of districts (Exhibit A-1). For example, in 2015, there were 56 principal respondents in Charlotte-Mecklenburg and 514 total principal respondents from all six districts, resulting in a survey weight of 1.53 (i.e., $(514/56)/6$). The 2013 and 2014 data from principals were weighted in a similar fashion.

Exhibit A-1: Principal survey weights by district and year

District	2013	2014	2015
Charlotte-Mecklenburg	1.37	1.73	1.53
Denver	1.05	1.33	1.56
Gwinnett County	1.51	1.92	2.09
Hillsborough County	1.07	2.15	1.86
New York City	0.58	0.34	0.35
Prince George's County	1.01	1.43	1.26

Exhibit reads: The survey weight for Charlotte-Mecklenburg principals in 2013 was 1.37; the survey weight for Charlotte-Mecklenburg principals in 2014 was 1.73; the survey weight for Charlotte-Mecklenburg principals in 2015 was 1.53.

The main method of analysis for this report was to analyze cross-sectional data from 2013 through 2015. The 2013 data included principals who started in their position from 2010 to 2013, the 2014 data included principals who started in their position from 2011 to 2014, and the 2015 data included principals who started in their position from 2012 to 2015.

Many comparisons in the report are presented along with information about whether the difference between estimated percentages is statistically significant. One method of determining the statistical significance of the difference between two percentages is to compare the confidence intervals of the two percentages. Confidence intervals provide information about the accuracy of the estimated percentages. If the confidence intervals for two percentages do not overlap, then the difference is deemed statistically significant.

POLICY STUDIES ASSOCIATES, INC.

1718 Connecticut Avenue NW

Suite 400

Washington, DC 20009

Commissioned by:



The Wallace Foundation®

5 Penn Plaza, 7th Floor

New York, NY 10001

January 2016