Westat

Improving lives through research™

Evaluation of the Illinois Performance Evaluation Reform Act: Final Report

Authors

Anthony Milanowski
Liam Ristow
Matthew Finster
Clarissa McKithen
Westat

Mimi Doll
Holly Lewandowski
Denise Roseland
Candeo Consulting

Bradford R. White
Illinois Education Research Council

Bradley Carl
Nandita Gawade
Hannah Kang
Daniel Marlin
Robert Meyer
Stephen Ponisciak
Clarissa Steele
Yan Wang
Wisconsin Center for Education Research

June 30, 2016

Prepared for:
Illinois State Board of Education
100 N. 1st Street
Springfield, IL 62777

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Contents

Table

Figure

Figure		Page

Executive Summary

In 2010, the Illinois Legislature passed and the Governor signed the Performance Evaluation Reform Act (PERA), which requires that all Illinois school districts establish new educator evaluation systems, ensure evaluators undergo rigorous training and obtain certification, and develop professional development or remediation plans for ineffective teachers. PERA mandated that districts implement new principal evaluation systems by the beginning of the 2012–13 school year. New teacher evaluation systems will be phased in between the 2012–13 and 2016–17 school years, at which time measures of student growth must be incorporated in evaluation ratings for all teachers.

PERA requires a research-based study of the implementation of new educator evaluation systems by Illinois school districts, and to meet this requirement, the Illinois State Board of Education (ISBE) contracted with Westat and its partners to conduct this study. An interim report, focusing on the implementation of PERA-compliant evaluation systems for the 2013–14 school year, was submitted in September 2015. The current report covers implementation during the 2014–15 school year, and addresses these questions:

- *How can PERA-compliant evaluation systems be implemented effectively and with fidelity?*
- *What lessons can we learn from principal evaluation that may be useful for teacher evaluations?*
- *How do districts fit measuring student achievement growth within their current system?*
- *What is changing in professional development and how is it affecting teachers?*

To address the questions, the study team administered a short survey to district administrators; conducted telephone or in-person interviews with knowledgeable district staff to collect more in-depth information on implementation, assessment, and professional development; conducted in-depth case studies of five districts that included interviews with teachers, principals, and district administrators; and analyzed personnel and student achievement data provide by ISBE.

The main findings of the study with respect to the broad research questions are:

- **How can PERA-compliant evaluation systems be implemented effectively and with fidelity?**
  - Joint committees that met regularly, had good communication and high teacher involvement, and were representative of all grade levels were most successful.

- Joint committees can take an active role in monitoring and streamlining the evaluation process; districts can establish other committees to address specific issues; districts can appoint a committee, individual, or set of individuals with specific responsibility for overseeing implementation.

- Districts should monitor their systems for compliance and seek feedback on the usefulness of their approach from teachers and principals.

- Prior use of the Danielson Framework for Teaching facilitated PERA implementation in some districts.

- The phase-in of teacher evaluation allowed districts to move forward without causing massive disruption or resistance and allowed joint committees to assess their initial decisions and modify systems to respond to implementation glitches.

- An analysis of state personnel data found no evidence that PERA implementation was followed by an increase in the attrition rate for principals.

- **What lessons can we learn from principal evaluation that may be useful for teacher evaluations?**

  - Some districts used the addition of student achievement growth measures to principals' evaluations to begin to address assessment and data collection issues that would later arise for teacher evaluation.

  - However, many districts have concentrated more on teacher than principal evaluation.

- **How do districts fit measuring student achievement growth within their current system?**

  - Many districts depend on commercially available assessments for meeting the requirement to measure student growth using Type I or II assessments.

  - Finding the time to develop assessments for grades/subjects without predeveloped commercially available tests was a challenge for most districts.

  - Most districts found they needed substantial additional training for staff on assessment literacy and development in order to implement measuring student growth with Type III assessments.

  - Most districts chose relatively simple methods to assess growth that do not require extensive technical expertise, are relatively easy to interpret, and do not demand extensive data systems.

  - Even when using simple methods, many districts had to make investments in infrastructure or expertise to include student achievement growth in educator assessment.

  - Districts may need more support to develop roster verification systems and data warehouses to facilitate measuring student achievement growth for teacher evaluation.

- There was not a strong or systematic relationship between the 2014–15 average growth ratings and 2014–15 school-level value-added growth in reading or mathematics.

- **What is changing in professional development and how is it affecting teachers?**

  - PERA's emphasis on student achievement growth has resulted in districts providing more professional development on student assessment and goal setting.

  - Most districts are now using performance evaluation results to plan district-wide professional development.

  - When done well, evaluation conversations and discussions are focused on improvements in teaching practices, and this is linked to teachers' professional development.

  - It is still too early to assess whether professional development changes promoted by PERA have had a major impact on teachers.

Introduction 1

This report presents final findings from the evaluation of the Illinois Performance Evaluation Reform Act (PERA). PERA requires a research-based study of the implementation of new educator evaluation systems by Illinois school districts, and to meet this requirement, the Illinois State Board of Education (ISBE) contracted with Westat and its partners to conduct this study. The study began in January 2013, and an interim report was provided in September of 2015. The interim report focused on the implementation of PERA-compliant evaluation systems for the 2013–14 school year. That report primarily focused on implementation in Race to the Top (RTT) and School Improvement Grant (SIG) districts that were required to begin implementing PERA-compliant evaluation systems for that year.

Based on input from the Performance Evaluation Advisory Council (PEAC) and the Illinois State Board of Education (ISBE) staff, the study design was substantially modified from that used for the interim report. The revised design focused on a set of questions developed by PEAC and refined by ISBE staff and the study team. The emphasis was on identifying how districts could successfully implement PERA. The four overall guiding questions established were:

- *How can PERA-compliant evaluation systems be implemented effectively and with fidelity?*
- *What lessons can we learn from principal evaluation that may be useful for teacher evaluations?*
- *How do districts fit measuring student achievement growth within their current system?*
- *What is changing in professional development and how is it affecting teachers?*

To answer these questions, the research team used a mixed-methods research design that incorporated document reviews, a survey of district administrators, interviews with district staff, intensive case studies of five school districts, and analysis of data maintained by ISBE. The remainder of this report is divided into four chapters:

- **Chapter 2: Data Collection.** This brief chapter describes the types of data that inform the findings presented in this report.
- **Chapter 3: Implementation of PERA Teacher and Principal Evaluation Systems.** This chapter describes the features of PERA teacher and principal evaluation systems and addresses the first two main evaluation questions.

- **Chapter 4: Assessments and Student Achievement Growth.** This chapter discusses the types of assessments and assessment strategies being used in study districts and addresses the third main evaluation question.

- **Chapter 5: PERA and Professional Development.** This chapter addresses the ways in which study districts are using teacher and principal professional development within the context of PERA and addresses the fourth main evaluation question.

In addition to addressing the four main evaluation questions, Chapters 3, 4, and 5 also discuss findings related to specific evaluation subquestions that PEAC posed.

What Is PERA?

In 2010, the Illinois Legislature passed and the Governor signed the Performance Evaluation Reform Act[1] (PERA), which requires that all Illinois school districts establish new educator evaluation systems that evaluate educators on a 4-point scale and include multiple components, including professional practice and student growth, with the latter a "significant factor" in determining ratings. In addition, districts must ensure that evaluators undergo rigorous training and obtain certification and develop professional development or remediation plans for ineffective teachers.

PERA mandated that districts implement new principal evaluation systems by the beginning of the 2012–13 school year (SY). New teacher evaluation systems were to be phased in: All districts had to have incorporated certain new teacher evaluation components by the 2012–13 SY (such as a four-level rating system), but most did not have to incorporate student growth until the start of the 2016–17 SY. Chicago Public Schools were required to fully implement in the 2013–14 SY, the lowest performing 20 percent of Race to the Top (RTT) districts in the 2014–15 SY, and the remaining RTT districts and other districts in the lowest performing 20 percent in the 2015–16 SY. Districts with School Improvement Grants (SIG) were required to implement at various points according to their grant agreements.

PERA established that teacher evaluation systems complying with its provisions be designed at the district level through a joint committee composed of an equal number of representatives selected by the district and its teachers, or the local teacher association. Joint committees are also expected to be involved in ongoing evaluation of the system and making needed modifications.

Related legislation, Senate Bill 7,[2] provided for the use of evaluation results in teacher licensure. The State Superintendent must now consider two unsatisfactory ratings within a 7-year period as evidence of incompetency that could trigger license revocation. It also provided for performance evaluation ratings to take precedence over seniority within categories based on qualifications (e.g., certifications) during reductions in force. These provisions added to the importance of performance ratings in making personnel decisions about teachers.

In addition to requiring districts to implement educator evaluation systems, PERA also established a Performance Evaluation Advisory Committee (PEAC) of teachers, principals, superintendents, and other stakeholders to advise the Illinois State Board of Education (ISBE) on the development and implementation of the new evaluation systems. PEAC and ISBE have also developed and disseminated numerous resources to help districts develop and implement the systems. Based on consultation with the PEAC, ISBE promulgated rules for educator evaluation, including defining the weight for student achievement growth to be considered a "significant factor" as at least 25 percent for the first 2 years of implementation and 30 percent thereafter. ISBE and PEAC also developed and disseminated state models for teacher and administrator evaluation that districts may use. Where district joint committees do not agree on a local model, the state model must be used.

[1] Public Act 96-0861, 96th Illinois General Assembly. Available at: http://www.ilga.gov/legislation/publicacts/96/PDF/096-0861.pdf

[2] Public Act 097-008, 97th Illinois General Assembly. Available at: http://www.ilga.gov/legislation/publicacts/97/PDF/097-0008.pdf

Data Collection 2

To address the questions posed by PEAC, the study team and ISBE staff decided to concentrate on the following data collection and analysis strategies: 1) a short survey of district administrators, to collect basic information on how PERA was being implemented for the 2014–15 school year (SY); 2) focused telephone or in-person interviews with knowledgeable staff to collect more in-depth information on assessment and professional development; 3) in-depth case studies of five districts that included interviews with teachers, principals, and district administrators; and 4) analysis of personnel and student achievement data provided by ISBE. Each data collection or analysis method is described further below.

District Administrator Survey

Fifty-nine districts were invited to participate in the survey, including all 32 RTT districts and 27 additional districts that were reputed to be making good progress toward full PERA implementation. RTT districts were included because they were required to implement the PERA requirements earlier than other districts. Twenty-seven additional districts were identified based on nomination by PEAC members, ISBE staff, or others knowledgeable about PERA implementation. The survey was designed by the study team and reviewed by ISBE staff. The survey was administered over the Internet between late April and late June of 2015. Invitations were sent by email to the PERA coordinator of record for each district. Emailed invitations included a link to the survey and a personal identification number needed to access the survey. Forty-two districts responded to at least some of the questions, for an overall response rate of 71 percent. Thirty-eight districts responded to all sections, for a complete response rate of 64 percent.

Interviews and Case Studies

Two strands of qualitative data collection were conducted: a set of interviews with school district personnel in a sample of 11 districts on specific study topics and in-depth case studies of five additional school districts covering a wider range of study topics.

Focused District Interviews

Interviews on focused topics (implementation, assessment, professional development) were done in 11 districts. Districts were chosen to cover all of the state's regions and to try to represent different district sizes, as well as to include districts suggested by PEAC or ISBE as positive examples of implementation. To reduce burden and encourage cooperation, the number of staff interviewed was limited for some districts. Interviews also concentrated on fewer topics in some districts, with the topics of focus chosen based on information from PEAC or ISBE about districts that were innovating in these areas. These interviews were conducted onsite (where more than one person was interviewed) or by phone during the spring of 2015.

Table 1 lists the topic focus and staff interviewed in each of these districts.

Table 1. Staff interviewed and topic focus in focused interview districts

				Topic focus			Staff interviewed			
District	Region	Number of schools in district	Race to the Top district	Imple-mentation	Assess-ment	Profes-sional develop-ment	Teacher	Principal	District admin.	Union rep.
1	6	1–5	✓	✓	✓	✓	✓	✓	✓	✓
2	5	1–5	✓	✓	✓	✓	✓	✓	✓	
3	3	1–5		✓		✓			✓	
4	1	1–5		✓		✓			✓	
5	2	1–5	✓	✓	✓			✓		
6	1	1–5	✓	✓	✓	✓	✓	✓	✓	✓
7	4	6–10	✓	✓	✓	✓	✓	✓	✓	
8	1	6–10	✓	✓	✓	✓			✓	
9	1	11–20			✓	✓			✓	
10	1	1–5		✓	✓				✓	
11	3	1–5		✓	✓				✓	

In-Depth Case Studies

In addition to the focused interviews, more extensive interviewing was done in a smaller sample of districts. The intent of these case studies was to get a more in-depth picture of how districts were implementing PERA. Five districts agreed to participate in these interviews, including three RTT districts and two not participating in RTT. In each district, interviews were held with two or three teachers, two or three principals, one or two district administrators, and a union representative. Interview questions covered all of the study topics. The interviews were done between April and July of 2015.

Analysis of Interviews

For both the focused interviews and in-depth case studies, responses from interview participants were integrated into a summary document organized by the major study questions that described the responses in each district. If districts participating in interviews or case studies provided documentation about their systems, those documents were reviewed as part of the development of district summaries. In addition, interview transcripts from the five case study districts were analyzed to identify important themes and insights into the steps the districts have taken to implement PERA evaluation system components. Highlights from this analysis are presented throughout this report in call-out boxes that focus on the experiences of teachers, principals, and district administrators interviewed.

Analysis of Principal Attrition

One question the PEAC raised was whether PERA implementation had affected the principal pipeline. While analysis of the effect of PERA on entry into the principal role was not possible due to state licensure changes made at the same time, it was possible to explore the potential impact on principals' propensity to leave the role. The hypothesis was that if principals found that more rigorous evaluation of their performance was highly burdensome or threatening, there might be an increase in attrition after PERA implementation. Therefore, principal attrition rates before and after PERA were compared and principal attrition was compared to teacher attrition, which was not expected to increase as much since the full impact of PERA on teachers will not be felt until the 2016–17 SY. The results of this analysis appear in Chapter 3.

This analysis used two personnel datasets collected by ISBE, the Teacher Service Record (TSR) from 2005-06 through 2011–12 and the Educator Information System (EIS) from 2012–13 through 2014–15. There were over a million records (1,259,996) from the TSR and over half a million (589,423) from the EIS. For the TSR, principals were identified through their position code and the dataset was restricted to only principals who were active and working full-time in a given academic year. For the EIS, principals were identified via their position name and only principals who were working full-time during the regular academic year (i.e. not solely night school or summer school) were selected. For both datasets, duplicate records of the same principal working in multiple schools during the same year were identified and eliminated for ease of analysis.

Both the TSR and EIS contained unique identifiers that allowed individuals to be tracked over time and between the datasets. These identifiers were used to merge the EIS data with the TSR data and to create a series of variables for each individual to indicate whether he or she was employed as a principal in an Illinois public school during the given year. Attrition rates were calculated by measuring the proportion of all principals from one year who were not principals in the subsequent year. An identical approach was taken to measure annual teacher attrition.

Growth Analysis

One question PEAC posed asked about the relationship between growth ratings assigned to teachers through PERA evaluation systems and measures of student achievement growth. To examine this relationship, Illinois state test data was used to estimate value-added at the school level, and these estimates were compared with growth evaluation ratings collected from study districts. The value-added model and the comparison of value-added estimates to growth ratings are briefly described below. A more comprehensive summary of the value-added analysis is provided in a technical supplement.

The value-added analysis focused on student achievement growth between the 2014 Illinois Standards Achievement Test (ISAT) assessment and the 2015 Partnership for Assessment of Readiness for College and Careers (PARCC) assessment in mathematics and reading for grades 4 to 8. Conceptually, value-added analysis is the use of statistical techniques to isolate the component of measured student knowledge that is attributable to schools. The analysis attempts to remove the influence of other factors such as prior knowledge and student characteristics. In estimating student growth, measurable student characteristics were controlled for using available data on gender, race, economic disadvantage, homelessness, English-language learner status, disability, and mobility during the school year. Thus, the model used a large set of student characteristics to identify the extent to which schools contributed to the improvement of student achievement outcomes.

The value-added model also corrected for measurement error in the pretest because uncorrected measurement error can lead to biased estimates of school value-added. The measurement error correction was performed using estimates of pretest measurement error obtained from the 2014 ISAT technical report. Another feature of the model was the use of Empirical Bayes (EB) estimation, also referred to as shrinkage estimation, in order to correct for the overrepresentation of schools with small numbers of students at both the top and bottom of the distribution of measured performance. The EB estimator minimizes the error associated with a given school's estimate by taking into account its signal-to-noise

ratio and effectively 'shrinks' each school's value-added estimate towards the value-added of the average school, which is set at zero. The higher the likely noise (i.e., the less reliable the estimate), the more a school's value-added estimate is moved toward that of the average school. Estimates for schools with small numbers of students are typically more affected by this adjustment than those for schools with larger numbers.

As a final step, the results were examined to evaluate the performance of the value-added model. The 2014 ISAT assessment proved to be a good predictor of performance on the 2015 PARCC assessment. Results showed a very low correlation between average prior proficiency—a measure of average performance in the previous year—and value-added. There were also substantive positive correlations between mathematics and reading value-added within each school.

The results of the value-added analysis were compared to teacher and principal evaluation ratings from the 2014–15 school year that were obtained from study districts. Eighteen districts provided evaluation results, though not all included ratings based on student achievement growth. These results were used to develop average growth ratings at the school level, which were then these to school-level estimates of value-added. The findings from this correlational analysis appear in Chapter 4.

For some comparisons, reading and mathematics value-added estimates were combined. This was done to simplify presentation of results and because the student growth scores averaged across teachers within schools could cover multiple subjects. As noted in Chapter 4, the school average teacher or administrator student growth ratings may be based on ratings of teachers of a variety of subjects, and some are likely to influence both mathematics and reading achievement. Mathematics and reading value-added were combined by creating a score based on the first principal component of a principal components analysis of the reading and mathematics value-added data. The first principal component accounted for 77 percent of the variation in the data, which suggests that this score represents a substantial proportion of the total variation in school effectiveness as measured by reading and mathematics value-added. This combined mathematics and reading value-added can be considered an overall measure of school effectiveness.

Implementation of PERA Teacher and Principal Evaluation Systems 3

Key Findings on Implementation

- Almost all districts have implemented PERA-compliant teacher practice evaluation, but many are still planning and piloting measuring student achievement growth.
- Prior use of the Danielson Framework for Teaching facilitated PERA implementation in some districts.
- Joint committees that met regularly, had good communication, and high teacher involvement were the most successful.
- Dedicating staff to PERA oversight and training facilitated implementation.
- The phase-in of teacher evaluation allowed districts to move forward without causing massive disruption or resistance. It has allowed joint committees to assess their initial decisions and modify systems to respond to implementation glitches.
- Almost all districts surveyed monitor their systems for compliance, and most seek feedback on the usefulness of their approach from teachers and principals.

This chapter addresses the evaluation questions related to the implementation of PERA-compliant evaluation systems. First, it describes the progress study districts have made implementing PERA-compliant teacher and principal evaluation systems by providing information on what the systems look like. It then reports on the challenges districts faced in implementing with fidelity and discusses lessons learned that could help other districts implement evaluation systems more efficiently. The chapter begins with the teacher evaluation systems and then discusses the principal evaluation systems.

Components and Processes of the Teacher Evaluation Systems

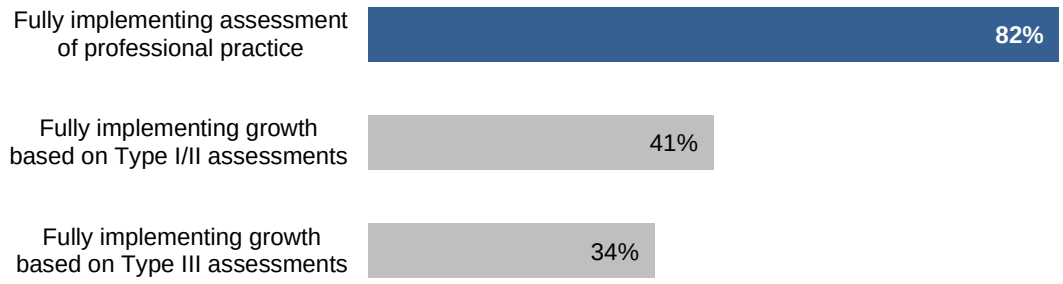
This section discusses the implementation of PERA-compliant teacher evaluation systems, including the progress districts have made in putting practice and growth components into place, the role of joint committees, and district approaches for monitoring their systems and addressing challenges.

How have districts implemented teacher evaluation?

PERA requires that districts implement several components as part of the teacher evaluation system: a practice rating, a measure of student achievement growth based on Type I and II assessments, and a measure of growth based on Type III assessments. Twenty-nine percent of the surveyed districts appeared to have fully implemented teacher evaluation systems, including assessment of professional practice and student growth measured by both Type I/II and Type III assessments. Another 49 percent were piloting either Type I/II or Type III student growth measures for evaluation, and 22 percent are still in the planning stages of measuring teacher performance using student growth.

As was the case in the interim report, we found that districts reported fully implementing assessment of professional practice to a greater extent than student growth assessments (Figure 1).

Figure 1. Districts reported fully implementing assessment of professional practice to a greater extent than student growth assessments



The Danielson Framework for Teaching was the dominant professional practice rubric used. All surveyed districts reported using it. Respondents in interviews and focus groups reported favorable impressions of the Framework and described it as "fair," "valid," and "meaningful."

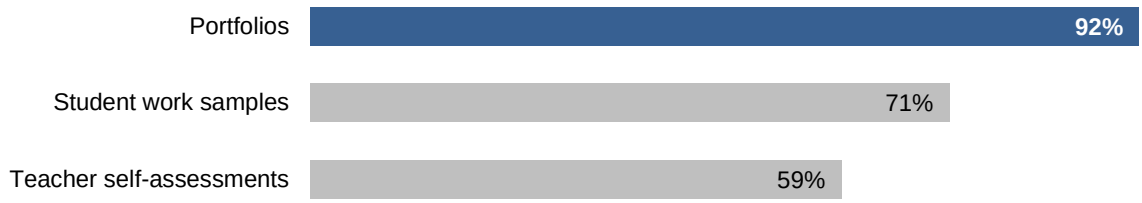
Most districts reported using similar procedures for assessing teacher practice. Some of the common procedures reported in the survey included the following:

- All responding districts evaluated new teachers every year.
- Over 90 percent of responding districts reported evaluating tenured "needs improvement" and tenured "unsatisfactory" teachers every year.
- In 95 percent of responding districts, tenured "proficient" teachers are evaluated every 2 years.

- Most responding districts made two formal observations of year 1–3 and year 4 nontenured teachers (59 and 69 percent, respectively) and one observation for tenured teachers (64 percent).
- The most frequent number of informal observations for nontenured (year 1–3 and year 4) and "proficient" tenured teachers was one. For tenured teachers with needs improvement or unsatisfactory ratings, the most common number of informal observations was two.

Districts also reported using similar types of supplemental information to evaluate teacher practice, with teacher portfolios being the most common type (Figure 2 and Table A-1 in Appendix A). District staff interviewed reported that artifacts round out observations and can give a fuller picture of performance when included in evaluations by providing an opportunity for teachers to demonstrate what they do in their role outside of what can be directly observed during classroom teaching (e.g., planning and prep, parent contact) and/or allow teachers to demonstrate their performance.

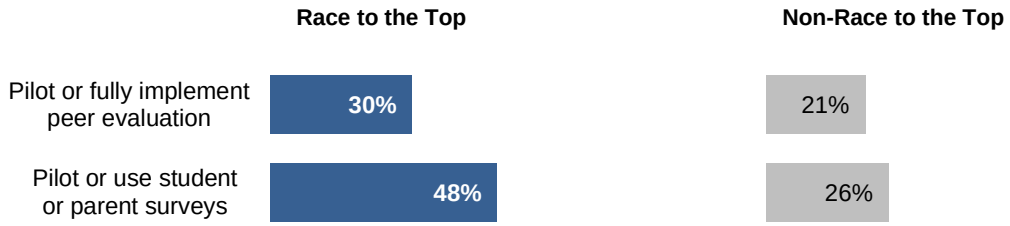
Figure 2. Portfolios were the most common type of supplemental information used to evaluate teacher practice



Most districts responding to the survey reported developing modified or specialized rubrics (mainly based on the Danielson Framework) for nonteachers, especially for counselors (75 percent), psychologists/social workers (83 percent), and nurses (48 percent). Based on the interviews in the case study districts, those being evaluated under the modified rubrics expressed less satisfaction or confidence in the validity of the evaluation system.

While not a PERA requirement, RTT districts were encouraged to adopt peer evaluation and additional measures beyond those based on classroom observation or growth. Among survey respondents, RTT districts were more likely than non-RTT districts to use peer evaluation and student or parent surveys as additional measures (Figure 3).

Figure 3. Race to the Top school districts were more likely than non-Race to the Top school districts to pilot or fully implement peer evaluation and pilot and use student/parent surveys



Some districts already had a peer assistance and review (PAR) process in place prior to PERA, for example, the Niles Township School District. There, nontenured teachers and those who are rated as "needs improvement" participate in PAR in which they are evaluated several times over the course of the year by Danielson-certified evaluators. Interviewees in Niles attributed the use of PAR and the Danielson Framework prior to PERA as a factor in contributing to their success with implementing PERA evaluation systems.

What factors contributed to the success or lack of success of the joint committees in developing and agreeing on district evaluation systems?

Factors identified that contributed to the success of joint committees include regular/frequent meetings, good communication, and teacher involvement.

PERA and Senate Bill 7 required the use of a joint committee to incorporate data and indicators of student growth into the evaluation plan. Joint committees were required to be "composed of equal representation selected by the district and its teachers, or where applicable, the executive bargaining representative of its teachers." In the interim report, joint committees were identified as an important factor in communicating evaluation processes across districts.

Interviews in districts suggested that regular/frequent meetings, good communication, and teacher involvement contributed to joint committee success. Interviewees from Unity Point Community School District reported that every constituency was represented on the joint committee, and all members had good background knowledge. Additionally, joint committee members all had leadership roles, and the process was teacher-led from the beginning. While teacher- and district-selected "equal representation" was a requirement of PERA, another school district interviewed reported that across-grade-level

representation of teachers was problematic. In this case, interviewees perceived the committee over-represented elementary teachers. Another positive factor was the use of additional committees or subcommittees to focus on specific issues. For example, the Urbana School District had several committees working on different aspects of the evaluation system besides the joint committee: one focused on student growth, one on evaluations, and one handles dismissals. Some districts also reported that it was useful to have specific skills on the joint committee. For example, the Niles Township School District included three mathematics teachers with interest in the measurement aspects on the committee. They helped the others understand the mathematics behind the measurement and the consequences of proposed decisions.

Have districts had the opportunity to revise or refine the decisions initially made by the joint committee?

Districts use the joint committees as a mechanism to review/consider feedback and recommendations and make corresponding changes in the evaluation systems for subsequent years. Changes often focused on procedural issues (e.g., deadlines, streamlining processes) or slight changes to the evaluation tools (e.g., clarifying language in rubrics) or facilitating the evaluation process (e.g., reducing burden).

Many districts have processes in place to review their evaluation systems and to make revisions accordingly. In the case study districts, these included: 1) having channels and procedures for communication about the evaluation systems; 2) collecting feedback formally via surveys and small group conversations with building leadership, union leaders, and/or the joint committee; and 3) joint committee reviewing/considering feedback and making changes in future actions.

Examples of process changes reported by study districts included:

- **Streamlining process steps and timelines.** The Sterling School District streamlined some of the steps of the process and adjusted the timing of activities to reduce the time and effort burden. Zion-Benton Township High School District adjusted some of the preconference requirements and simplified the Domain 4 professional profile.
- **Modifying observation rubrics.** Sandoval School District found the observation instrument was especially problematic for new teachers, so it made some minor tweaks. It also made minor modifications for special education teachers to recognize that some practices are not feasible for lower functioning students. Unity Point School District also reported making revisions based on piloting and feedback, adding clarifying language to help make the process more "foolproof." Interview participants in Zion-Benton Township reported making revisions to the rating

instrument based on feedback from surveys of teachers about the pilot. They identified some ambiguous areas and worked to make them clearer using enhanced Internet-based rubrics with mouse-overs and clickable examples. The Rantoul City School District initially weighted all of the components of the evaluation rubric equally, but then decided that there were four components that were most important and weighted them higher.

- **Refining student growth measures.** Sandoval Community Unit School District will allow more flexibility in the student growth component; next year it will allow teachers to add interim assessments, instead of just pre- and posttests, and count students as demonstrating mastery if they meet goals at any point during the school year, not just on the posttest. Unity Point also revised its student learning objective (SLO) process, beginning with the state default language and then customizing and clarifying it after the piloting process. Through the SLO pilot, Unity Point realized the importance of a mid-year check-in and the need to create consistency in how evaluators were conducting and documenting it. In response, the district created forms to make this process easier for evaluators and to provide teachers with a guide to the questions the evaluators will be asking about their SLOs to help teachers validate their assessments.

- **Communication.** Sterling defined additional communication channels and increased the frequency of communication about implementation. The district provided an electronic document library with information about the system for staff.

What resources did teachers, principals, and evaluators find to be valuable to understand PERA requirements and understand their districts' evaluation process?

Districts relied on a variety of sources to provide information on the evaluation components and provide training and support on specific elements. The most common resources referenced for understanding PERA requirements were the PEAC website, webinars, conferences, and meetings. To provide teachers and principals with information about the district evaluation process, almost all districts provided informational sessions/trainings.

In interviews, district staff reported that, besides PEAC- and ISBE-provided resources, they also referenced or consulted with the Consortium for Educational Change (CEC), staff of local universities, Kids at the Core, Danielson Framework experts, the Danielson Framework for Teaching book, the Illinois Association of School Administrators (IASA), and other districts to develop and implement the components of the evaluation systems. These districts considered the CEC consultations to be the most helpful resource for designing and implementing their systems. CEC consultants helped districts understand and establish both the professional practice and student growth components of their systems (e.g., practice measures, SLOs, Type III assessments), and provided examples from other locales implementing PERA-like evaluation systems (e.g., those in Austin, Denver, and Rhode Island). Interviewees from Washington School District reported using videos provided by IASA as being very

helpful for summarizing important points from the Danielson Framework, providing examples of various performance levels, and providing an overview of evaluation processes. Similarly, Urbana School District participants indicated using Teachscape training courses and videos to provide training on the Danielson Framework. Interviewees from Rich Township School District reported that, in addition to information from ISBE and CEC, they observed other districts to help develop their evaluation plan. They also disseminated knowledge from state trainings by having a select group of teachers attend the training and then share the information with other district teachers.

In addition to state resources, university staff, and consultants, some districts have also developed their own resources to facilitate implementation of the teacher evaluation systems. For example, interviewees from Zion Elementary School District reported building a framework around years of growth data to facilitate establishing growth goals. Interviewees from East Richland School District reported designating specific personnel as "PERA experts" to develop documents to provide an overview of PERA to be used in evaluations (e.g., documents to calculate student growth) and for individual and group trainings (e.g., developing SLOs). The interviewees reported these designated personnel as a strong asset in implementing PERA. Interviewees from Evanston School District reported that its assessment department compiled a set of tools to assist teachers with the SLO process. Similarly, interviewees from Sandoval School District reported building a bank of SLOs to make it easier for teachers in the future.

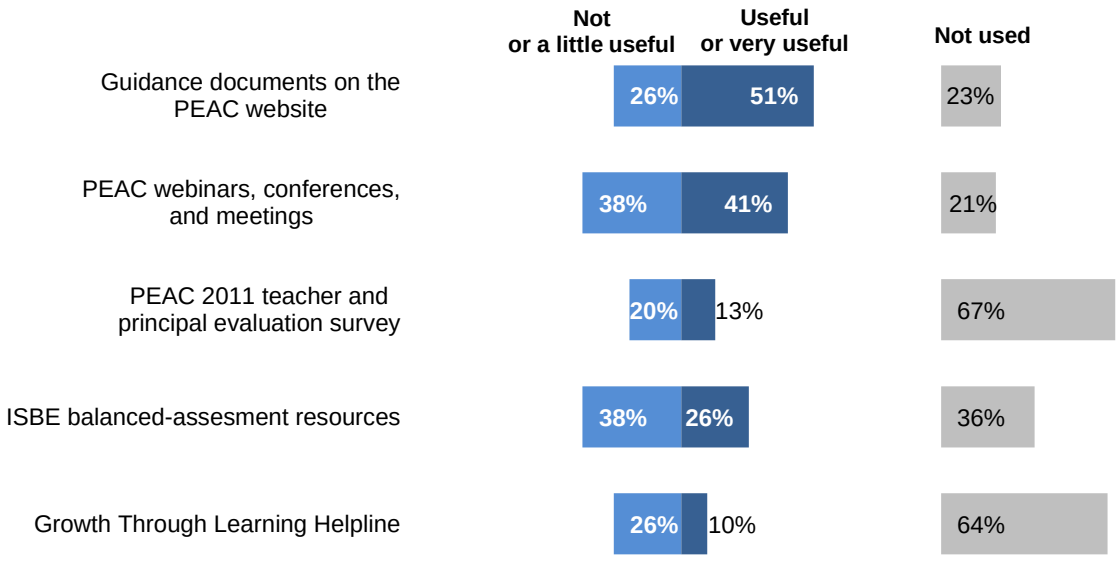
To help teachers understand the components of the evaluation system, districts responding to the survey reported they provided training on the various components (Figure 4).

Figure 4. Almost all school districts reported providing informational sessions/trainings to teachers to help them understand growth measures and informational sessions/trainings for the summative rating

How useful did districts and joint committees find resources provided by PEAC for developing and implementing their systems?

Districts responding to the survey rated the guidance documents on the PEAC website and PEAC webinars, conferences, and meetings as the most useful (Figure 5).

Figure 5. Guidance documents on the PEAC website and PEAC webinars, conferences, and meetings were rated as among the most useful resources



While most of those interviewed appreciated PEAC resources, some respondents were critical of the timeliness and usefulness of the resources provided, stating that they were not enough to help them develop a working system. While some interviewees recognized the challenge faced by PEAC and ISBE in providing resources for PERA, they still mentioned that resources were not always available on time, which delayed implementation in their districts.

How are districts monitoring whether their evaluation systems are being implemented as intended?

Almost all districts reported monitoring their evaluation systems for procedural compliance (e.g., number and frequency of observations). Districts also reported using a variety of activities to monitor the quality of the evaluation systems. Often, the joint committee was used as a mechanism to monitor the implementation of the systems.

- All districts responding to the survey reported monitoring implementation of the teacher evaluation system in some way. Over 90 percent of the districts reported: a) tracking the training of teacher evaluators to ensure all were properly trained (97 percent), b) monitoring whether evaluations were completed on time (97 percent), c) tracking whether teachers were observed the requisite number of times (95 percent), d) examining distribution of teacher evaluation ratings (92 percent), and e) assessing whether evaluators have necessary competencies to implement evaluations.

- Ninety-two percent of districts reported appointing a person or committee to be responsible for ensuring the teacher evaluation system is implemented as designed.

- Ninety percent of districts used three or more strategies to assess implementation

- All of the districts reporting they had fully implemented both the professional practice and growth components of the teacher evaluation systems (n=12) used multiple types of information to monitor the quality of implementation. All tracked whether teachers had been observed the required number of times, tracked whether evaluators had been trained, assessed evaluators' understanding of the evaluation system, and reviewed final evaluation rating forms for quality.

Table A-2 in Appendix A provides additional information on actions districts reported taking to monitor implementation.

In interviews, district staff reported using their joint committees to monitor the evaluation systems. For example, in Sandoval Community Unit School District, teachers can share concerns with committee members at any time during the school year, and there is an annual review by a standing committee to recommend or make major changes. Sandoval also uses its standing committee's annual review as an opportunity to examine outliers/mismatches in cases where practice and growth measures do not align. Other districts supplemented the joint committee. For example, Unity Point School District established a committee that meets quarterly to monitor frequencies of observations, sort teachers into tiers for Senate Bill 7 (reduction in force), and develop SLO tools (e.g., a bank of SLOs and an assessment inventory). One important monitoring concern is rater consistency. Many districts went beyond the training provided by Growth Through Learning to help promote consistency. As an example, Rock Island-Milan School District provided extensive training for teacher evaluators beyond the modules mandated by ISBE. In

Rock Island-Milan, professional development has focused on inter-rater reliability of the teacher observation rubric through watching videos demonstrating specific domains, rating them, and comparing ratings. Professional development has focused on ensuring that evaluators understand what goes into each rating and how artifacts and student growth scores are incorporated into a summative score. Interviewees in East Richland School District reported principals take an inter-rater reliability online training, and in administrator team meetings, they discuss how they develop ratings.

While 62 percent of the districts responding to the survey reported having evaluators rate videos or artifacts and compare ratings, fewer (42 percent) used multiple observers to rate actual observations. An interesting variation was used in East Richland, where teachers' union representatives accompany school administrators when doing observations to provide a reliability check.

How are districts evaluating the usefulness of their evaluation systems for improving teaching practice and for making other personnel decisions?

Assessing the usefulness of PERA evaluation systems is a common practice, with 83 percent of districts responding to the survey reporting such assessment. Districts take a variety of approaches to this assessment, including examining whether performance ratings improved over time or whether teachers participated in professional development related to weaknesses identified by the evaluation, and by interviewing/surveying teachers about improvements in practice (Figure 6 and Table A-3 in Appendix A).

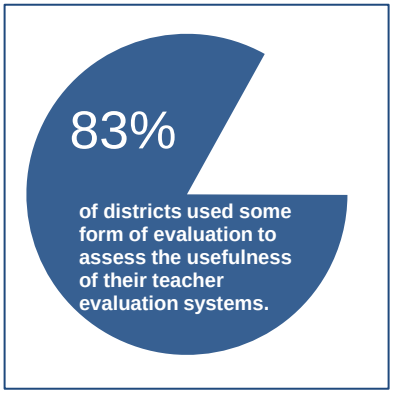
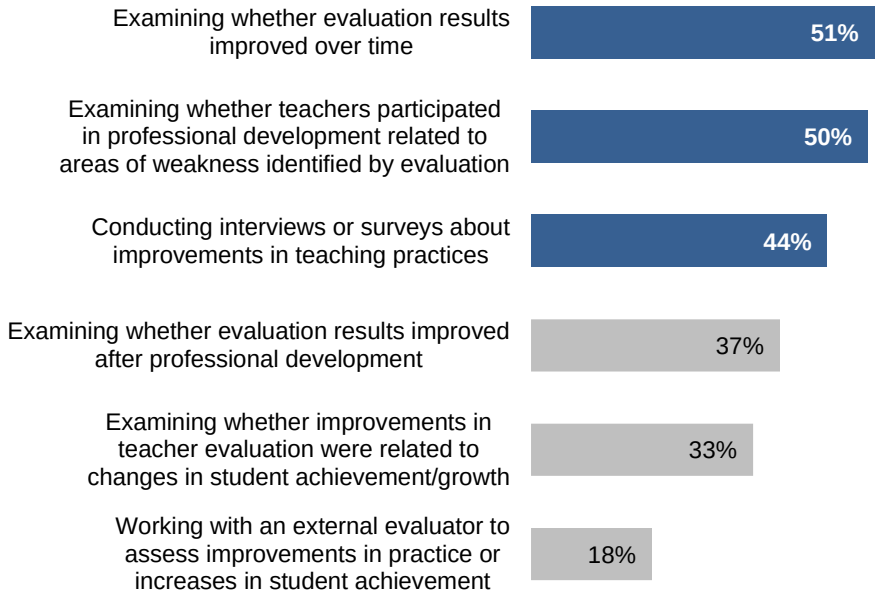


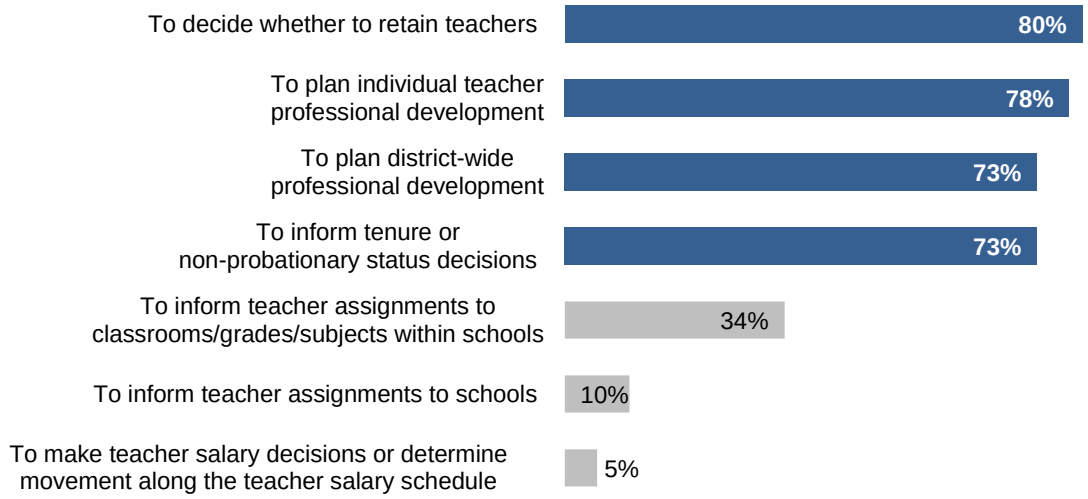
Figure 6. Common ways districts evaluate their systems include examining whether evaluation results improved over time and teacher participation in professional development related to areas of weakness



In addition, almost all of the 12 districts that reported fully implementing both the professional practice and growth components of the teacher evaluation systems used multiple types of information to assess the effects of their evaluation systems. The most common method these districts used was to examine whether evaluation results had improved over time.

Most districts are using evaluation data to inform personnel decisions such as retention, individualized and district-wide professional development, and tenure status decisions. Districts report using evaluation data less frequency for teaching assignments (e.g., grade level), promotion decisions, and/or movement on the salary schedule (Figure 7 and Table A-4 in Appendix A).

Figure 7. Most districts are using evaluation data to inform personnel decisions such as retention, individualized and district-wide professional development, and tenure status decisions



While evaluation data can inform many human resource decisions, PERA specifically requires that all districts use evaluation data to make reduction-in-force (RIF) decisions. Interviewees in Zion-Benton Township School District shared that based on requirements in Senate Bill 7, a RIF based on the evaluation results would lead to differences in who would be released compared to the old system, although they had not engaged in a RIF to that point.

What challenges have districts encountered implementing PERA-compliant teacher evaluation systems, and how have these been addressed?

The most frequently cited challenge was lack of time.

Across districts, the chief concern was the time burden stemming from all of the required activities to develop and implement the systems (e.g., the amount of time required to schedule, prepare, and conduct observations; compile artifacts; complete paperwork; establish and review growth goals; and conduct conferences, etc.). The time burden associated with these activities and practices surfaced repeatedly in the interviews. For example, interviewees reported that simply scheduling observations is time consuming, but conducting the observations, in addition to any pre- and/or post-conferences, is more so, especially because teacher evaluators (usually principals) must complete multiple observations, formal and informal, for multiple staff within a specified timeframe. So for many, the task is overwhelming and feels rushed. Additionally, as some districts noted, the amount of documentation and paperwork associated with the observations is an additional time burden.

Teacher evaluators were not the only persons burdened by demand on time and workload. A few districts explained that teachers being evaluated also have a limited time to prepare for evaluations and that teachers may be spending too much time collecting evidence and artifacts to demonstrate practice and completing other required tasks (e.g., self-assessments).

Lack of time to complete all evaluation activities also was perceived as undermining the benefit of the evaluation process.

The implications of "rushing" through the evaluation activities can potentially undermine the intended impact of the evaluation system. For example, interviewees noted that time limitations potentially reduce the perceived validity of the observation component, since evaluators and teachers question the quality and fairness of the observations when activities are rushed. Also, if due to time constraints teachers receive only formal (i.e., announced) observations, some believed this makes it possible for teachers to "put on a show," which also decreases the validity of the observations.

Another challenge is teacher and principal perceptions of lack of validity and/or reliability of evaluation tools. To some extent, this has been addressed by providing training to evaluatees and evaluators in the observation rubric, and/or the student growth measures.

Interviewees expressed concerns to varying degrees with the validity and reliability of the evaluation components. While, in general, teachers and evaluators alike believe that the Danielson Framework is a valid and thorough review of professional practice, teachers expressed concerns about evaluators' rating biases, lack of knowledge of the Danielson Framework, and/or consistency in applying the ratings, each of which can decrease validity and/or reliability. Some interviewees also expressed validity and reliability concerns regarding the information/artifacts used to supplement observations for the assessment of professional practice. While some interviewees believed they were useful because they help to "round out" observations by allowing teachers an opportunity to provide further evidence of their performance and effectiveness that might not have been captured during the observations, others had concerns about the validity of artifacts because they believed teachers could "doctor" them to achieve higher ratings. Also, interviewees expressed concern about the consistency of integrating artifacts into a rating. For example, there might be variation in how teachers present artifacts, as different teachers and administrators may have different ideas of what should be included in their presentations or how comprehensive they should be. As one district noted, further guidance around the level of detail for artifacts may be necessary.

Implementation of teacher practice measurement was facilitated by experience with the Danielson Framework for Teaching.

Interviewees also indicated that where their districts were using the Framework for Teaching prior to PERA (either as an observation or induction/mentoring tool), the transition was easy because teachers were already familiar with it.

Identifying staff and dedicating their time to PERA helped make implementation smoother.

For example, in one district, a percentage of two staff members' time was dedicated to PERA implementation. These two staff people became fully trained in PERA-compliant evaluation and served as PERA content experts, trainers, and technical assistance providers within their district to both principals and teachers. They also developed resources to educate staff and make the evaluation process more user friendly and efficient. Both teachers and principals alike spoke to the benefits of these roles.

Insights From the Case Studies: Reflections on New Teacher Observation Rubrics

Overall, teachers and administrators indicated that using PERA-compliant evaluation rubrics, such as the Danielson Framework, has provided a more holistic and objective description of what is included in the high-quality instruction, rather than a subjective checklist as some prior rubrics were considered. Districts that implemented their PERA-compliant observation process as no stakes during the first year were more likely to have more clearly articulated processes for formal and informal observations and for the role and use of artifacts in subsequent years than were districts that implemented the new process more recently for stakes. As one teacher said:

> *"They are very clear with us on the four domains, and they give us examples of what should be observed and what artifacts can show how you're meeting each of the domains."*

All teachers and evaluators also pointed out the value in a balance of planned (or formal) observations and informal (drop-in) observations as a means of capturing a more thorough viewpoint of a teachers' true performance and as a way to guard against someone putting his/her best foot forward during a planned observation and then not delivering instruction at that level when an evaluator is not present. According to an evaluator:

> *"I like the observations but…any teacher can put on a show for a day. So that's where I think the informals are so valuable. This is where we see what really happens. I think doing the formal along with the informal is a pretty accurate picture."*

While teachers and their evaluators universally praised the overall quality of the Danielson Framework, several noted limitations that created challenges. For example, as one stakeholder said, the Framework seems less appropriate to use with teachers in certain grade levels:

> *"The one challenge we have is that it's supposed to be valid and reliable across all grade levels and a variety of instructional roles. We find it's really difficult for kindergarten through second grade to have the same level of engagement as high schoolers."*

Some evaluators also expressed concerns that they did not always have the time to effectively implement a rigorous observation process, including advisement and coaching of teachers on opportunities for improvement. For example:

> *"Finding the time to get into classrooms as often as is useful in the process is the biggest challenge."*

> "[Because of the tight timeline] *Evaluators are not taking the lead on how they are going to remediate* [a teacher with performance deficiencies] *before it's time to let that teacher go."*

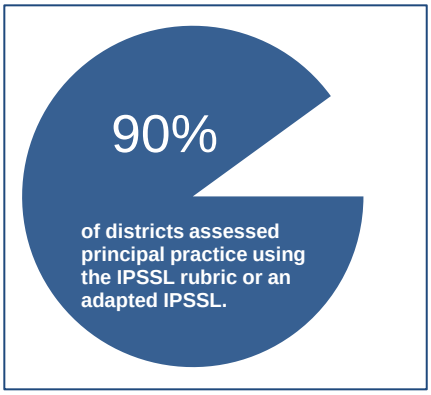
Components and Processes of the Principal Evaluation Systems

This section discusses the principal evaluation system following a similar format to the discussion of teacher evaluation systems. It discusses the components and processes, use of resources and communication, monitoring processes, evaluation practices, and challenges with implementation of the principal evaluation systems.

How have districts implemented principal evaluation?

While PERA-compliant teacher evaluation has been phased in across the state, districts were required to begin evaluating all principals each year using a combination of measures of professional practice aligned with the Illinois Performance Standards for School Leaders and student achievement growth beginning in the 2012–13 SY. As reported in the interim report, we previously found that most of the districts we studied had fully implemented these requirements, and this year, the findings were similar.

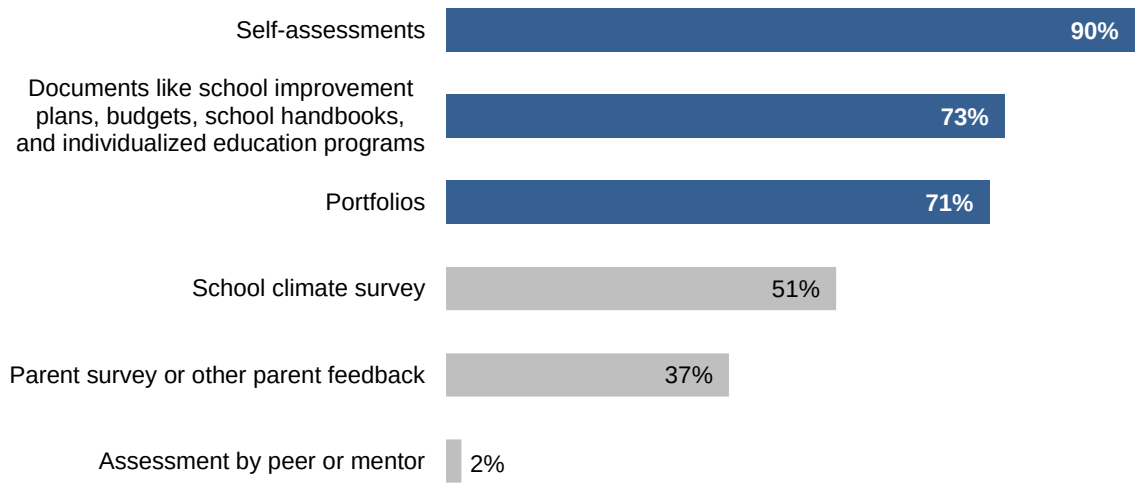
Most of the districts participating in the survey (90 percent) reported assessing principal practice using the Illinois Performance Standards for School Leaders (IPSSL) rubric or an adapted/modified version of the IPSSL. (Two districts reported using the Governors State University Model.) All responding districts reported using student achievement growth based on either Type I, II, or III assessments. Interestingly, 57 percent of the districts reported using Type III assessments to measure student growth for principal evaluation, and one district reported using only Type III assessments.



Only 13 percent of responding districts reported they assess inter-rater agreement of principal evaluators. Of the districts that assess inter-rater agreement, half reported that they do so by having evaluators rate a set of information or artifacts or two or more evaluators observe a principal and compare scores. In East Richland, interviewees reported the superintendent and assistant superintendent meet and review the evaluations and reach an agreement. For smaller districts, assessing inter-rater agreement of principal evaluators is problematic because of low numbers (e.g., only one principal evaluator).

In addition to observation, surveyed districts reported using a variety of information/artifacts to evaluate principal professional practice, including self-assessments, school operation documents (such as school improvement plans or budgets), and portfolios (Figure 8 and Table A-1 in Appendix A).

Figure 8. Self-assessments, school operation documents, and portfolios were commonly used to evaluate principal practice

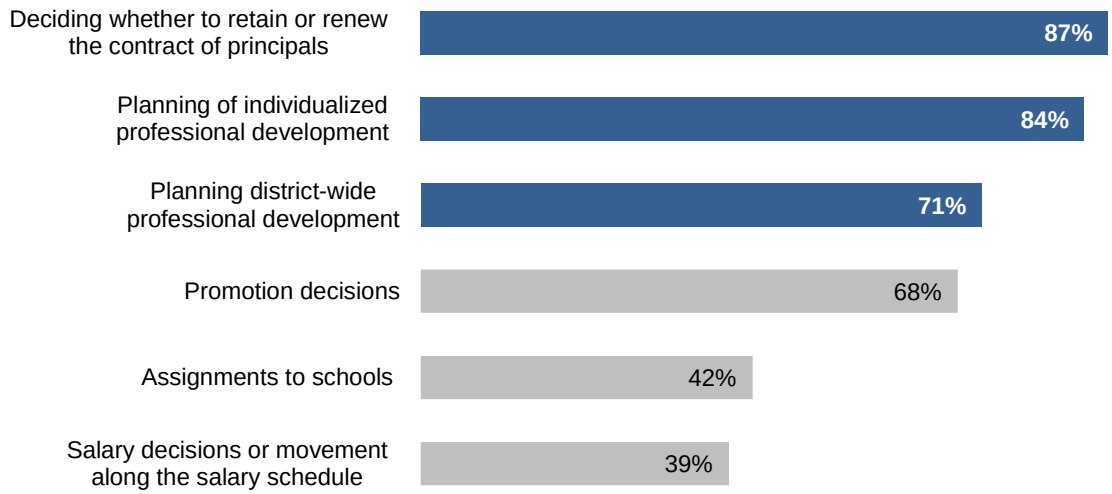


All surveyed districts reported evaluating both principals and assistant principals every year. Districts typically formally observe principals twice within an evaluation cycle and assistant principals once. Districts most frequently reported requiring one informal observation for principals and assistant principals, although there was greater variation in policy between districts for informal compared to formal observation.

Only 24 percent of responding districts used different measures to evaluate assistant principals from those used for principals. Twenty-two percent used different student growth measures and 11 percent used different nonacademic student outcome measures. Some interviewees expressed concern that their districts do not tailor the goals for assistant principals enough since they are required to use the same goals as principals. Only one of the districts responding reported using a different method of student growth measurement for evaluating first-year principals.

All districts responding to the survey indicated that they use principal evaluation data to make human resource decisions. Retaining or renewing the contract of principals, planning of individualized professional development, and planning district-wide professional development were among the most common ways districts do this (Figure 9 and Table A-4 in Appendix A).

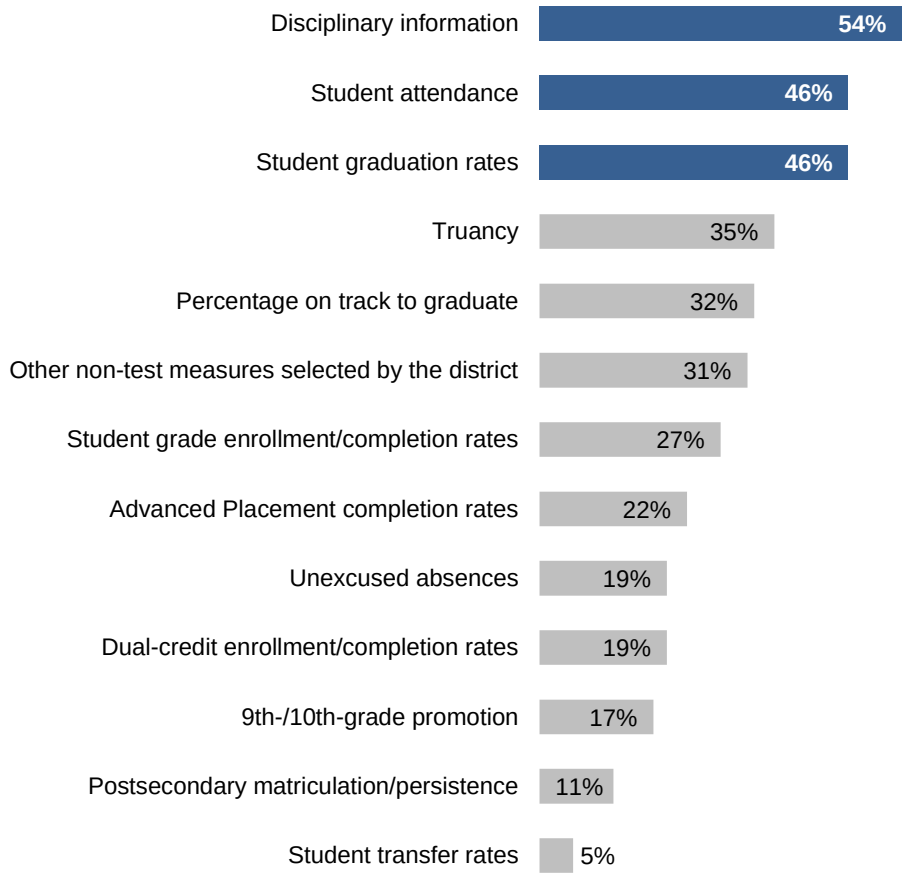
Figure 9. Districts commonly reported using principal evaluation data to inform retaining or renewing principal contracts, planning individualized professional development, and planning district-wide professional development



What nonacademic set of student outcome measures do districts use?

Surveyed districts reported using a wide range of nonacademic student outcomes measures to assess principals' performance, including disciplinary information, student attendance, and student graduation rates (Figure 10).

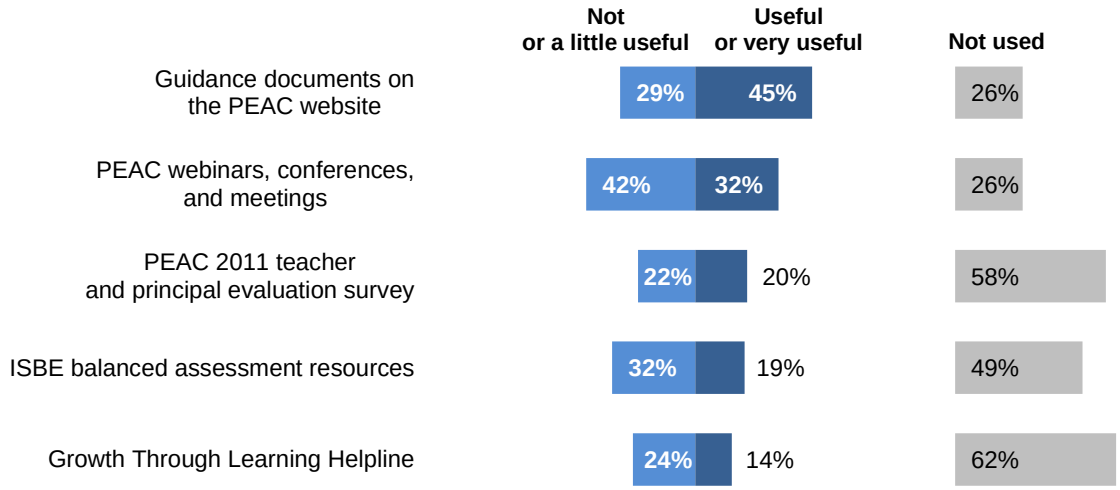
Figure 10. Disciplinary information, student attendance, and student graduation rates were the most common nonacademic student outcomes used to assess principal performance



How useful did districts find resources provided by PEAC for developing and implementing their principal evaluation systems?

As for teacher evaluation, districts responding to the survey rated the guidance documents on the PEAC website as the most useful (Figure 11). Some districts participating in interviews described other resources that were useful. For example, Rich Township School District reported using principal mentorships within the district to provide support. Rich Township also relied on training and support from the local university (Governor's State University).

Figure 11. Districts rated the guidance documents on the PEAC website as the most useful resource for developing and implementing their principal evaluation system



What did districts learn from their experiences designing and implementing their systems? How did they apply what they learned to teacher evaluation?

In general, districts did not find their experiences with principal evaluation provided many useful lessons for teacher evaluation. Some interviewees noted that although districts are supposed to be further along implementing principal evaluation systems than teacher evaluation systems, the principal evaluation systems have received less attention than the teacher evaluation systems. As one district explained, principal evaluation has taken a back seat to teacher evaluation. As a result, the training for the principal evaluation system has been weaker, and formal observation of principals has not been a priority; they have instead relied on informal conversations. This follows a historical pattern of teacher evaluation

systems receiving more attention and scrutiny, likely due to the far greater number of teachers affected, and collective bargaining agreements, which often set the terms and negotiate processes for evaluations.

Nevertheless, some districts did cite lessons that applied to teacher evaluation. Interviewees in Rock Island-Milan School District learned based on their experience with principal evaluation that substantially more communication and training would have to be provided when the system was rolled out for teachers. Glen Ellyn School District reported that starting with principal evaluation allowed the district to develop it at a smaller scale and to change data warehousing and develop student-teacher links before implementing teacher evaluations. In Rich Township School District, interviewees reported they learned the need for better alignment of principal and teacher evaluation systems. That is, giving more careful consideration to how principal and teacher student growth assessment can best work together to enhance student performance. They reported that a lack of alignment between principal and teacher student growth/level goals makes it difficult to work toward common goals and among teachers and principals.

How has PERA influenced the willingness of teachers to prepare and apply for school administrator positions?

District staff had mixed opinions on whether PERA has influenced the principal pipeline.

Interviewees in Rich Township reported PERA has a minimal influence on the pipeline for teachers to seek administrative roles. However, only those teachers already certified as evaluators are invited to apply for administrative positions since certification is expensive ($600), and districts do not want to pay for it. In Rock Island-Milan, interviewees were divided on this question. Some interviewees felt that PERA negatively influenced teachers' willingness to pursue an administrator certification, specifically due to the demands PERA puts on principals to evaluate teachers and staff. Other interviewees felt that if a teacher wanted to be an administrator, she/he would pursue it regardless with eyes wide open. Interviewees in the East Richland School District reported that the additional burden of teacher evaluation likely dissuades people from pursuing administrative work. Another district reported fewer applicants for an administrative opening compared to a couple of years ago.

State data show no change in the rate of principal attrition following the implementation of PERA.

An analysis of principal attrition using Illinois state personnel data was conducted with the aim of exploring changes in the rate of principal attrition following the implementation of PERA evaluation

requirements for principals in 2012.[1] The analysis showed that principal attrition increased from 2012 to 2013, but teacher attrition increased at an even higher rate (Table 2). However, principal attrition returned to pre-PERA levels from 2013 to 2014 and in the subsequent school year. The unusual spike in attrition rates in 2012-13 coincides with the implementation of a new state personnel data system, which may be responsible for the uptick.

Ignoring the 2012–13 school year, principal attrition rates from the post-PERA era are consistent with those from the pre-PERA era—around 14.5 percent annually—whereas teacher attrition rates appear to have increased (from about 9.5 percent annually to around 13 percent annually). Thus, there does not appear to be an increase in principal attrition rates relative to teacher attrition rates in the PERA years. Principal attrition might increase later, when all principals feel the full burden of teacher evaluation. But, based on these data, it seems unlikely that PERA's principal evaluation requirements have affected principal attrition.

Table 2 Principal and teacher attrition rates: 2006 to 2015

	Attrition rate			
	Principal		Teacher	
Year	Percent	Number	Percent	Number
Pre-PERA				
2006 to 2007	18.5	682	9.8	12,694
2007 to 2008	16.8	593	9.3	12,137
2008 to 2009	12.9	460	8.5	11,352
2009 to 2010	12.3	473	7.8	10,553
2010 to 2011	12.0	451	11.9	16,167
2011 to 2012	14.5	550	9.5	12,127
Average of 6 years pre-PERA	*14.5*		*9.5*	
Post-PERA				
2012 to 2013*	20.7	771	17.6	22,274
2013 to 2014	14.8	527	11.8	14,009
2014 to 2015	14.0	505	14.1	16,569
Average of 3 years post-PERA	*16.5*		*14.5*	
Average of 2 years post-PERA (excluding 2012 to 2013)	*14.4*		*13.0*	

Source: Illinois State Board of Education personnel data

*ISBE implemented a switch in data systems (from the Teacher Service Record to the Educator Information System) between 2012 and 2013, which may be responsible for the increased rates of attrition in 2012–13.

[1] Chapter 2 contains a discussion of the data used for this analysis and the steps taken to carry it out.

Principal Evaluation System Implementation Challenges

Some districts noted that turnover of district leadership made it difficult to sustain progress implementing principal evaluation. In addition to staff having to adjust to a new administrator/evaluator, turnover in leadership can influence the evaluation system, shifting the vision and evaluation goals as well as the way the rubrics are interpreted. Interviewees in at least one district noted that turnover in administrative leadership delayed implementation of the evaluation system.

Districts also noted that it has been a challenge to effectively evaluate principals because their roles are so varied (e.g., instructional leaders, disciplinarians, meeting facilitators, operations manager, etc.). As with teacher evaluation, finding time also poses a challenge. Interviewees noted it has been difficult for principals and evaluators to find time to complete observation-related tasks (e.g., self-reflections, assembling artifacts, etc.). Conducting multiple observations of principals has been especially challenging because there are often a limited number of principal evaluators in the district, and, in many cases, there is only one—the superintendent.

The dual difficulty of assessing the varied role of a principal and the time constraints can make appropriately evaluating a principal more problematic than evaluating a teacher. Although some principals believe the observation portion of the evaluation to be high quality and valid, others note that it is sometimes difficult to get a true sense of principals' performance using observations because the role of the principal is so diverse. This problem was further exacerbated by the limited amount of time evaluators and principals have to conduct observation tasks, and/or, in some cases, the insufficient number of observations conducted by evaluators.

One district explained the observations are valid when the principal evaluator is actually in the building enough to get a full picture of practice/performance. Otherwise, as another district noted, the observation component becomes a "luck of the draw," which may unfairly benefit or disadvantage a principal. For this reason, some believe that the artifacts principals submit to help to "round out" the observation are more necessary than those teachers submit to help evaluators better understand principals' day-to-day activities and performance. One district addressed the issue of the representativeness of principal observations by focusing on a few aspects of performance, but by doing more observations. In this district, principals are observed conducting the entire cycle of teacher evaluation. Superintendents and principals alike praised the value of this aspect of the observation process.

Other issues mentioned by principal evaluators or principals include:

- **Goal alignment**. Some interview participants expressed dissatisfaction with being required to use common principal and assistant principal goals across the district, while building-level factors that could affect outcomes might vary significantly.

- **The more casual nature of conversations about principal professional development compared to teacher professional development**. While conversations with teachers tended to adhere to formal, scheduled sessions, conversations between principals and superintendents tended to be more informal, flexible, and less often scheduled to a specific time. While some participants appreciated this informality and saw it as an opportunity to have an ongoing conversation rather than a one-time event, others saw this practice as risky since time dedicated to reflection and planning for improvement with principals was not being set aside as consistently as it was for teachers.

- **A lack of reliability assessment across evaluators**. One district attempted to address this issue by having the superintendent and assistant superintendent confer on ratings. However, assessing inter-rater reliability can be a challenge if there is only one principal evaluator—the superintendent—especially in small and rural districts. In one district, turnover in the administrative team prevented assessing agreement. Compared to PERA teacher evaluation systems, principal evaluation systems are monitored less for inter-rater consistency. Only 13 percent of the districts responding to the survey reported assessing inter-rater agreement.

Summary: How can the PERA evaluation systems be implemented effectively and with fidelity?

Overall, districts are making strides toward implementing their teacher and principal evaluation systems. Although, like last year, more districts have fully implemented the professional practice component, they are still phasing in the student growth component for teachers. Districts that had used the Danielson Framework for Teaching before PERA found it easier to adapt to the requirement for evaluating teacher practice. The phasing-in of teacher evaluation has undoubtedly allowed the study districts, most of which have limited resources, to move forward without causing either massive disruption or resistance. This has provided the opportunity for joint committees to assess their initial decisions and modify systems to respond to implementation glitches. Having staff dedicated as "PERA experts" to provide training and answer questions also facilitated implementation.

By design and in practice, the joint committees are an important influence on the implementation of PERA-compliant teacher evaluation systems. Factors that contributed to the success of joint committees are regular/frequent meetings, good communication, and teacher involvement. Districts used the joint committees as a mechanism to review/consider feedback and recommendations and make corresponding

changes in the evaluation systems for subsequent years. Often, the joint committee was also used as a mechanism to monitor the implementation of teacher systems.

Many districts have set up other feedback loops to help head off implementation problems. Almost all districts reported monitoring their evaluation systems for procedural compliance (e.g., number and frequency of observations). Most districts were actively evaluating the usefulness of their evaluation systems, by examining whether performance ratings improved over time or whether teachers participated in professional development related to weaknesses identified by the evaluation and by interviewing/ surveying teachers about improvements in practice. Districts are not using evaluation results for potentially controversial and disruptive purposes such as teacher assignment or movement on the salary schedule. Few if any districts had yet to use evaluation results for RIFs. Further, an analysis of state personnel data found no evidence that PERA implementation was followed by an increase in the rate of principal attrition.

While districts found guidance provided by PEAC and ISBE useful, they still experienced many challenges related to implementation, and many reported that more support would have been useful. The most frequently cited challenge was lack of time. Districts that implemented PERA more successfully appear to have drawn on outside resources, including local universities, consultants, regional educational agencies, and other districts.

Several districts mentioned that the initial "no stakes" implementation of growth for teachers was a positive feature of PERA. It helped to get teachers used to the idea and to work out bugs in assessments, data collection, and growth measures before making these measures count.

Assessments and Student Achievement Growth 4

Key Findings on Assessments

- Most districts chose relatively simple methods to assess growth that do not require extensive technical expertise, are relatively easy to interpret, and do not demand extensive data systems.
- Even when using simple methods, many districts had to make investments in infrastructure or expertise to include student achievement growth in educator assessment.
- Overall, a substantial number of districts have not yet fully implemented student achievement growth measurement into teacher evaluation.
- Districts may need more support to develop roster verification systems and data warehouses to facilitate measuring student achievement growth for teacher evaluation.
- There was not a strong or systematic relationship between the 2014–15 average growth ratings and 2014–15 school-level value-added growth in reading or mathematics.

This chapter discusses the assessments districts have used to respond to PERA's requirement to include measures of student achievement growth in teacher and principal performance evaluation. Illinois, like the majority of states, has been working to include measures of student achievement growth as one component of a "multiple measures" approach to teacher evaluation. PERA guidelines stipulate that student growth measures must represent at least 25 percent of a teacher's performance rating in a district's first 2 years of PERA implementation and at least 30 percent thereafter. Under PERA, three types of assessments are specified for use in measuring student achievement growth:

- **Type I:** "a reliable assessment that measures a certain group or subset of students in the same manner with the same potential assessment items, is scored by a nondistrict entity, and is administered either statewide or beyond Illinois. Examples include assessments available from the Northwest Evaluation Association (NWEA), Scantron Performance Series, Star Reading Enterprise, College Board's SAT, Advanced Placement or International Baccalaureate examinations, or ACT's EPAS® (i.e., Educational Planning and Assessment System)."

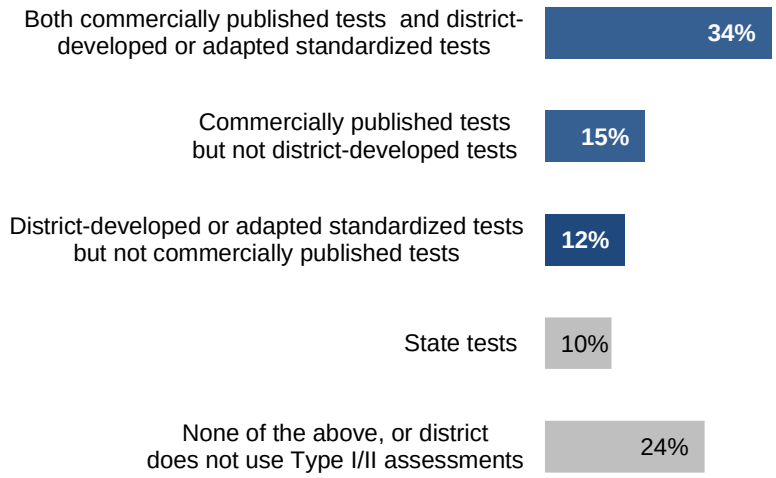
- **Type II:** "any assessment developed or adopted and approved for use by the school district and intended to be used on a district-wide basis by all teachers in a given grade, course or subject area. Examples include collaboratively developed common assessments, curriculum tests and assessments designed by textbook publishers."

- **Type III:** "any assessment that is rigorous, that is aligned to the course's curriculum, and that the qualified evaluator and teacher determine measures student learning in that course. Examples include teacher-created assessments, assessments designed by textbook publishers, student work samples or portfolios, assessments of student performance, and assessments designed by staff

who are subject or grade-level experts that are administered commonly across a given grade or subject. A Type I or Type II assessment may qualify as a Type III assessment if it aligns to the curriculum being taught and measures student learning in that subject area…."[2]

What types of assessments are used in districts for measuring student achievement growth?

Figure 12. For Type I/II assessments, most districts use commercially developed tests, district-developed tests, or both types



About a third of districts responding to the survey used both commercial and district-developed tests. Six districts reported using only commercial tests, while five used only district-developed tests. In total, 25 of 41 used both or either (61 percent). Six districts used state tests, and 10 districts reported using some other type of test. Districts typically supplemented commercial tests for teachers of grades/subjects for which commercial tests are not available.

Interviews with district staff suggested that many districts found selecting and/or developing assessments a challenge. Districts began this task from very different starting points, with some having used commercially developed Type I/II assessments (such as NWEA Measures of Academic Performance) for years or having developed their own assessments prior to PERA, while others have little prior experience either in buying or developing their own assessments. Teachers and administrators in the latter group of districts faced a steeper "learning curve" for becoming familiar with and using different types of test data, as well as with the test development process itself. While many districts noted that they have invested in assessment *literacy*, few have worked on capacity-building about assessment *development*.

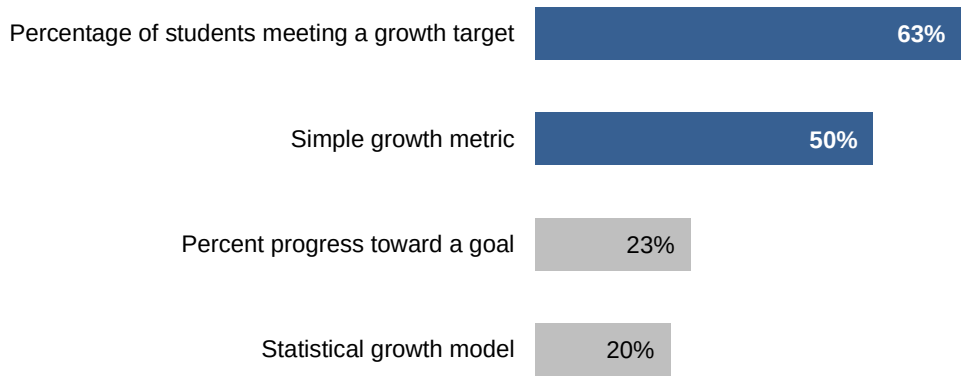
[2] 23 Illinois Administrative Code, Section 50.30. http://www.isbe.net/rules/archive/pdfs/50ARK.pdf

A number of districts noted during interviews that they have been considering the tradeoffs between purchasing and developing assessments. Many indicated a preference for the "build your own" approach and identified a variety of benefits that they anticipated to extend beyond PERA (including more opportunities for professional collaboration and greater capacity for data usage among teachers). It was not obvious, however, that the districts were fully aware of the time/resource commitments and technical support required to make the "build your own" approach workable, nor had they adequately considered where this time and support would come from. Other districts indicated that they were leaning, for at least some grades/subjects, toward purchasing available assessments, an approach which comes with tradeoffs of its own: higher technical quality on the plus side, but less teacher buy-in, missed opportunities for teacher collaboration, and an indefinite dependence on outside vendors (as opposed to building capacity from within on the negative side).

How is student achievement growth measured using Type I/II assessments?

Of the 31 districts using Type I/II assessments, 63 percent used the percentage of students meeting a growth target, half used a simple growth metric, and 23 percent used the percentage of progress toward a goal (Figure 13). Some districts used multiple methods, depending on the test and group of teachers. Table A-7 in Appendix A shows the percentage of districts that used various combinations of Type I/II assessments and measurement models to measure student achievement growth.

Figure 13. The most common ways districts measured student achievement growth was to use simple growth or the percentage of students meeting a growth target (as in SLOs)



How are districts converting Type I/II assessment results into student growth ratings?

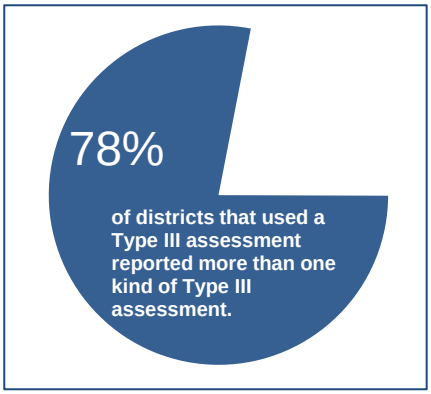
The most common method of converting growth measures to student achievement growth ratings for teachers in districts using Type I/II assessments was to set growth goals for teachers and base the teacher growth score or rating on either: a) the number or percentage of students meeting or exceeding the goals or b) the number or percentage of goals met.

- 58 percent of the districts using Type I/II assessments used the number or percentage of students meeting or exceeding goals.
- 39 percent used the number or percentage of goals met.
- 49 percent used a combination of methods.

In most of the interview districts, teacher growth ratings were based on the percentage of students who reach growth targets (though the extent to which the targets themselves are rigorous and data-informed varies widely). While some of these districts followed the state default model, under which teachers are rated as "excellent" if 80 to 100 percent of their students meet their growth targets, and "proficient" if 60 to 79 percent of students meet their growth targets, others use variations of the state model, mostly using different percentile thresholds to determine ratings.[3]

[3] For example, in one district, if 75-100 percent of students met their growth targets, the teacher was rated as "excellent"; if 50–74 percent of students met their growth targets, the teacher was rated as "proficient"; and so on. Another set the bottom of the "proficient" range at 66 percent and the bottom of the "excellent" range at 80 percent, and one set the "proficient" threshold as high as 80 percent.

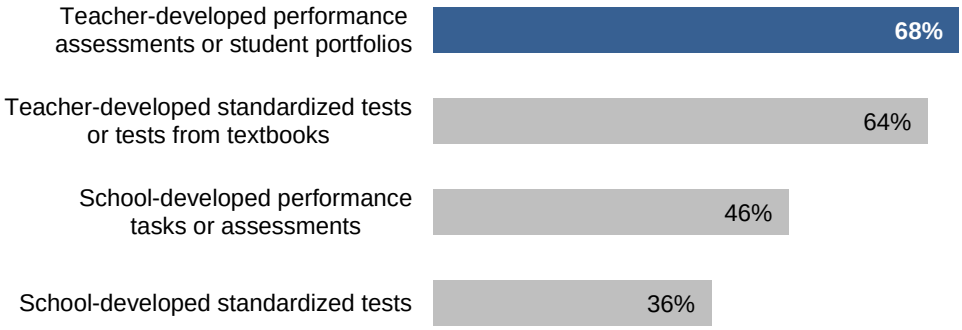
What Type III assessments are districts using?



Among surveyed districts that were using Type III assessments, a mixture of school-developed and individual teacher-developed assessments was the most common approach. Seventy-eight percent of these districts used more than one kind of Type III assessment.

Of the 28 districts that reported using Type III assessments,[4] teacher-developed assessments or student portfolios were the most common, followed by teacher-developed standardized tests or tests from textbooks, school-developed assessments, and school-developed standardized tests (Figure 14).

Figure 14. Teacher-developed performance assessments or student portfolios were the most commonly used Type III assessment



How is student achievement growth measured using Type III assessments?

As was the case with Type I/II assessments, most districts used simple growth or the percentage of students meeting a growth target to measure student achievement growth using Type III assessments.

Fifty-seven percent of districts that reported using Type III assessments used one of these two ways to measure student achievement growth. Table A-8 in Appendix A provides a more detailed breakdown of the percentage of districts using each measurement model by the kind of Type III assessment used.

[4] Thirty-two percent of the districts (13 of 41) surveyed did not use any Type III assessments for teacher evaluation in 2014–15. Nine of these districts were not yet calculating student achievement growth, and while the other four were doing so, they were not using Type III assessments.

How are districts converting Type III assessment results into student growth ratings?

Most districts converted Type III student growth scores to ratings based on the percentage of students meeting goals.

- Sixty-eight percent of the districts using Type III assessments for teacher evaluation used the number or percentage of students meeting goals.
- Forty-six percent of districts converted scores to ratings by considering the percentage of students showing some growth.
- Thirty-nine percent of the districts used multiple methods.
- Sixty-four percent of these districts used an SLO process for setting growth goals on Type III assessments; these same districts also used an SLO process for setting growth goals on Type I/II assessments.

Do district administrators, principals, and teachers believe that student achievement growth can be measured validly by Type III assessments?

Staff in all districts in which we interviewed had concerns related to the validity of Type III assessments.

Staff interviewed expressed varied opinions about the validity of growth measures based on Type III assessments. On one hand, they recognized that Type III assessments can have a high degree of content validity and buy-in on the part of teachers, since they are generally developed by teachers themselves and can be closely aligned to what teachers do in their daily instruction. On the other hand, having teachers develop assessments on their own, often with minimal previous training—and then having these measures used as part of their own evaluations—was seen as questionable.

Most staff were concerned about how to ensure that Type III assessments were of sufficiently high quality. Most districts readily acknowledged that at least some of the assessments developed and used for PERA implementation were of unknown or questionable technical quality in terms of key properties such as individual test items being bias-free, having pretest and posttest measures equated for difficulty, and the ability of teachers to set rigorous (yet attainable) growth targets based on historical data. In addition, districts report difficulty in finding historical data or useful reference points for determining what an appropriate amount of student growth might be. One district noted that scores from these assessments

were skewed toward either very high or very low performance. Some districts have found it challenging to develop meaningful scoring rubrics for performance assessments and to establish training to ensure inter-rater reliability. Another factor affecting perceived validity was whether it was a "neutral" teacher who scored the performance assessments or the teacher being evaluated. Finding neutral teachers can be difficult, particularly in smaller schools where there may only be one art or music teacher in a building.

Insights From the Case Studies: Important Factors in Successful Development of Type III Assessments

One factor that was central to understanding districts' success with developing Type III assessments was the extent of a district's mastery of assessment literacy and assessment development. Districts that had been using the same assessments for several years *and* had also developed Type III assessments prior to implementing PERA tended to have the most success with preparing teachers to develop, use, and make decisions from Type III assessments. A teacher from a district with strong assessment described the development process:

> *"We spent the whole year. My school has professional development about creating assessments; so like the knowledge and what kind of questions you're asking. With district assessments, we need to figure out where they fall, what kind of questions it takes; what kind of questions do we need to put on our assessments? Plus all of our professional development, which* [has] *been our main objective this year. It's been training for us and then we go back and we train the rest of our staff. That is, how to ask questions,* [what questions] *do we need to have on these assessments? We are kind of experimenting with it; right now we are using the district assessments for reading and math."*

Districts such as this one were also able to pilot their student growth models gradually over two school years. They used a staggered approach to onboard buildings, thereby 'working out the kinks' with regard to Type III assessment development before all buildings were implementing them.

Another factor that seemed to determine a district's success was having at least one staff member who had prior experience with assessment development. In-district expertise allowed for a district to provide teachers with ongoing coaching and support. This had the added benefit of influencing teacher practice, as one principal described:

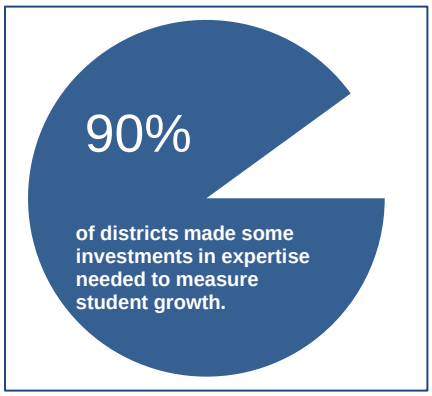
> *"Yes, it has improved* [teacher practice] *because they worked together on developing their growth assessments. It kind of forced, we're pushing common summative assessments already, that it almost forced them into having common* [assessments] *because they liked working together on the sets and looking at that pre-assessment data and coming up with the target together. If they didn't have that common assessment, they wouldn't be able to look at that same data. It kind of worked well and forced them into PLC* [professional learning communities] *discussions on coming up with talking about the assessment how they rate, what they know and targets and goals for their kids."*

Successful districts also offered a significant amount of training and allocated resources toward assessment development. In fact, all of the successful districts provided a full year of professional development opportunities as well as dedicated time for teachers to learn together. It is worth noting that struggling case study districts also provided professional development and time for teachers to collaborate on Type III assessment development, but had less prior experience and resources to draw upon. Struggling districts realized the learning curve they would have to overcome, as illustrated by a comment from a union representative from one such district:

> *Well, that's a part, we just have never thought in terms of course we've looked at student growth, we want to move kids that's what we're in the business of doing, but the idea of putting an actual measure on it, no practice in that and we have very little historical data, reliable data, to even look at trends. What can I expect because we're adopting so many new assessments so that historical kind of longitudinal data isn't available?*

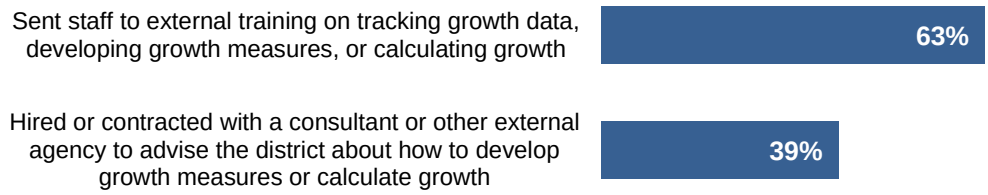
What resources are required to develop assessments and measure student achievement growth for use in educator evaluation?

Time was an important resource. Almost all the districts we interviewed indicated that sufficient time to develop assessments was an important resource for measuring student achievement growth.



Almost all districts surveyed (90 percent) made some investments in expertise needed to measure student achievement growth. Providing staff with outside training opportunities and hiring contractors with the expertise to measure and calculate growth or develop data systems were common types of investments made (Figure 15). Table A-9 in Appendix A provides more detailed information on the percentage of districts that invested in various resources to measure student achievement growth.

Figure 15. Providing training opportunities for staff and hiring outside experts were the most common district investments related to student growth measurement



Assessment literacy, experience with assessment development, and assessment training were important resources.

Interviews suggested that substantial training and identifying existing staff with expertise to coach teachers, especially at the beginning of implementation, was beneficial getting teachers up and running with Type III assessments and SLOs. Districts that had success in implementing SLOs and Type III assessments adopted a gradual implementation approach. These districts also had a high level of prior assessment literacy and experience with assessment development. Teachers from struggling districts tended to mention a need for more professional development or coaching from outside experts to assist them in developing competencies with SLOs and assessment writing. These districts had no in-district expertise to assist in assessment writing.

Many districts are developing roster verification systems and data warehouses, but even some districts fully implementing student achievement growth measures do not have these systems in place.

In order to accurately attribute student growth to teachers, especially when using value-added models, districts need to verify which students were taught which subjects by which teachers (called roster verification). A data warehouse integrating student and educator data is another resource that can help districts handle the workload of data collection, storage, and reporting that comes with measuring student achievement growth. Among districts responding to the survey:

- Twenty-six percent reported having a roster verification system ready for use or in use, while another 42 percent are planning or developing one. However, 32 percent did not have or plan to develop such a system.
- Forty-six percent store data used to measure student achievement growth in a single repository, typically a data warehouse. Seven percent have multiple data systems that need to be linked to access the needed data. The remaining 41 percent lack these resources, relying on ad hoc methods such as spreadsheets, reports from test vendors, and paper files.

Of the 12 districts fully implementing student achievement growth measures, 33 percent had roster verification systems up and running, and 50 percent (primarily the larger districts) had data warehouses in operation.

What challenges did districts face in choosing assessments and measuring growth, and how were these challenges addressed?

Time to develop assessments for grades/subjects without a predeveloped commercially available test was identified as a challenge by most districts.

Numerous districts reported that they had not had sufficient time to develop high-quality assessments for use in meeting PERA requirements and anticipated continued challenges finding time and technical support for meeting obligations related to assessment development. One district, for example, reported that it has not had enough time to give pretests or to train teachers to develop technically sound assessments; others stated that teachers do not feel that they have done an adequate job writing assessments, but also did not anticipate having sufficient time to do this considering the myriad other

responsibilities teachers have and the many initiatives that they need to implement (including Response to Intervention, new curricula, after-school activities, etc.).

Several districts indicated that they provided dedicated time for teachers and other staff to work on assessment development. One district committed time to help teachers develop district assessments aligned to the Common Core and then allowed principals and teachers to factor those assessment results into their growth targets. Another provided teachers with release and substitute time to collaborate to create and select assessments, and another district reported being able to provide time daily for teachers to hold grade-level, team, and departmental meetings to create assessments. This same district reported that it has been able to use outside funding to bring in experts to train staff on how to develop Type III assessments. A number of districts have been able to develop meaningful partnerships with external providers (such as local universities or Regional Offices of Education) to support their needs around PERA implementation, and in other instances, districts have joined together in informal collaborations to address common challenges such as training in assessment literacy and/or development. Small districts in rural areas, however, had fewer options for reallocating time or accessing external support.

Many districts are also finding measuring student achievement growth for use in principal evaluation challenging.

Several districts reported that principal goal-setting has been more challenging than they had expected. Many districts indicated, for example, that they were concerned with principals being able to choose their own assessments and growth goals, which could make ratings biased and hard to compare across schools. Another district perceived challenges to validity when applying the same goal-setting system to elementary, middle, and high school principals. Additionally, in one district, growth data actually pushed principal ratings into a higher rating category because their schools perform well, even though the district central office believed that principals' practices were not as strong as student growth ratings suggested.

Given the variation that emerged with respect to using student growth measures for principal evaluation, it was not surprising that districts reported things being "in flux," with ongoing efforts to experiment and adapt (perhaps even more so than with teacher evaluation). One reported, for example, that it continues to tweak its model, look for exemplars elsewhere, and talk to stakeholders. Another indicated that it is trying several different methods and will eventually determine which is most appropriate. Another district has identified potential challenges in using schoolwide performance measures as part of principal evaluation in cases when the principal is new to his/her building; in these cases, the district recognized that results in the early years of a new principal's tenure could more accurately reflect the effects of his/her predecessor since it takes time for a principal to influence the many factors that affect building-level student learning.

The need to finish evaluations by March was seen as a challenge in many districts.

Districts in Illinois continue to be bound by a longstanding requirement to make decisions around the retention of nontenured teachers (as well as most principals, who work on annual contracts) by March 1 of each year. Practically, many districts try to make an evaluation rating in March for most educators. This deadline poses obstacles to measuring student growth for use in both teacher and principal evaluations. In most cases, districts do not have an entire year's worth of data and evidence with which to judge an educator's effectiveness by the end of February. Accordingly, they must choose between a number of nonoptimal choices for evaluating nontenured teachers that include using only one semester's worth of student growth data (or fall-winter growth on benchmark assessments such as MAP) or using "old" data from the previous year. Thus, some if not all teachers are being evaluated based only on student growth during a portion of each school year (August to February). One district pointed out that the March 1 requirement essentially forces schools to use two sets of data in evaluating their teachers—August through February with nontenured teachers and August through May for tenured teachers. Another district worried that the need to meet the March 1 nonrenewal deadline might prevent nontenured teachers from setting more rigorous student growth goals, as it would be unrealistic to expect their students to meet high growth targets in a short timeframe.

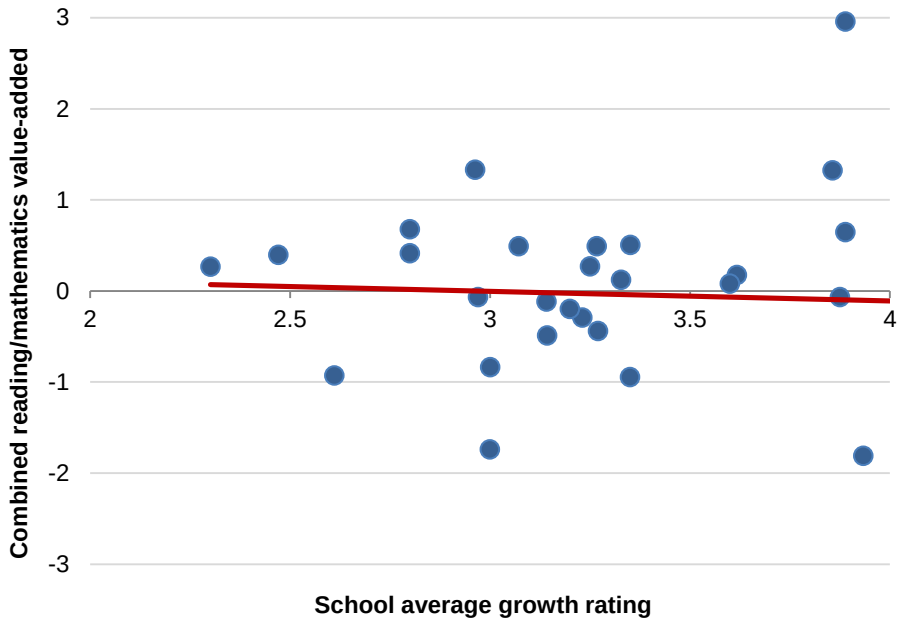
Do schools where teachers are rated high on local measures of student achievement growth also have high levels of growth on statewide student achievement measures?

To address this question, evaluation results were requested from the study districts for the 2014–15 school year. Eighteen districts provided evaluation results, though not all included ratings based on student achievement growth. School-level averages were then calculated across the teachers within schools for whom results were provided. To represent growth on statewide achievement measures, value-added estimates of school-level reading and mathematics achievement were developed. These data were available for grades 4-8. The techniques used to develop these estimates are described briefly in Chapter 2, and in greater detail in a technical supplement to this report.

There was not a strong or systematic relationship between the 2014–15 average teacher growth ratings based on district growth measures and 2014–15 school-level value-added growth in reading or mathematics based on state tests.

Figure 16 depicts the relationship between school-level combined reading and mathematics value-added[5] (centered around each district's mean value-added and the average teacher growth rating[6]), for the schools outside of the Chicago Public School District.[7] The relationship between average growth ratings and school mathematics value-added was very similar.

Figure 16. Relationship between average teacher student growth ratings and school-level combined reading/mathematics value-added



The relationship between school-level averages of teacher ratings on district growth measures and school value-added varied substantially across districts.

Figure 17 shows the correlations between average growth ratings and school value-added in reading and mathematics for the eight districts[8] for which data were available, and correlations could be calculated. Positive correlations show that schools with higher average teacher growth ratings have higher school-

[5] Mathematics and reading value-added estimates were combined to allow economical presentation of results and because the student growth scores averaged across teachers within schools could cover multiple subjects. The combined mathematics and reading value-added represents an overall measure of school effectiveness. Chapter 2 includes a discussion of how the estimates were combined.

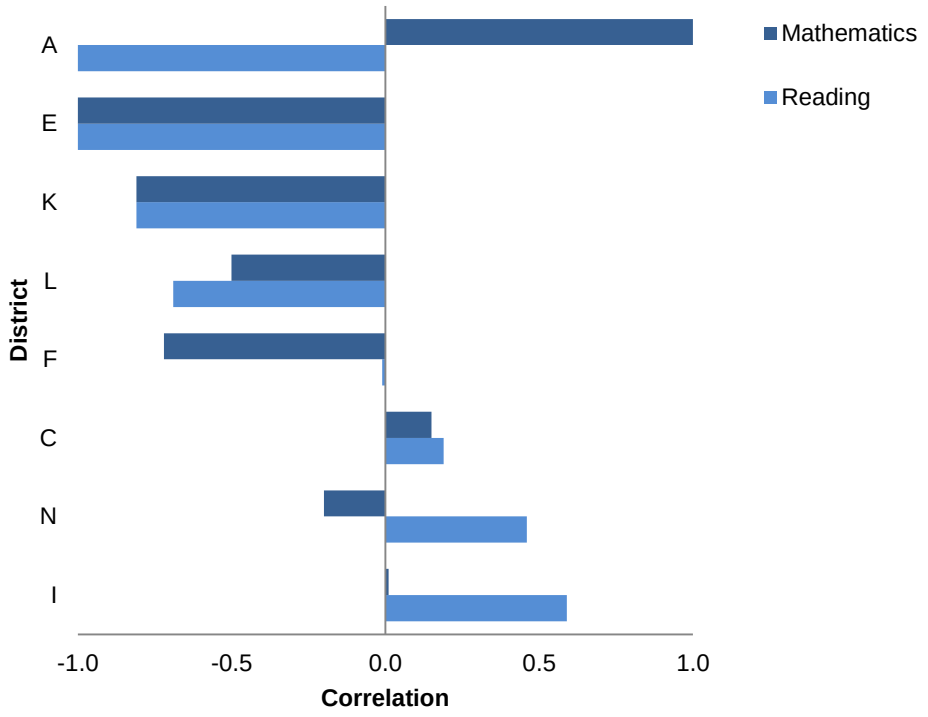
[6] Centering the value-added estimates around the district mean removes systematic differences between districts in mathematics or reading value-added growth. Since districts generally used different standards for deciding how much growth would be needed to receive each rating, as well as different assessments to calculate growth, it was appropriate to compare growth ratings to value-added relative to each district's average.

[7] Data from Chicago Public Schools are excluded due to the size of the district, which would obscure the results from other school districts. When examined separately, a positive relationship between combined reading and mathematics value-added and average school-level growth ratings is seen for Chicago.

[8] Note that for two additional districts, there was no variation in the growth ratings, so correlations could not be calculated.

level value-added. Some districts, however, show negative correlations, indicating that schools with higher average teacher growth ratings have lower value-added. Note that for two districts, data were available for only two schools, so correlations could only take the values of 1 or -1. Some of the variability shown in the figure would be expected, due to the relatively small numbers of schools in most of the districts for which data were available. Table B-1 in Appendix B shows the actual correlation coefficients.

Figure 17. Correlations between school average teacher student growth ratings and school-level reading and mathematics value-added



Because more study districts reported overall teacher summative ratings than growth ratings, the relationship between school average summative ratings and school-level reading and mathematics value-added was also examined. While the overall relationships were more positive, there was still substantial variation across districts, and no consistent pattern of results, as shown in Figure 18 below.

Figure 18. Correlations between school average overall summative ratings for teachers and school-level reading and mathematics value-added

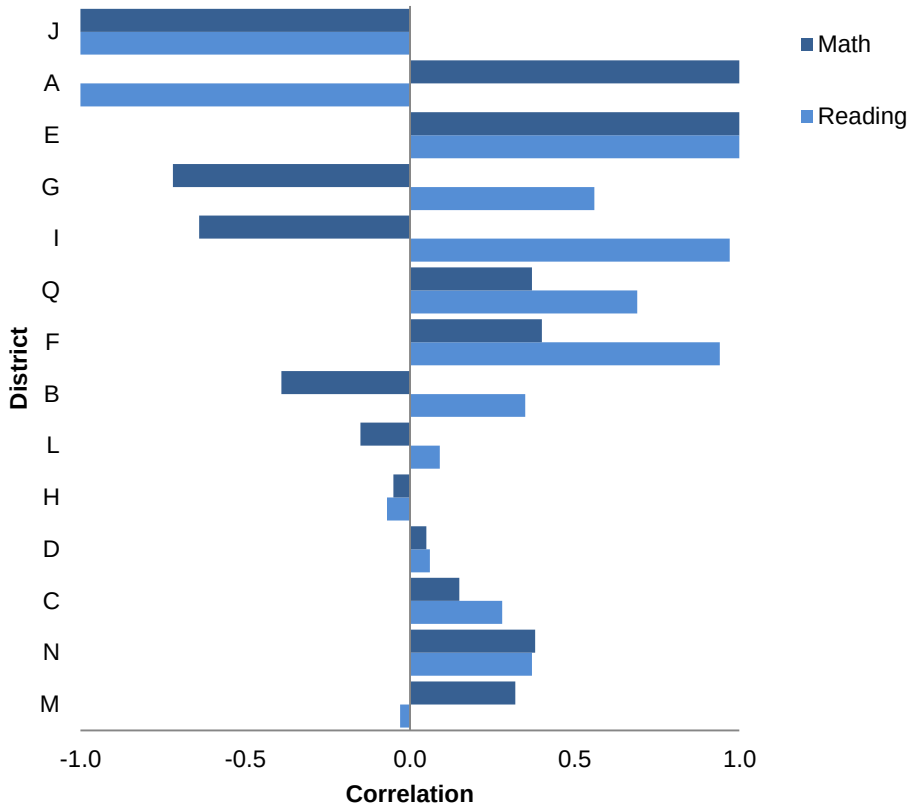
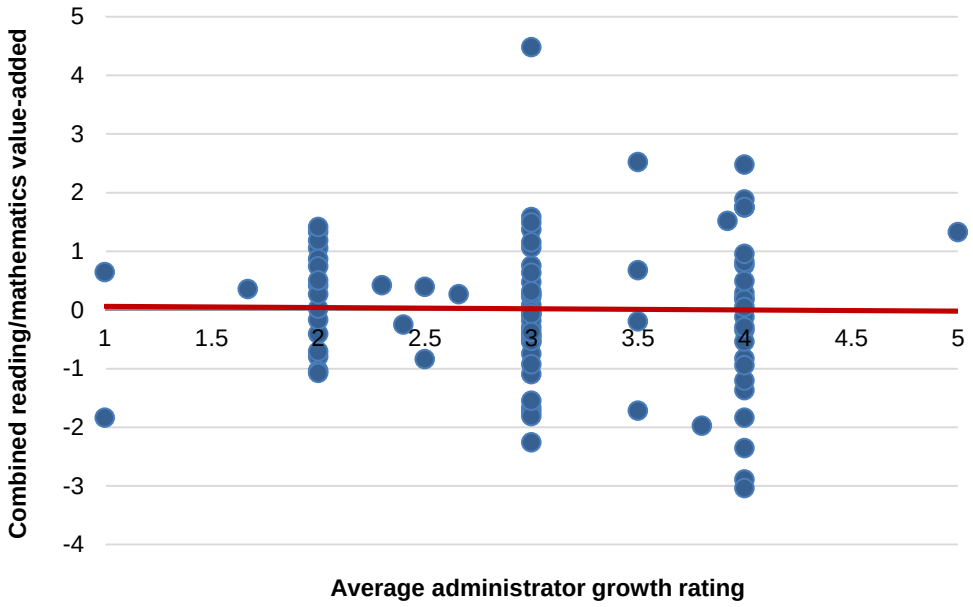


Table B-2 in Appendix B shows the actual correlation coefficients.

The relationship between average principal growth ratings and school value-added was also inconsistent.

The relationship between growth ratings and school value-added for administrators was also examined. Growth ratings were provided by 18 districts. Value-added mathematics and reading achievement were available for 91 schools in 15 of these districts. (Three districts were high school districts, for which value-added estimates could not be computed.) Figure 19 below shows the relationship between combined school value-added (centered around each district's mean value-added) and the average administrator growth rating (the average of the principal and associate/assistant principal ratings, for schools with associate and assistant principals). Again, there is no evidence of a positive relationship between growth ratings and value-added across the schools for which data were available.

Figure 19. Relationship between average administrator growth ratings and school-level combined reading/mathematics value-added



The correlations between administrator growth ratings and school value-added in reading and mathematics for the individual districts are shown in Table B-3 in Appendix B.

These results should be interpreted with caution.

First, because the tests used by districts for measuring growth are different than the tests used to construct the value-added estimates, high correlations between school-level averages of educator growth ratings and school value-added would not be expected. Research has shown that the growth estimates can differ depending on the test used (see Papay, 2011; McCaffrey, 2013; Stuitt et al., 2014). These studies found that even using similar growth models (i.e., value-added), growth results based on different tests were correlated between .2 and .65. Because districts typically used growth models that were quite different from value-added, as well as different tests, positive correlations would not likely be greater than the lower end of this range
(i.e., .2–.4).

Second, some teachers whose evaluations are included in the average may not influence student achievement in reading and mathematics as much as others. Most districts did not provide information about the subjects and grades teachers taught, so it was not possible to include only teachers responsible for mathematics or reading instruction, or weight teachers in grades other than 4–8 less. Also, since

PERA does not require that all teachers be evaluated each year, the set of teachers represented in a school average may not be the same as the set responsible for mathematics or reading instruction.

Third, many districts had only a small number of schools represented. Partly this was because school value-added was available only for grades 4–8, so ratings of teachers or administrators in high schools could not be included in the analyses. In addition, many of the districts that provided teacher evaluation ratings based on student achievement growth were relatively small. Outside of the Chicago Public Schools, the number of schools per district ranged from 1 to 7 for the growth ratings and 1 to 22 for the overall ratings. The large variation in the correlations reported above is likely due in part to the low number of schools per district.

Summary: How do districts fit measuring student achievement growth into their current systems?

Including student achievement growth in teacher and principal evaluation is a challenge that study districts are working steadily to master. A substantial number of districts are still working to fully implement student achievement growth as part of teacher evaluation. Districts are relying on both commercially developed and district-developed Type I and II assessments. Districts used a mixture of school-developed and individual teacher-developed Type III assessments. It appears to take more time to implement this aspect of PERA, as teachers and principals need to be trained and assessments developed.

Once assessments are acquired or developed, most districts used relatively simple and easy-to-follow methods of measuring growth, such as simple growth or the percentage of students meeting a growth target. Few use more complex growth measurement models such as value-added or student growth percentiles. The most common method of converting growth measures to student achievement growth ratings for teachers is to set growth goals for teachers and base the teacher growth score or rating on the number or percentage of students meeting or exceeding the goals, or the number or percentage of goals met.

Even using relatively simple methods for measuring growth, many districts still made investments in infrastructure or expertise to include student achievement growth in educator assessment. Districts also had to devote a substantial amount of time to developing assessments, which required training educators in assessment as well.

One challenge to fitting student achievement growth measurement into educator evaluation is the need to finish evaluations by March. This makes it necessary to test earlier than the state testing window and to acquire or develop, deploy, and score additional tests.

While some, especially larger, districts have developed the data warehouses and roster verification systems that facilitate measuring student achievement growth for educator evaluation, others may need more support to develop these resources.

Insights From the Case Studies: Assessing Student Growth for Principal Evaluations

All principals and district leaders from case study districts who were interviewed noted that systems and processes for measuring student growth related to principal evaluations were present in their districts. Most principals interviewed also indicated confidence in their ability to achieve their student growth goals using school-level data, although several questioned whether rating principals on student growth was even appropriate given that principals have no "direct involvement in the instruction." Nevertheless, in the initial year or years using student growth as a measure in principal performance, nearly all principals interviewed met their targets. As one principal said:

> *"Right out of the gate, attaining my student growth goal was really pretty attainable. It helps when you have good teachers too, but it's definitely attainable."*

Several factors seemed to contribute to principals' abilities to attain their student growth goal:

1. the principals' ability to have input into the goal that was set for them;
2. their ability to use prior performance on the assessments to inform goal setting; and
3. that the very idea of growth allows them to examine the school's contribution to a student's performance in a single year regardless of the level of academic achievement a student brings in from prior years.

Several challenges were also highlighted. For example, districts reported that implementation of student growth measures for evaluation had taken place in stages, with measures of student growth for principals coming before measures for teachers. This staggered implementation process led to some misalignment between the growth goals set for teachers and principals. In the view of some stakeholders, this made progress difficult:

> *"If you want to move a system forward, you're going to align everything so that everybody is working towards the same goal. So the way it is now, we have our principals working on one thing and we have our teachers working on something different."*

Another challenge expressed by all districts was in identifying Type I or II assessments that measured gains appropriate to all grade levels or within all subject areas. For example, when a principal's growth measure included literacy, districts were challenged to identify measures at the secondary level since literacy instruction is typically embedded in multiple course types and not assessed in a standardized, stand-alone way. In other cases, stakeholders viewed the skills addressed by available current Type I and Type II assessments as inadequate. As one principal said, measures that were more closely aligned to currently emphasized student skills would be of greater utility:

> *"We need better tools to measure math growth. And reading growth. It's hard for me to get excited about how many words per minute a child can read at the beginning and the end of the year. We need something more tied to Common Core and a skill that matters like comprehension."*

PERA and Professional Development: How Is Professional Development Changing and How Is It Affecting Teachers?

5

Key Findings on Professional Development

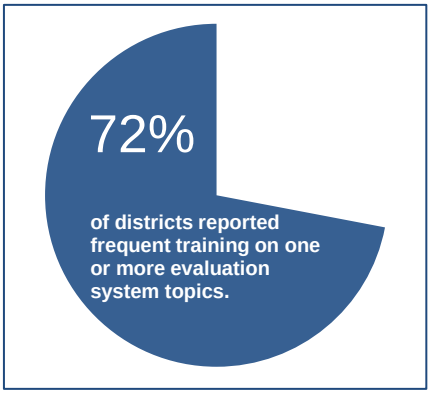
- PERA's emphasis on student achievement growth has resulted in districts providing more professional development on student assessment and goal setting.
- Most districts are now using performance evaluation results to plan district-wide professional development.
- When done well, evaluation conversations and discussions are focused on improvements in teaching practices, and this is linked to teachers' professional development.

This chapter addresses the evaluation questions related to the effects of PERA on professional development. A major goal of PERA has been to improve educator effectiveness, and to do so school districts have been expected to improve their professional development to help teachers and principals meet the teaching and leadership standards underlying the evaluation process. In order to maximize the contribution of the evaluation practices required by PERA to improving instruction and student achievement, evaluation and professional development need to be closely aligned. Three important links between evaluation and professional development include:

- Ensuring that professional development relevant to the standards of practice is available to help educators improve their performance;
- Using evaluation results to plan professional development, so as to build on educators' strengths and address areas in need of improvement;
- Using the performance evaluation process, especially conversations between educators and their evaluators, for providing feedback and discussing how to improve performance.

What changes in professional development practices have districts made as a result of implementing a PERA-compliant evaluation system?

Districts have spent considerable time and resources training educators on the system itself.



Survey responses showed that training on the evaluation system is common across districts:

- All districts responding to the survey reported discussing new evaluation practices in group professional development sessions or professional learning community meetings.
- 90 percent of responding districts reported providing teachers with training about the professional practice and student achievement growth measures.
- 84 percent provided training to teachers' evaluators on the professional practice rubric and 71 percent on providing feedback to teachers about their practice (see Appendix A, Table A-6).
- 72 percent of the responding districts reported frequent training on one or more evaluation system topics. Training on understanding the professional practice rubric was most often identified as a frequent training topic (see Appendix A, Table A-5).

Insights From the Case Studies: Training Principals to Perform PERA Evaluation Duties

For case study districts, administration team meeting time was often used to help principals strengthen their understanding of the teacher practice rubric domains and how to collect evidence to support their evaluation ratings and feedback. One district administrator shared the process of training principals on the teacher practice rubric:

> *"We said OK let's look at this standard, let's look at this component of the framework so instead of spending time with training, what we did is we spent time interpreting what we said this language was really saying. We actually talked about that. We spent our very first year doing just that. That was important because that was staff development within itself. I wanted to hear what my six principals felt like this particular component was saying and what it would look like and so forth, and so that interpretation piece was an internal professional development that we used to get a better understanding of what this particular domain and component, what it looked like, what the indicator would be if there was formally one."*

Principals were provided multimodal learning opportunities such as watching videos of teacher practice, shadowing principals during observations, scoring rubrics together, and most important, principals were provided consistent, dedicated time to discuss issues that came up during the evaluation cycle with district leadership and other principals. In addition to providing peer learning opportunities throughout the evaluation cycle, one district used this meeting time for principals to collaborate on developing professional development opportunities for teachers based on evaluation findings, as one principal shared:

> "*From the collaboration among our administrative team, we usually come up with some pretty good ideas, pretty good plans for professional development and conversations to have with our staff to help everybody improve."*

Principals thought that these professional development opportunities were effective in supporting them in their role as teacher evaluators. One principal shared *"I think that's been our best professional development to tell you the truth."*

To further support principals in their role as teacher evaluators, two districts adopted an approach that had district administration shadow principals through the entire evaluation cycle. Principals in these districts valued the added support they received as illustrated in this quote:

> *"Our superintendent is our evaluator, and he evaluates us formally twice in a year. Once he follows us through our evaluation process of a teacher, so he will sit through the entire thing and it helps me because it reassures* [me] *that I'm doing a good job.* [He] *will give us feedback, and we'll be able to compare our notes to what we saw and that helps me just reaffirm that inter-rater reliability, that he and I are both seeing the same thing and giving the same rating."*

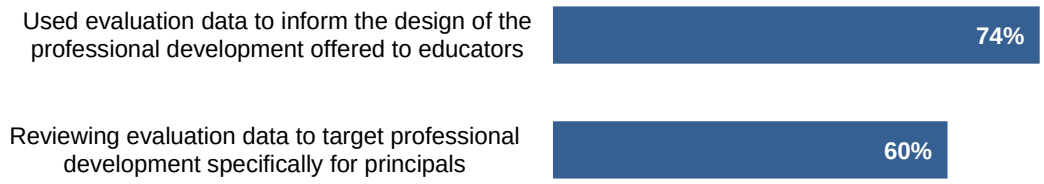
PERA has also encouraged districts to provide training on assessment and setting student achievement goals.

District staff we interviewed told us that professional development around assessment and student growth was often viewed as a highlight of the new evaluation systems. As they began designing and implementing the student growth component of the evaluation, many districts realized that teachers and administrators needed additional training and practice on data analysis and developing high-quality assessments and SLOs in order for the systems to succeed. For example, one district that was heading into the student growth pilot dedicated the year's district-wide training to assessment writing and data-based decision-making, which provided built-in time for staff to collaborate on grade-level benchmarks.

Several districts acknowledged that they had been "historically" weak in this area and that this training was well overdue. Many principals and teachers reported attending workshops on assessment design and SLOs. Some teachers reported that this additional training in assessment and data analysis has helped them feel more comfortable and competent in writing SLOs and produced a more fair and productive evaluation system. One respondent stated that training on assessment writing was the most useful aspect of his/her professional development; another noted that implementing the student growth component had led to higher quality assessments and provided additional time for data analysis; and yet another said that PERA facilitated an increase in the richness of discussion around data literacy and use.

What professional development do districts provide that is linked to teacher evaluation results? Do districts provide specific professional development based on practice ratings? On growth results?

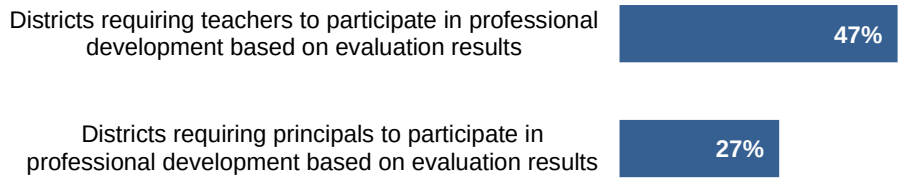
Figure 20. Most districts are now using evaluation results to inform the design of professional development offered to teachers and principals



Our interviews with district staff revealed that many districts are beginning to use evaluation data to direct professional development at the building or district level. For example, one district used data from informal walkthroughs (using the same rubric as full observations) that indicated that teachers' discussion techniques were below district goals to plan district-wide professional development around this area. Other districts were beginning to aggregate building-level evaluation results, and principals were meeting with district leadership to determine professional development plans that target areas that are the lowest across buildings. Other districts, however, reported that the professional practice component of the evaluation did not identify any across-the-board weaknesses, making it difficult for them to pinpoint where to focus district-wide professional development. Several of the larger districts interviewed also noted that they had yet to begin linking professional development to evaluation results district-wide.

Many districts also target individual professional development based on evaluation results.

Figure 21. Almost half of school districts require teachers to participate in professional development based on evaluation results; about a quarter of school districts have the requirement for principals



Additionally, 26 percent required teachers to participate in a minimum number of professional development hours each year, but only 5 percent indicated that the required professional development hours changed based on evaluation ratings.

Most of the districts in which we interviewed explicitly linked evaluation results with professional development at the individual level by targeting training toward areas identified as weaknesses. Interviewees in these districts reported that the professional practice component—and the Danielson Framework in particular—was most helpful for directing professional development at the individual level. Professional practice ratings were often used to push teachers to work with instructional coaches or seek additional resources. For example, Sandoval Community Unit School District reported that the teacher professional practice component of the evaluations, in particular, has helped administrators to be "more intentional" about directing professional development for teachers. Previously, professional development in the district tended to be district-wide, whereas it is now more customized to individual or building-wide strengths and weaknesses. Similarly, Unity Point School District reports that district teachers used to be able to do whatever professional development they wanted, but this is not the case now because professional development goals are derived from the summative evaluation and identified target areas. Some districts, such as Rich Township School District and Washington School District used the "PD 360" program to facilitate this process, which allows teachers to watch videos or get recommendations in areas identified as needing improvement based on their own evaluation results.

In some districts, such as Niles Township School District, Carbon Cliff-Barstow School District, and West Central Community Unit School District, data for directing individual professional development had come from the student growth measures, as teachers and administrators analyze test results and curricula to help determine target areas and avenues for improvement. However, linking professional development

with data from the student growth component of educators' evaluations was less common, likely because many districts had yet to fully implement student growth measures. One high school district noted that its professional development content was typically based on student growth rather than on professional practice data, and one elementary-middle school district reported that it used growth data to identify areas for improvement. Several other districts reported that they placed more emphasis on the professional practice measures than on growth, under the rationale that improvements in practice would eventually lead to improvements in growth or that improvements in growth could not occur in the absence of improvements to practice.

Insights From the Case Studies: Evaluation-Informed Teacher Professional Development

Case study interviews revealed that while professional development takes place at many levels, including district, building, team, and individual levels, it is at the individual and team levels where evaluation findings truly appear to influence professional development. For example, teachers whose evaluations identified specific areas for improvement were offered the chance to connect with teachers who excelled in these areas for observation and informal peer mentoring. Areas of growth revealed through the evaluation process were also used to determine learning modules or external development opportunities to attend, such as PD 360, conferences, and speaking events. As one principal shared:

> *"I had a conversation with a pretty new teacher, and she is not doing so well with discipline, and on the tool that is where she is getting dinged the hardest. I told her, 'I will free you up a half day, I want you to pick some teachers that you know are doing well or I'll pick them and go watch them. Or go to a workshop on discipline….' So it has driven some professional development."*

In addition to addressing areas for growth, individual professional development served as a venue where specific questions could be addressed and where teachers' concerns about the evaluation system could be eased. This was especially the case for teachers of nontested subjects or with special student populations such as art or special education teachers. As one special education teacher noted:

> *"It's been nice because we've had group in-services and meetings, but then we also have had the opportunity to have individual meetings so we can ask very specific questions…. That's really what I've found most beneficial, because I teach special needs kids. There are two of us who teach* [special education]*…. My situation is nowhere remotely the same as everyone else's. So, I've asked very different questions than I think some of the other people have and it's been nice to be able to be in an individual situation where I can just go and do that."*

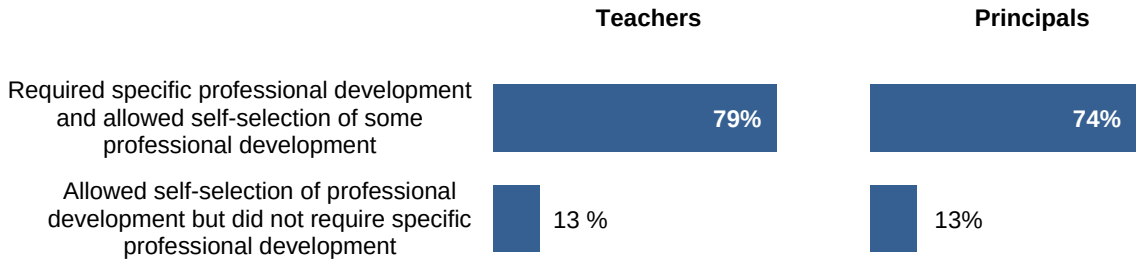
While most case study districts held professional development at the district and building levels to inform staff about how PERA-compliant evaluation would be implemented, fewer offered professional development that was rooted in evaluation results at these levels.[1] Often, this was because systems to gather and/or analyze findings across buildings were not present. Further, teachers, evaluators, and administrators noted that it was usually the responsibility of teachers to seek out professional development opportunities rather than of the district to provide them. One teacher described the process of identifying professional development opportunities as follows:

> *"If I needed support in classroom management or writing lesson plans, I would probably have to find my own, write it up, submit it to the administrator, and say this is what I would like to do and submit it for approval."*

[1] It is worth noting that other districts participating in the study have made progress in this area, indicating some variation between districts in offering district- or building-wide professional development that is informed by evaluation results.

The vast majority of the surveyed districts allowed teachers and principals to self-select some professional development opportunities, but most also required teachers and principals to participate in some specified professional development (Figure 22).

Figure 22. In most districts surveyed, professional development choice is a mix of required professional development and professional development chosen by the teacher or principal



Some districts have not yet linked individual educator professional development to evaluation results.

In 6 of the 15 districts interviewed about the impact of PERA on professional development, administrators reported that the new teacher evaluation systems had no impact on professional development. In these districts, professional development continues to be focused on district-wide initiatives rather than evaluation results. These districts are aware that they ought to be linking professional development and evaluation results more consistently to link training with observed needs and plan to do so in the near future.

Districts cited several challenges, including competing state and district initiatives and leaving professional development decisions up to teacher and principal discretion. However, finding time for planning and collaboration was viewed as the primary obstacle to effective professional development, and developing SLOs, in particular, was viewed as an extensively time-consuming process. Successful implementers were able to overcome these obstacles by rolling out training deliberately and in a strategic order to help teachers see how all district initiatives were connected and aligned.

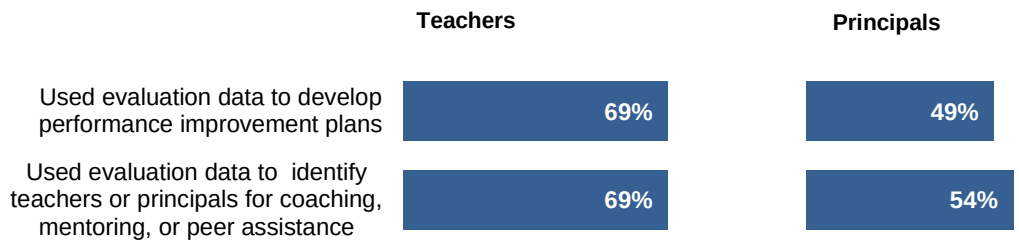
Districts cited several challenges that made it difficult to link professional development with evaluation results. For example, districts reported that training to implement Common Core has dominated the professional development focus of many districts in recent years, leaving less time for other professional learning opportunities. Further, some districts have a tradition of leaving some professional development decisions up to teacher and principal discretion (often with administrator approval). In one district,

teachers believed that using evaluation data for teacher-level decisions is not appropriate and that it should only be used at greater levels of aggregation. Finally, in some districts, principals and district administrators felt they lacked the experience or appropriate skills to direct professional development to their teachers or principals. For example, one principal with a middle school background felt unprepared to provide professional development for elementary teachers in her building.

What professional development is being provided to "needs improvement" teachers?

Teachers rated as "needs improvement" are typically put on a performance improvement plan, and some districts have chosen to focus specific professional development recommendations solely on those teachers.

Figure 23. Most districts used evaluation results to develop performance improvement plans or offer support to teachers rated as "needs improvement." About half of districts took a similar approach for low-performing principals



District staff interviewed indicated that teachers rated as "needs improvement" are typically put on a performance improvement plan. These plans often require additional, individualized coaching or mentoring from peers or administrators; setting growth targets for professional practice or student achievement, more frequent meetings to monitor progress, and more prescriptive professional development. Less frequently, such teachers are asked to observe the classrooms of teachers who had high ratings in the target area. There is typically a limit on the number of days a teacher can stay on a professional development plan, and teachers rated as "needs improvement" could be dismissed if they did not improve after a specified period of time. In 74 percent of districts responding to our survey, there is a limit on the number of days a teacher can stay on a professional development plan.

In one district, teachers rated as "needs improvement" must work with their evaluators to develop a remediation plan within 30 days after the evaluation is completed. The plan must identify specific

instructional areas that need improvement and describe the supports the district will make available, and teachers have 90 days to remediate their deficiencies and must receive an additional evaluation within 10 days of this time period. (Teachers rated as "needs improvement" in some districts are also subject to more frequent meetings and check-ins to monitor their progress, may be required to set growth targets, and typically receive more prescriptive professional development than other teachers.)

According to our interviews, it was relatively rare for a tenured teacher to be rated as "needs improvement," so policies related to performance improvement plans and consequences have yet to be implemented in most districts.

Many of the interview districts offered individualized coaching or job-embedded professional development to support teachers identified as needing improvement. For example, Rock Island-Milan School District provides peer mentoring, observations, and reflective conversations in addition to the teacher's professional development plan. In some districts, teachers needing improvement will receive support from a mentor from their building who is strong in an area needing improvement, based on their evaluation results. Niles Township School District, for example, uses a PAR model to support struggling teachers (for more information on PAR in Niles, see White et al., 2012). In other districts, teachers who need improvement can observe the classrooms of teachers who had high ratings in the target area, which one teacher singled out as the most useful aspect of his professional development.

Some administrators from smaller districts felt that their size facilitated additional opportunities to work one-on-one with teachers who need improvement. For instance, Unity Point uses a peer-to-peer, nonevaluative process to work with a mentor to review the Danielson Framework and watch video-recorded lessons to help teachers see what the evaluator is seeing, and then go through a self-assessment process with the superintendent. Other, typically larger, districts tended to look to external support for struggling teachers, or make recommendations for external training and resources. Districts with active teachers unions were also able to turn to these associations to support struggling teachers.

Several districts noted that teachers rated as "needs improvement" could be dismissed if they did not improve after a specified period of time and that performance ratings were taken into consideration when districts were forced to make a RIF. A district administrator from one district that dismissed two nontenured teachers for performance last year noted that the new evaluation system provided ample documentation and more than enough support for this process. However, another superintendent reported that PERA and Senate Bill 7 have not substantially changed teacher retention decisions and that districts still have difficulty dismissing low-performing teachers.

What are the central themes of conversations between teachers and evaluators? What makes for a high-quality conversation?

High-quality conversations between teachers and evaluators were collaborative, were based on the rubric, and provided specific feedback.

Several interview respondents noted that conversations between teachers and evaluators were the most useful aspect of the new evaluation systems. Teachers generally valued the mid-cycle and end-of-year meetings to discuss evaluation findings with their evaluator and stressed that quality feedback was vital for instructional improvement. Pre-observation conferences typically involved a discussion of the lesson to be delivered, the features that the evaluator would be looking for, and features that the teacher hoped to focus on or improve. Postobservation conferences usually involved a discussion of the lesson with reference to the professional practice rubrics and provided an opportunity for teachers to present artifacts or additional evidence needed for context, and for both parties to share any questions or concerns they might have. Evaluators and teachers often discussed examples of performance at various levels (e.g., pointing out differences between proficient and excellent). The conference typically concludes with mutual identification of strengths and areas for improvement, time for reflection, and feedback about how to improve practice or recommendations for professional development and support. However, as noted above, if the teacher is rated as "needs improvement," the postobservation conference is typically more directive and specific about what the teacher needs to do to improve.

Several districts noted that the Danielson rubrics provided concrete language to help guide improvement and have led to better, richer conversations about instructional improvement.

Teachers generally felt that conversations with evaluators were valuable, and both teachers and administrators emphasized that high-quality feedback was vital for instructional improvement. As one teacher put it, the whole evaluation process can hinge on the quality of these conversations and how well the evaluator embodies the role of an instructional leader. Respondents noted that one basic element of high-quality conversation is having all requisite paperwork filled out ahead of time. Beyond that, productive conversations between teachers and evaluators were viewed as more of a conversation or "another set of eyes" on performance, rather than a one-sided lecture. Collaborative discussions where the evaluator lets teachers guide the conversation and asks them how they thought the observed lesson went and where the evidence fell, leading them to language from the rubrics—rather than the evaluator leading the conversation and simply pointing out what the teacher did wrong—were viewed as especially productive. For example, one district noted that the evaluation system has helped teachers know what's expected of them and what it takes to improve and helped them feel as though they are being treated as

professionals. They say this is due not to teachers' evaluation ratings, but to the conversations the district is having around teacher practice.

Insights From the Case Studies: The Value of Teacher-Evaluator Conversations

Conversations that occur both prior to classroom observation and after an observation were described as one of the most valued aspects of PERA teacher evaluations. Stakeholders in all case study districts appreciated how the process emphasizes high-quality teaching and how it has shifted the nature of performance discussions to focus on student learning.

Teachers generally described conversations with their evaluators as starting with a discussion of what did or did not go well during a recent observation (based on the Danielson Framework). Evaluators shared their ratings based on the observation rubric; the teachers shared their perspectives or reaction to the ratings and in some cases offered additional artifacts to further demonstrate their performance. From there, the conversations shifted to discussions of specific resources and strategies to improve performance. As one evaluator said:

> *"We focus on what went well and where things need improvement, and we really put more focus on resources and strategies to help that teacher improve. It's a collaboration. It's a good time to brainstorm together, to come together, to problem solve together."*

Several teachers commented on the benefits of such conversations:

> *"Most helpful was the administrator's eagerness to provide assistance, genuine assistance* [in coming up with ideas to improve]*."*

> "[The feedback from observations] *can be really, really specific and not just general classroom management. They can be specific on exactly what it was that caused the good classroom management. The comments I got were really insightful, and they really knew, and were right on target, with the problem areas that I need to work on."*

Both teachers and their evaluators also valued how the evaluation system seems to have spurred conversation and collaboration between and among teachers that is more focused on teaching and learning. As one evaluator said:

> *"The collaboration between teachers has been really good. At first, I thought they might feel competitive and not share curriculum because this system creates competition. Really, I find that did not happen, and I'm so glad because I didn't want our culture to, you know, the bottom to fall out because of a competitive nature. I find they are collaborating, still sharing. I've observed them in meetings sitting with the tool and having discussions about how they are going to be excellent in this area or that area. It's been really good."*

On the other hand, evaluative conversations often fail to reach their full potential in instances where consideration of additional artifacts is inconsistent or incomplete. Two domains in the Danielson Framework (Planning and Preparation and Professional Responsibilities) involve review of additional artifacts. Artifacts could also provide a more thorough view of domains or subdomains than what could be observed in a single lesson. Several case study districts reported inconsistent use of additional artifacts in the evaluation process. As a teacher noted, this can be frustrating:

> *"I do like the ability to submit lesson plans and student work samples and things like that. I think it broadens the view of my teaching. It takes a lot of time to gather those things, but I think it gives me the opportunity to have a more meaningful conversation with the administrator…. it's frustrating when we're too rushed to even review them."*

What are the features of conversations that take place during principal evaluation?

Conversations between principals and their evaluators are also likely to be important for improving leadership practice. Interview participants from Sterling Community Unit School District reported that productive principal-evaluator conversations focused on instructional leadership, rather than on day-to-

day building management and were the most useful aspect of principal observation. The district felt it was important to reflect the importance of instructional leadership in the observations for principal evaluation. Principals interviewed in Sterling indicated this helped them be better evaluators of teacher performance. The interviewees also found the self-reflection component an important aspect of the principal observation.

In some other districts, interviewees had less positive views about conversations. Concerns included that flaws or limitations of the rating tool (Vanderbilt Assessment of Leadership in Education, or VAL-ED) skewed the conversation; the goals for the principals' evaluations were set at the district level and were not necessarily aligned with or reflective of building factors or an individual principal's or assistant principal's role, and that formal conversations between principal and their evaluators were not always scheduled (as is done in the teacher system) but rather more informal, flexible.

Insights From the Case Studies: Content of Principal-Evaluator Conversations

Most principal-evaluator conversations included a discussion of what went well or didn't go well during the past school year in the building, professional strengths and weaknesses, reflections on performance, and ways to enhance practice. One superintendent described observing and providing feedback to a principal on a challenging area:

> *"We had a principal, one of his weak areas he acknowledges was having difficult conversations and, you know, conflict. He says that, 'I'm not good, I don't like it,' and none of us do. So one of the things that, for that principal in particular, we went to a teacher that…it was pretty obvious was not a very strong teacher. So, I said, 'we're going to watch this teacher together,' and so… but we did this with a lot of different teachers…I just want to see what the principal took on and what they choose not to take on... the principal and I then meet afterwards and we go over everything of things that I saw through the process and all that… . That's been really good, that's been a really good process."*

Principals and evaluators indicated that using PERA-compliant evaluation rubrics, such as the Illinois Performance Standards for School Leaders, has focused their conversation more on instructional leadership, goal setting related to professional practice, student growth, and nonacademic goals such as student behavior. PERA-compliant rubrics have provided a more holistic and objective description of what is included in the role of a principal compared to previous rubrics used. As one principal evaluator indicated:

> *"it's one of the most important things that our principals do, to lead instruction and up to two years ago, three years ago, I never evaluated the most important thing that our principals do…. Having six buildings, I can manage it, and so it's been again extremely valuable for me and for our principals."*

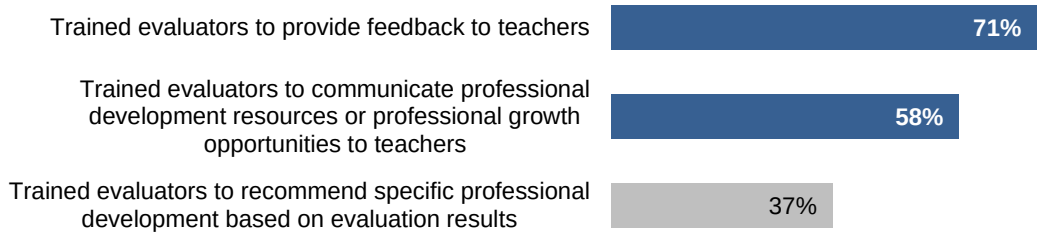
Principal evaluation conversations were less beneficial when the evaluation rubric was considered faulty or even invalid, application of evaluation standards were overly rigid, or where the relationship between a principal and the evaluator was strained. Some principals expressed dissatisfaction with being required to use common goals across the district that may not reflect important building-level factors.

Finally, in contrast to teachers, principal evaluation conversations tended to be more informal and less often scheduled to a specific time. While some participants appreciated the flexibility afforded by this approach, others saw it as risky since time dedicated to reflection and planning for improvement was not being set aside consistently. As one principal said:

> *"Formally sitting down and having those conversations, I don't think we're doing a very good job. Whether that's, and I think that's kind of across the board with our administrators. We've got a really strong administrative team, and I think that kind of turns into not taking the time to formally sit down and have those conversations like we should. We talk about it a lot, and we are going to do better at that, and at that, but I don't think we've moved there yet."*

What types of training do principals get to support teacher growth? How effective has this training been?

Figure 24. Most districts trained evaluators to provide feedback to teachers or communicate professional development resources or professional growth opportunities to teachers



Some interviewees observed that the transition to the new evaluation system was especially hard for new principals because they needed additional training about how to handle "tough conversations" using language from the Danielson Framework and providing feedback in terms of questions that reflect the Framework's "distinguished" category. However, one district representative stated that she was not aware of any high-quality training on having these "tough conversations," adding that the training offered was more focused on being efficient with the Framework and occasionally offered potentially counterproductive guidance (e.g., "Start out with Domain 3c, if they're proficient with that, they're probably proficient with everything else….").

What do districts do to evaluate how well the professional development provided is working to improve teacher performance?

While most districts reported evaluating their professional development efforts, in most cases this was done primarily through surveys or other input from teachers or principals.

Ninety percent of the districts responding to the survey reported that they evaluated how professional development affects educator performance.

- Typically districts evaluate the impact of professional development by surveying educators' satisfaction with current professional development and suggestions for future opportunities. Ninety percent of districts reported collecting

this information from teachers and 76 percent from principals. Eighty-nine percent used feedback from teachers and 76 percent from principals to adapt future professional development.

Interviews confirmed that districts evaluated the influence of professional development on teacher practice largely by collecting information on educators' satisfaction with professional development and their suggestions for future opportunities.

How are the changes in professional development affecting teachers? Has the professional development provided affected teaching practice?

Most districts reported that they believed evaluation and the associated professional development had a positive impact on teachers. However, many districts cannot specifically attribute changes to PERA.

In several districts, the staff interviewed provided examples of how conversations with evaluators and professional development around assessment have affected teaching practice, For example, one district reports that the level of conversations in the district is much better than prior to implementation of the evaluation system and that teachers have a better understanding of the rationale behind their instruction and of their own strengths and weaknesses.

In most of the interview districts, interviewees stated that they believed PERA was having a positive impact on teacher growth. According to one district administrator, the impact of the teacher evaluations has "not been as earth-shattering as one would expect, but rather a process that happens gradually." However, interviews also revealed that many districts had difficulty attributing changes specifically to PERA, and many lacked the capacity to systematically track the impacts of their evaluation systems and the associated professional development.

A primary obstacle to effective professional development, mentioned by multiple districts, seemed to be finding time for planning and collaboration. Several districts noted that they are being asked to implement many new initiatives simultaneously and have insufficient time to devote to each of these the time they deserve. Designing teacher and principal evaluation systems, conducting multiple classroom observations with associated conferences, and, in particular (in districts using this method) developing SLOs, were often viewed as extensively time-consuming processes that left little time for additional professional development. Unity Point School District has tried to overcome these obstacles by being strategic about connecting all the initiatives going on in the district and rolling out training deliberately and in order to

help staff see how they were aligned—beginning with induction and mentoring, moving to evaluation of professional practice, then Common Core, and now assessments and student growth.

Summary: How is professional development changing and how is it affecting teachers?

Districts have spent substantial amounts of time on training educators about the new evaluation systems. PERA's emphasis on student achievement growth has resulted in districts providing more professional development on student assessment and goal setting. Most districts are now using performance evaluation results to plan district-wide professional development and especially to provide professional development for individual educators. In many of the study districts, the major benefits of PERA-compliant evaluation systems appear to be providing clear expectations of good professional practice and clear and consistent communication to both teachers and evaluators about what it looks like in practice. Interviewees report that when done well, evaluation conversations and discussions are focused on improvements in teaching practices, and this is linked to teachers' professional development.

District staff interviewed reported benefits of the professional practice evaluation, including providing a common language, establishing clear expectations in terms of behaviors and practices, improving the level conversations between teachers and evaluators, and more clearly identifying teaching strengths and weaknesses.

References

McCaffrey, D. F. (2013). Will teacher value-added scores change when accountability tests change? Carnegie Knowledge Network. Retrieved from http://www.carnegieknowledgenetwork.org/briefs/value-added/accountability-tests/

Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, *48*(1): 163–193.

Stuit, D., Berends, M., Austin, M. J., & Gerdeman, R. D. (2014). *Comparing estimates of teacher value-added based on criterion- and norm-referenced tests* (REL 2014–004). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

White, B. R., Cowhy, J., Stevens, W. D., & Sporte, S. E. (2012). *Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts.* Research Brief. Chicago, IL: Consortium on Chicago School Research.

Appendix A. District Survey Tables

Table A-1. Percentage of districts reporting use of information sources other than observations to evaluate teacher or principal practice (*N*=41)

Source of information	Percent using for information source for teacher evaluation	Percent using for information source for principal evaluation
Self-assessment	59	90
Portfolio	92	71
Assessment by peer or mentor	10	2
Student work samples	71	NA
Documents like school improvement plans, budgets, school handbooks, individualized education plans	NA	73
Student survey or other student feedback	27	NA
Parent survey or other parent feedback	22	37
School climate survey	NA	51

NA=not applicable.

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-2. Actions taken by districts to monitor the implementation of their teacher or principal evaluation systems

Action	Percent of districts that reported monitoring teacher evaluation implementation (*N*=39)	Percent of districts that reported monitoring principal evaluation implementation (*N*=30)
Appointed a person/committee responsible for ensuring teacher evaluation system is implemented as designed/intended	92	67
A system was in place to track implementation problems	69	70
Assessed whether staff have necessary competencies to implement evaluations	92	80
Assessed evaluators' understanding of components of the evaluation system	82	67
Assessed evaluatees' understanding of the components of the evaluation system	67	70
Tracked training of evaluators to ensure all were properly trained	97	97
Tracked whether educators were observed the number of times required	95	93
Monitored whether evaluations were completed on time	97	100
Surveyed/interviewed evaluatees about their experiences with the system	54	47
Surveyed/interviewed evaluators about their experiences with the system	54	45
District staff sat in on observations or feedback sessions	41	30
Reviewed evaluator observation summaries and written feedback forms for quality	74	73
Reviewed final evaluation ratings for quality	85	80
Examined distribution of evaluation ratings	92	80
Monitored for evaluation bias and/or common errors	54	47

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-3. Percentage of districts reporting specific activities used to evaluate the usefulness of evaluation systems for improving practice

Activities	Percent of districts that reported evaluating the usefulness of teacher evaluation for improving practice (*N*=39)	Percent of districts that reported evaluating the usefulness of principal evaluation for improving practice (*N*=37)
Examined whether evaluation results improved over time	51	51
Examined whether teachers participated in professional development related to aspects of performance for which their evaluation ratings were low	50	NA
Conducted interviews/surveys about improvements in teaching practices	44	32
Examined whether evaluation results improved after professional development	37	NA
Examined whether improvements in teacher evaluation were related to changes in student achievements/growth	33	49
Worked with an external evaluator to assess improvements in practice or increases in student achievement	18	11
Analyzed results from 5 Essentials Survey	NA	76
Other	3	8

NA=not applicable.

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-4. Percentage of districts reporting using evaluation information to make human resource management decisions for teachers and principals

Decision	Percent using evaluation information for decision about teachers (*N*=41)	Percent using evaluation information for decision about principals (*N*=38)
Deciding whether to retain/renew	80	87
Planning individual professional development	78	84
Planning district-wide professional development	73	71
Tenure/probationary status decisions for teachers	73	NA
Teacher assignments to classrooms/grades/subjects within schools	34	NA
Promotion decisions	32	68
Assignments to schools	10	42
Salary decisions or movement along the salary schedule	5	39

NA=not applicable.

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-5. Percentage of districts reporting different frequencies of training for teachers on the teacher evaluation system (*N*=39)

	Frequently	Infrequently	Not at all	Not applicable
Understanding the professional practice rubric	69	28	3	-
Understanding how the district measures student growth	51	36	8	5
Understanding the overall evaluation system	46	51	3	-
Creating student learning objectives or other measures of student learning	44	33	10	13
Understanding how to obtain baseline data to set growth targets and measure performance	35	49	10	5

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-6. Percentage of districts providing evaluation system training to evaluators of teachers and principals

Training topic	Percent providing for evaluators of teachers (*N*=39)	Percent providing for evaluators of principals (*N*=38)
Evaluators were trained on the components of the evaluation system	82	87
Evaluators were trained on the professional practice rubric	84	89
Evaluators tested to assess their accuracy of using the practice rubric	42	42
Evaluators trained to provide feedback to educators about their professional practice	71	57
Evaluators trained to communicate professional development resources and/or professional growth opportunities	58	53
Evaluators trained to recommend specific types of professional development activities to teachers/principals based on their evaluation ratings	37	42

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-7. Percentage of districts using different combinations of Type I/II assessments and measurement models to measure achievement growth (*N*=31)

Type I/II assessment	Statistical model (e.g., value-added)	Simple growth (pretest-posttest)	Percent of progress toward goal	Percent of students meeting growth goal/target	Other
State test	3	6	6	19	5
Commercially published tests	10	26	19	39	0
District-developed tests	13	42	16	35	0
District-developed performance tasks or assessments	10	35	10	26	0

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-8. Percentage of districts using different combinations of Type III assessments and measurement models to measure achievement growth (*N*=28)

Type III assessment	Simple growth (pretest-posttest)	Percent of progress toward goal	Percent of students meeting growth goal/target	Other
School-developed standardized tests	29	14	36	7
School-developed performance tasks, performance assessments, or student portfolios	32	14	50	4
Teacher-developed standardized tests, tests from textbooks	39	21	50	4
Teacher-developed performance tasks, performance assessments, or student portfolios	43	14	57	4

SOURCE: Survey of school districts participating in the study (spring 2015).

Table A-9. Percentage of districts reporting acquiring various additional resources to measure student achievement growth (*N*=41)

Resource	Percentage of districts
Hired additional staff with expertise in these area	17
Hired or contracted with a consultant or other external agency to advise the district about how to develop growth measures or calculate growth	39
Hired or contracted with a consultant or other external agency to calculate your growth measures	24
Purchased new computer software to store/track growth data or analyze growth data	7
Sent staff to external training on tracking growth data, developing growth measures, or calculating growth	63
Hired or contracted with a consultant or other external agency to develop a data system or data warehouse for tracking growth data	22
Any of these resources	90

SOURCE: Survey of school districts participating in the study (spring 2015).

Appendix B. Growth Analysis Correlation Coefficients

Table B-1. Correlations between school average growth ratings for teachers and school-level reading and mathematics value-added: 2014–15

District	Correlation of school average growth rating with reading value-added	Correlation of school average growth rating with mathematics value-added
C	.19	.15
F	-.01	-.72
I	.59	.01
K	-.81	-.81
L	-.69	-.05
N	.46	-.20
E[1]	-1.00	-1.00
A[1]	-1.00	1.00
O[2]	NA	NA
P[2]	NA	NA

NA: Not applicable

[1] Data were available for only two schools in these districts.

[2] Data were available for just one school in these districts

SOURCE: Calculations based on comparison of district-provided growth ratings and value-added analysis.

Table B-2. Correlations between school average overall summative ratings for teachers and school-level reading and mathematics value-added: 2014–15

District	Correlation of school average summative rating with reading value-added	Correlation of school average summative rating with mathematics value-added
A[1]	1.00	1.00
B	.35	-.39
C	.28	.15
D	.06	.05
E[1]	-1.00	1.00
F	.94	.40
G	.56	-.72
H	-.07	-.05
I	.97	-.64
J[1]	-1.00	-1.00
L	.09	-.15
M	-.03	.32
N	.37	.38
O[2]	NA	NA
P[2]	NA	NA
Q	.69	.37

NA: Not applicable.

[1]Data were available for only two schools in these districts.

[2]Data were available for just one school in these districts

SOURCE: Calculations based on comparison of district-provided growth ratings and value-added analysis.

Table B-3. Correlations between school average administrator growth ratings[1] for and school-level reading and mathematics value-added: 2014–15

District	Correlation of school average growth rating with reading value-added	Correlation of school average growth rating with mathematics value-added
F	-.77	-.99
I	.56	.06
Q	.81	.87
K	-.16	-.05
L	-.61	-.73
N	-.26	0.00
H	-.31	-.07
M	.38	.39
R	.06	-.28
P	.95	.99
E[2]	NA	NA
D[2]	NA	NA
O[2]	NA	NA
S[2]	NA	NA
J[3]	NA	NA

NA: Not applicable

[1]Ratings for principals and associate/assistant principals were averaged.

[2]No variation in growth ratings between schools in these districts.

[3]Data were available for only one school in this district.

SOURCE: Calculations based on comparison of district-provided growth ratings and value-added analysis.